AI and Spatial Sorting

Ziqian Song [†]

Sciences Po

This Draft: Jan 2026

Artificial intelligence (AI) affects cognitive tasks performed by skilled labor, potentially altering the skill-biased productivity advantages of large cities and spatial sorting. In this paper, we quantify the aggregate and distributional effects of AI-induced spatial sorting using a spatial equilibrium model with heterogeneous labor. We find that AI is relative substitute for skilled labor and relative complement for unskilled labor. Counterfactual analysis shows that AI automation overturn the spatial sorting of skilled labor, with the skill ratio and skill premium decrease most in large cities. We also find that AI automation increases aggregate productivity via knowledge spillover channels, driven by the relocation of unskilled labor to large cities.

Keywords: artificial intelligence, spatial sorting, agglomeration, migration, productivity
JEL Codes: R12, R23, O33, J24

[†]Email: ziqian.song@sciencespo.fr

1 Introduction

A dominant pattern in urban economics in recent decades is the disproportionate spatial sorting of skilled labor into large cities, accompanied by a dramatic increase in the skill premium (Diamond, 2016; Diamond and Gaubert, 2022). Emerging technologies such as artificial intelligence (AI) may disrupt this pattern by automating cognitive tasks that underlie urban agglomeration economics (Combes et al., 2012; Davis and Dingel, 2019), thereby reshaping the comparative advantage of skilled labor in large cities and fundamentally altering patterns of spatial sorting. This raises two key questions: How does AI influence the spatial distribution of workers? And what are the resulting welfare and distributional implications? Addressing these questions is critical for understanding how to enhance welfare and reduce inequality in the era of AI.

In this paper, we document AI-induced spatial sorting in China and quantify its aggregate and distributional effects. China provides a unique setting for this research due to several reasons. First, China is advanced in industrial AI adoption, with widespread adoption across regions and industries. Second, the country has undergone rapid urbanization in recent decades, characterized by large-scale internal migration and the disproportionate concentration of skilled labor in large cities. These two factors generate substantial variation that allows us to analyze the effects of AI on spatial sorting. Additionally, high migration barriers across regions exist due to China's hukou system. This context allows us to evaluate migration policies in the era of AI and derive normative implications for policymakers.[1]

We integrate a task-based model of artificial intelligence (AI) into a Rosen-Roback spatial equilibrium framework with heterogeneous labor (Acemoglu and Restrepo, 2018; Rosen, 1979; Roback, 1982). Representative firms in heterogeneous cities optimally choose input combinations and task allocations, given the elasticity of substitution between AI capital and each type of labor. Workers of different skill types make optimal migration decisions based on migration costs and idiosyncratic preferences over locations. We also directly impose a functional form for the skill-biased productivity advantage in large cities through endogenous agglomeration forces (Combes et al., 2012; Davis and Dingel, 2019). This allows us to clearly analyze the relationship between AI and spatial sorting via the knowledge spillover channel.

We measure AI capital by combining software asset data from Chinese publicly listed firms with AI-related research grants from the National Natural Science Foundation of China (NSFC). Specifically, our measure isolates the variation in firms' software assets explained by their exposure to AI-related public R&D shocks. We then estimate the elasticity of substitution between AI

[1]The hukou system is a household registration system in China based on an individual's family background. It determines where people are officially permitted to live, work, and access public services. In particular, the hukou system restricts migrants' access to public goods in urban areas, with local governments varying the strictness of enforcement over time.

and both skilled and unskilled labor. Structural estimation indicate that AI is a relative substitute for skilled labor and a relative complement for unskilled labor, in contrast to the predictions of canonical capital–skill complementarity (Krusell et al., 2000). Our findings suggest that AI exhibits preliminary characteristics of a skill-substituting technology, with potential implications for mitigating labor market inequality.

In this paper, we focus on two types of AI automation: (1) The extensive margin of AI automation, represented by an expansion of the task automation frontier, and (2) The intensive margin of AI automation, denoted by a uniform increase in the productivity of AI capital. To quantify the spatial equilibrium effects of AI automation, we construct a novel measure of AI exposure based on the textual similarity between occupational tasks and U.S. AI patents. We then use this AI exposure measure to recover the extensive margin of task displacement in Chinese cities from 2013 to 2018, exploiting the AI automation-induced component of changes in AI capital share and skilled wages. In addition, we use relative task performance between humans and AI from Kiela et al. (2021, 2023) to recover the intensive margin of AI automation.

We conduct a series of counterfactual simulations of AI automation within our spatial equilibrium model. We find that AI automation reverses the spatial sorting of skilled labor into large cities, reducing the skill ratio and skill premium most sharply in these cognitive hubs. We also show that AI automation increases aggregate output, reduces skilled welfare, and lowers the skill premium. By shutting down the endogenous skill-biased agglomeration force, we find that AI automation has negligible aggregate effects. This highlights that knowledge spillovers are a key mechanism driving the aggregate productivity gains from AI-induced spatial sorting.

While the extensive and intensive margins of AI automation share similar effects on skilled sorting, they differ fundamentally in their impact on unskilled labor. At the extensive margin, the substitution effect of AI automation dominates the productivity effect. As AI displaces skilled labor without generating compensating productivity gains, it causes unskilled labor to out-migrate from large cities. In contrast, the intensive margin generates strong productivity effects that attract unskilled labor to sort into large cities. We find that the aggregate impact of the intensive margin is two to four times greater than that of the extensive margin.

This research relates to several strands of literature. First, this paper contributes to the literature on the spatial sorting of workers across cities. Seminal works by Combes et al. (2008) and Diamond (2016) document the increasing concentration of high-skilled workers in high-wage, high-amenity cities, contributing to the dramatic increase in the skill premium and regional inequality. A key mechanism underlying spatial sorting is knowledge spillovers (Combes et al., 2012; Gaubert, 2018; Davis and Dingel, 2019). Theoretically, Eeckhout et al. (2014) explain this pattern through extreme skill complementarity in large cities, while Behrens et al. (2014) propose a framework disentangling how agglomeration economies, sorting, and selection jointly deter-

mine labor allocation. Our framework builds on Giannone (2017), who study how skill-biased technical change interacts with agglomeration to reduce regional convergence. We extend this literature by incorporating AI automation into the production function within a setting of skill-biased agglomeration. While existing studies emphasize the disproportionate sorting of skilled labor into large cities, we are among the first to show that AI automation could reverse this pattern and reduce inequality. A novel implication of our analysis is that AI-induced spatial sorting contributes to aggregate productivity gains primarily through the knowledge spillover channel.

Second, this paper complements the growing literature on the labor market effects of technological change. Krusell et al. (2000) estimate capital-skill complementarity within a production function to explain the increase in the skill premium. Autor and Dorn (2013) document that the automation of routine tasks displaces middle-skill jobs, generating job polarization . Building on this, Eeckhout et al. (2021) show that the routine cognitive employment share decreases more in large cities due to IT adoption. More recent work has focused on AI: Webb (2019) and Acemoglu et al. (2022) analyze the early labor market impacts of AI, while experimental studies investigate how large language models (LLMs) affect productivity and task performance (Noy and Zhang, 2023; Brynjolfsson et al., 2023; Peng et al., 2023; Eloundou et al., 2024). Our paper complements this literature by providing the first estimate of the elasticity of substitution between AI and heterogeneous labor. We also contribute by documenting AI-induced spatial sorting in China and quantifying its aggregate and distributional implications.

Finally, this paper contributes to the development literature on internal migration. While the prior work has focused on migration barriers (Bryan and Morten, 2019; Tombe and Zhu, 2019), insurance barriers (Munshi and Rosenzweig, 2016), pollution (Khanna et al., 2025), land security (Adamopoulos et al., 2024), and family migration mode (Imbert et al., 2023) in shaping spatial distributions of workers, we propose a new perspective on how AI affects internal migration decisions and their aggregate implications. Our findings show that in developing countries with low human capital, unskilled labor that complements AI is fundamental in driving the aggregate productivity effects of AI. Our paper also provides normative implications for policymakers on designing migration policy to manage AI-induced spatial sorting.

The remainder of the paper is organized as follows. Section 2 introduces the theoretical framework. Section 3 presents the structural estimation and model fit. Section 4 conducts counterfactual analyses. Section 5 concludes.

2 Theoretical Model

In this section, we develop a model to examine how AI affects the spatial distribution of heterogeneous labor across cities. To do so, we incorporate a task-based framework of artificial

intelligence (AI) into a Rosen–Roback spatial equilibrium model (Acemoglu and Restrepo, 2018; Rosen, 1979; Roback, 1982). This framework provides a micro-foundation for capturing how AI displaces tasks traditionally performed by skilled labor.

The economy consists of cities indexed by $j \in \mathcal{D}$. Representative firms are located in heterogeneous cities and employ workers of different skill types in production. Workers are born in cities indexed by $i \in \mathcal{O}$ and are exogenously assigned to be either skilled or unskilled. Each worker makes a nested location decision: in the upper nest, they choose whether to stay in their origin city or migrate; in the lower nest, they choose to migrate to destination city j conditional on migrating.

2.1 Production

In each city j, a representative firm operates under perfect competition and takes the price of the final good as the numeraire. Output Y_j is produced using a nested Constant Elasticity of Substitution (CES) production function that combines unskilled labor (U_j) with a composite input of skilled labor and AI capital (ES_j). In the inner nest, we incorporate a task-based framework in which tasks are performed by AI capital (K_j) and skilled labor (S_j). These inputs are then aggregated into a CES composite (ES_j).

Outer CES Nest The outer nest aggregates unskilled labor and the AI–skilled labor composite as follows:

$$Y_j = \left[\eta_{uj}(U_j)^{\sigma} + \eta_{sj}(ES_j)^{\sigma}\right]^{\frac{1}{\sigma}}, \tag{1}$$

where U_j is unskilled labor, ES_j is the AI–skilled labor composite, and η_{uj} and η_{sj} are city-specific productivity terms for unskilled labor and the AI–skilled composite, respectively. The elasticity of substitution between unskilled labor and the AI–skilled composite is $1/(\sigma - 1)$.

Building on the literature on skill-biased agglomeration (Combes et al., 2012; Giannone, 2017; Davis and Dingel, 2019), we model agglomeration forces as a function of the local skill composition, total population, and exogenous fundamentals:

$$\eta_{sj} = \left(\frac{S_j}{U_j}\right)^{\lambda_s} (S_j + U_j)^{\phi_s} \exp(\xi_{sj}) \tag{2}$$

$$\eta_{uj} = \left(\frac{S_j}{U_j}\right)^{\lambda_u} (S_j + U_j)^{\phi_u} \exp(\xi_{uj}) \tag{3}$$

where ξ_{sj} and ξ_{uj} denotes fundamental productivity of city j for the AI-skilled labor composite and unskilled labor, respectively. S_j and U_j represents skilled and unskilled labor. The parameters (λ_s, ϕ_s) and (λ_u, ϕ_u) govern the agglomeration force associated with the skill ratio and total

population, respectively.

Inner CES Nest The inner nest is assembled from a continuum of tasks $i \in [0,1]$. The AI–skilled labor composite ES_j is defined as:

$$ES_j = \left[\int_0^1 z_j(i)^\rho di \right]^{\frac{1}{\rho}} \tag{4}$$

where $z_j(i)$ denotes the intermediate input for task i in city j, and ρ governs the elasticity of substitution across tasks.

The intermediate good $z_j(i)$ can be produced by either skilled labor or AI capital. All tasks $i \in [0,1]$ can be performed by skilled labor, but only those in the interval $[0, \beta_{kj}]$ are automatable and can be performed by AI capital. The task-level production function is given by:

$$z_j(i) = \begin{cases} A_s l(i), & \text{if } i \in [0,1] \\ A_k \psi_k(i) k(i), & \text{if } i \in [0, \beta_{kj}] \end{cases} \tag{5}$$

where $l(i)$ is the skilled labor input, $k(i)$ is the AI capital input. $z_j(i)$ is the AI capital–specific productivity for task i in city j. A_s and A_k are the factor-augmenting terms for skilled labor and AI capital, respectively.

Producer of the AI–skilled labor composite ES_j maximizes profits taking prices as given, so the inverse demand for each intermediate task i is:

$$p_j(i) = \left[\frac{ES_j}{z_j(i)} \right]^{1-\rho} \tag{6}$$

The price of each task $p_j(i)$ is equal to the marginal cost of the input used for its production:

$$p_j(i) = \begin{cases} \dfrac{w_j^s}{A_s}, & \text{if} \quad i \in [\beta_{kj}, 1], \\ \min\left\{ \dfrac{w_j^s}{A_s}, \dfrac{1}{A_k \psi_k(i)} \right\}, & \text{if} \quad i \in [0, \beta_{kj}]. \end{cases} \tag{7}$$

where w_j^s is the skilled wage in city j, and the price of AI capital is normalized to one.

Consistent with the framework of Acemoglu and Restrepo (2018), given that skilled labor holds a strict comparative advantage in higher-indexed tasks, there exists a unique threshold $\tilde{\beta}_{kj} \in (0,1)$ such that:

$$\frac{w_j^s}{A_s} = \frac{1}{A_k \psi_k(\tilde{\beta}_{kj})} \tag{8}$$

The city-specific threshold $\tilde{\beta}_{kj}$ denotes the task index at which the unit cost of producing with AI capital equals that of skilled labor. For all tasks $i \in [0, \beta_{kj}]$, we have $\frac{1}{\psi_k(\tilde{\beta}_{kj})} < \frac{A_k w_j^s}{A_s}$, implying these tasks are performed by AI capital. We impose the following assumption to guarantee that the automation threshold is bounded by the technological frontier:

Assumption 1 (Binding Automation Frontier)**.** *The automation frontier is strictly below the economic profitability threshold, $\tilde{\beta}_{kj}$, for all cities:*

$$\beta_{kj} < \tilde{\beta}_{kj}, \quad \textit{for all } j \in \mathcal{D}.$$

Assumption 1 implies that the technological constraint is binding. Therefore, any expansion of the automation frontier β results in the immediate reallocation of tasks from skilled labor to AI capital.

Solving the firm's cost minimization problem and integrating over tasks yields the composite AI–skilled input as:

$$ES_j = \left[\left(\left(\Gamma_{kj}\right)^{\frac{1-\rho}{\rho}} K_j \right)^{\rho} + \left((1-\beta_{kj})^{\frac{1-\rho}{\rho}} A_s S_j \right)^{\rho} \right]^{\frac{1}{\rho}} \tag{9}$$

where Γ_{kj} is the effective productivity of AI capital, defined as:

$$\Gamma_{kj} = \int_0^{\beta_{kj}} \psi_k(i)^{\frac{\rho}{1-\rho}} di \tag{10}$$

Substituting this expression into the outer CES function gives the final aggregate production function:

$$Y_j = A_j \left\{ \eta_{uj} (U_j)^{\sigma} + \eta_{sj} \left[\left(\Gamma_{kj}^{\frac{1-\rho}{\rho}} A_k K_j \right)^{\rho} + \left((1-\beta_{kj})^{\frac{1-\rho}{\rho}} A_s S_j \right)^{\rho} \right]^{\frac{\sigma}{\rho}} \right\}^{\frac{1}{\sigma}} \tag{11}$$

where $K_j = \int_0^{\beta_{kj}} k(i)\, di$ and $S_j = \int_{\beta_{kj}}^1 l_j(i)\, di$. The parameter $1/(1-\rho)$ represents the elasticity of substitution between AI and skilled labor, while $1/(1-\sigma)$ denotes the elasticity of substitution between unskilled labor and the AI–skilled labor composite.

Our specification nests AI capital with skilled labor, following the approach of Krusell et al. (2000) and Beaudry et al. (2010), who embed capital within the skilled labor nest. This framework provides a flexible structure that allows for differential substitution between AI and heterogeneous labor. We then structurally estimate the two key elasticities, ρ and σ, in Section 3.1. If $\rho < \sigma$, AI exhibits capital-skill complementarity, consistent with technological change such as

information technology. In contrast, $\rho > \sigma$ implies that AI acts as a relative substitute for skilled labor and a relative complement to unskilled labor. Emerging evidence shows that skilled labor is more exposed to AI (Restrepo, 2024), which implies that AI could be fundamentally different from prior skill-biased technical change.

2.2 Worker

Workers of skill type $e \in \{u, s\}$ born in city i make a nested migration decision (Imbert et al., 2023). In the upper nest, they choose whether to stay in their origin city i or migrate to another city j, drawing idiosyncratic preference shocks from a Fréchet distribution with shape parameter γ_e. In the lower nest, conditional on migrating, they choose a specific destination city j, drawing from a Fréchet distribution with shape parameter η_e.

The indirect utility of a type e worker from origin city i who chooses to migrate to destination city j is given by:

$$\ln V_{ij}^e = \ln w_j^e - \alpha_h^e \ln r_j - \tau_{ij} \tag{12}$$

where w_j^e denotes the wage of type e workers in city j, α_h^e is the share of expenditures on housing for type e workers, r_j is the housing price in city j.

The indirect utility of a type e worker who stays in their origin city i is given by:

$$\ln V_{ii}^e = \ln w_i^e - \alpha_h^e \ln r_i \tag{13}$$

By the properties of the Fréchet distribution, the unconditional probability π_{ij}^e that a type e worker born in city i migrates to city j is given by:

$$\pi_{ij}^e = \left(\frac{V_{J,i}^e}{V_i^e}\right)^{\gamma_e} \cdot \left(\frac{V_{ij}^e}{V_{J,i}^e}\right)^{\eta_e} \tag{14}$$

where π_{ij}^e can be decomposed to two components. The first component $\left(\frac{V_{J,i}^e}{V_i^e}\right)^{\gamma_e}$ is the probability that a type e worker born in city i chooses to migrate rather than stay. The second component $\left(\frac{V_{ij}^e}{V_{J,i}^e}\right)^{\eta_e}$ is the probability of selecting destination j conditional on migrating from city i.

The term $V_{J,i}^e$ denotes the inclusive value for a type e worker migrating from city i, defined as the CES aggregator of indirect utilities across all potential destinations $k \neq i$:

$$V_{J,i}^e = \left[\sum_{k \in \mathcal{D}/i} (V_{ik}^e)^{\eta_e}\right]^{\frac{1}{\eta_e}} \tag{15}$$

where V_{ik}^e is the indirect utility of a type e worker from origin city i who chooses to migrate to

destination city k. η_e governs the responsiveness to destination choice conditional on migration.

The term V_i^e denotes the inclusive value for a type e worker born in city i, aggregating the indirect utility of staying (V_{ii}^e) and the inclusive value of migrating ($V_{J,i}^e$):

$$V_i^e = \left[(V_{ii}^e)^{\gamma_e} + (V_{J,i}^e)^{\gamma_e}\right]^{\frac{1}{\gamma_e}} \tag{16}$$

The parameter γ_e governs the responsiveness of the migration decision in the upper nest (stay versus migrate).

The population in each destination city j is the sum of inflows from all origin cities i, weighted by the migration shares:

$$\begin{aligned} S_j &= \sum_{i \in \mathcal{O}} S_{i,0} \pi_{ij}^s, \\ U_j &= \sum_{i \in \mathcal{O}} U_{i,0} \pi_{ij}^u. \end{aligned} \tag{17}$$

where $S_{i,0}$ and $U_{i,0}$ denote the initial populations of skilled and unskilled workers in city i, respectively. π_{ij}^s and π_{ij}^u denote the migration shares of skilled and unskilled workers between city i and city j.

2.3 Housing Market

We model housing as a local non-tradable good. Its price is determined by the intersection of local demand and an upward-sloping supply curve (Diamond, 2016).

Housing Supply We assume a perfectly competitive housing supply market, where the equilibrium price equals the marginal cost of providing housing. The inverse housing supply function is given by:

$$r_j = HS_j^{\sigma_r}, \tag{18}$$

where HS_j denotes the housing supply in city j, and σ_r is the inverse housing supply elasticity. A higher value of σ_r implies a more inelastic housing supply.

Housing Demand Housing demand is derived from the utility maximization problem of workers under Cobb-Douglas preferences. Total demand in each city is thus equal to the aggregate housing expenditure of all residents divided by the local housing price:

$$HD_j = \left(\frac{\alpha_h^s w_j^s S_j + \alpha_h^u w_j^u U_j}{r_j}\right) \tag{19}$$

Equilibrium Housing Prices Market clearing requires that housing demand equals housing supply $HD_j = HS_j$. Substituting the demand expression into the supply equation yields the equilibrium housing price:

$$r_j = \left(\alpha_h^s w_j^s S_j + \alpha_h^u w_j^u U_j\right)^{\frac{\sigma_r}{1+\sigma_r}} \tag{20}$$

Taking logarithms, the equilibrium price becomes:

$$\ln r_j = \frac{\sigma_r}{1+\sigma_r} \ln\left(\alpha_h^s w_j^s S_j + \alpha_h^u w_j^u U_j\right) \tag{21}$$

3 Structural Estimation

In this section, we structurally estimate the key parameters of the theoretical model. Specifically, we estimate the elasticity of substitution between AI and heterogeneous labor, labor supply elasticities, model-implied migration costs, and the housing supply elasticity. Additionally, we recover underlying fundamentals, such as city-specific productivity.

3.1 Labor Demand

The objective of this section is to estimate two key elasticities in the production function: the elasticity of substitution between unskilled labor and the skilled labor–AI composite, given by $1/(1-\sigma)$, and the elasticity of substitution between skilled labor and AI capital, given by $1/(1-\rho)$. To identify these parameters, we exploit exogenous variation in labor supply shocks, following the approach of Baum-Snow et al. (2018).

For tractability in the quantitative analysis, we relax the assumption on task-specific productivity and simplify the AI capital task set by assuming a uniform mapping, $\psi_k(i) = 1$. Under this assumption, the production function simplifies to:

$$Y_j = \left\{ \eta_{uj} \left(U\right)^{\sigma} + \eta_{sj} \left[\left(\left(\beta_{kj}\right)^{\frac{1-\rho}{\rho}} A_k K_j \right)^{\rho} + \left(\left(1-\beta_{kj}\right)^{\frac{1-\rho}{\rho}} A_s S_j \right)^{\rho} \right]^{\frac{\sigma}{\rho}} \right\}^{\frac{1}{\sigma}} \tag{22}$$

By solving the representative firm's first-order conditions under perfect competition in the final goods market, we derive a system of equations characterizing the relationships among the capital–output ratio, skilled wages, and the skill premium. We leverage exogenous variation in the relative supply of skilled labor ($\Delta \ln\left(S_j/U_j\right)$) to identify the labor demand elasticities by

jointly estimating the following system:

$$\begin{cases} \Delta \ln\left(\dfrac{K_j}{Y_j}\right) = \beta_1 + \dfrac{(\rho-\sigma)(1-\omega_j^k)(1-\omega_j^{ks})}{Z_j} \cdot \Delta \ln\left(\dfrac{S_j}{U_j}\right) + \Delta\epsilon_{j1} \\ \Delta \ln\left(w_s\right) = \beta_2 + \dfrac{(1-\rho)(1-\sigma)(1-\omega_j^{ks})}{Z_j} \cdot \Delta \ln\left(\dfrac{S_j}{U_j}\right) + \Delta\epsilon_{j2} \\ \Delta \ln\left(\dfrac{w_s}{w_u}\right) = \beta_3 + \left[(\sigma-1) + \dfrac{(\sigma-\rho)(1-\sigma)\omega_j^k(1-\omega_j^{ks})}{Z_j}\right] \cdot \Delta \ln\left(\dfrac{S_j}{U_j}\right) + \Delta\epsilon_{j3} \end{cases} \tag{23}$$

where ω_j^k denotes the share of AI capital within the AI–skilled nest in city j, and ω_j^{ks} represents the share of expenditure on the AI–skilled composite relative to total cost in city j.

The share of AI capital within the AI–skilled composite is given by:

$$\omega_j^k = \frac{\left(\left(\beta_{kj}\right)^{\frac{1-\rho}{\rho}} A_k K_j\right)^{\rho}}{\left(\left(\beta_{kj}\right)^{\frac{1-\rho}{\rho}} A_k K_j\right)^{\rho} + \left(\left(1-\beta_{kj}\right)^{\frac{1-\rho}{\rho}} A_s S_j\right)^{\rho}} \tag{24}$$

The share of the AI–skilled component in the overall labor aggregate is:

$$\omega_j^{ks} = \frac{\eta_{sj}\left[\left(\left(\beta_{kj}\right)^{\frac{1-\rho}{\rho}} A_k K_j\right)^{\rho} + \left(\left(1-\beta_{kj}\right)^{\frac{1-\rho}{\rho}} A_s S_j\right)^{\rho}\right]^{\frac{\sigma}{\rho}}}{\eta_{uj}\left(U\right)^{\sigma} + \eta_{sj}\left[\left(\left(\beta_{kj}\right)^{\frac{1-\rho}{\rho}} A_k K_j\right)^{\rho} + \left(\left(1-\beta_{kj}\right)^{\frac{1-\rho}{\rho}} A_s S_j\right)^{\rho}\right]^{\frac{\sigma}{\rho}}} \tag{25}$$

Estimating the elasticities of substitution between ρ and σ requires precise measurement of AI capital to capture firm responses to labor supply shocks. However, direct firm-level measures of AI investment are not available in China, and AI-related investments cannot be separately identified from other types of investment.

We overcome this limitation by constructing a proxy for AI capital (see Appendix B.3). We begin by extracting software asset data from the balance sheets of all publicly listed firms. Next, we identify AI-related public research grants using a curated set of AI-related keywords,[2] and we compute the technological proximity between each firm and universities using the cosine similarity of their respective technology portfolios. We then isolate the variation in the software assets of firms due to public AI grant shocks based on technological proximity, yielding a firm-

[2]The keyword list includes both English and Chinese terms, such as: 人工智能 (Artificial Intelligence), 机器学习 (Machine Learning), 监督学习 (Supervised Learning), 深度学习 (Deep Learning), 强化学习 (Reinforcement Learning), 神经网络 (Neural Networks), 自然语言处理 (Natural Language Processing, NLP), 计算机视觉 (Computer Vision, CV), 语音识别 (Speech Recognition), 知识图谱 (Knowledge Graphs).

level measure of AI capital. Finally, these estimates are aggregated to the city level and combined with wage bill data to construct a city-level proxy for AI capital.

Given China's rapid acceleration in AI adoption after 2015, we use data from two periods: 2012–2013 (Period 1) and 2017–2018 (Period 2) to construct AI capital K_j, skilled wages w_j^s, and unskilled wages w_j^u.[3] We measure the skill ratio ($\frac{S_j}{U_j}$) using population census data from 2010 and 2015.[4]

To isolate exogenous variation in changes to the skill ratio, we construct two separate instrumental variables: one for skilled labor and one for unskilled labor. We instrument skilled labor supply using thermal inversion shocks following Khanna et al. (2025), which serves as an exogenous driver of skilled out-migration. For unskilled labor, we adopt the agricultural shock-induced migration instrument from Imbert et al. (2022). By combining agricultural price shocks in rural origin areas with historical migration networks, we construct the instrument for unskilled labor flow to destination city j as follows:

$$\Delta \ln\widehat{(U_j)} = \ln\left(\sum_{i \in O \setminus \{j\}} \lambda_{ij}\, \Delta z_i\right) \tag{26}$$

where Δz_i is the agricultural shock in origin city i, and λ_{ij} denotes the historical stock of migrants with agricultural hukou from i residing in j prior to 2000.

Combining the two sources of exogenous variation, the first-stage equation for the change in the relative skill supply is:

$$\Delta \ln\left(\frac{S_j}{U_j}\right) = \Delta \ln\widehat{(S_j)} + \Delta \ln\widehat{(U_j)} + \xi_j \tag{27}$$

where $\widehat{S_j}$ and $\widehat{U_j}$ are the predicted skilled and unskilled labor inflows, and ξ_j is an error term. Table A2 presents the first-stage estimation results, using both instruments for the changes in the skill ratio.

[3] To recover wage levels, we combine skill premia estimated from Mincer regressions in Appendix B.4, using the CMDS dataset, with total wage bill data at the city level from the China Urban Construction Statistical Yearbook.

[4] Although the census data precede the estimation window, they represent the most recent period with comprehensive microdata, as population censuses in China are conducted every five years.

Table 1: Labor Demand Elasticity Estimates

Parameter	Coefficient					
	(1)	(2)	(3)	(4)	(5)	(6)
ρ	0.704	0.879	0.673	0.798	0.853	0.693
	(0.117)	(0.112)	(0.127)	(0.089)	(0.066)	(0.146)
σ	0.166	0.634	0.101	0.346	0.429	-0.240
	(0.234)	(0.329)	(0.226)	(0.236)	(0.227)	(0.529)
Method	FGNLS	FGNLS	FGNLS	GMM	GMM	GMM
IV1: thermal inversion	Y	Y	N	Y	Y	N
IV2: agricultural shock	Y	N	Y	Y	N	Y
Observations	122	122	122	122	122	122

Notes: Standard errors in parentheses. FGNLS is feasible generalized nonlinear least squares; GMM is generalized method of moments. IV1 is from (Khanna et al., 2025), IV2 is from agricultural shock (Imbert et al., 2022). The regressions are weighted by the 2000 population of each city.

Table 1 presents our estimates of labor demand elasticities. We find that $\hat{\rho} = 0.798$ and $\hat{\sigma} = 0.346$, and a formal test shows that $\hat{\rho}$ is significantly greater than $\hat{\sigma}$ ($p < 0.05$). This implies that AI is a relative complement to unskilled labor and a relative substitute for skilled labor.

Our estimates indicate that AI technology diverges from the predictions of standard capital–skill complementarity (Krusell et al., 2000). We show that even before the emergence of Large Language Models, AI possessed the unique potential to automate both routine and non-routine cognitive tasks traditionally performed by skilled labor.

3.2 Labor Supply

Workers are born with type $e \in \{s, u\}$ in origin city i, and make a nested location choice. In the upper nest, they choose whether to stay in city i or migrate to another city. This decision is governed by the upper nest elasticity parameters $\{\gamma_s, \gamma_u\}$. Conditional on choosing to migrate, workers then select a destination city j in the lower nest, with this choice governed by the lower nest elasticity parameters $\{\eta_s, \eta_u\}$.

Lower Nest We take the logarithm of the second component of the equation (14):

$$\ln\left(\pi^e_{ij} | \text{ migrate from } i\right) = -\eta_e \ln V^e_{J,i} + \eta_e \left[\ln w^e_j - \alpha^e_h \ln r_j - \tau^e_{ij}\right] \tag{28}$$

Where π_{ij}^e denotes the share of type e workers who migrate from origin city i to destination city j, conditional on choosing to migrate from i. The term $V_{J,i}^e = \left[\sum_{k \in \mathcal{D} \setminus i} (V_{ik}^e)^{\eta_e}\right]^{\frac{1}{\eta_e}}$ represents the inclusive value of migration for a type e worker in city i, capturing the aggregate attractiveness of all potential destination cities. Additionally, τ_{ij}^e denotes bilateral migration costs and other unobserved factors that affect migration decisions.

To estimate labor supply elasticities, we control for origin-city fixed effects and estimate a Poisson regression model in (28). This approach has the advantage of rationalizing the high frequency of zero migration flows between cities. A key challenge in identifying labor supply elasticities is the presence of unobserved factors that jointly determine wages and migration decisions, which may lead to biased estimates. To address this endogeneity, we employ two distinct instrumental variables. First, we exploit cross-city variation in AI augmentation shocks as an instrument for skilled wages. Details on the construction of this instrument are provided in Appendix B.2. Second, for unskilled labor, we exploit China's WTO accession as an exogenous shock to unskilled wages, following Facchini et al. (2019)

Table 2: Labor Supply Elasticity Estimates

	$\ln \pi_{ij}^e$			$\ln w_j^e - \alpha_h^e \ln r_j$
	Poisson		Poisson-IV	First stage-OLS
	(1)	(2)	(3)	(4)
Panel A: Skilled Labor				
$\ln w_j^s - \alpha_h^s \ln r_j$	2.157	2.353	4.515	
	(0.204)	(0.221)	(1.096)	
AI Aug$_j$				2.389
				(0.014)
Observations	11,130	11,130	11,130	11,130
Panel B: Unskilled Labor				
$\ln w_j^u - \alpha_h^u \ln r_j$	1.805	1.838	2.673	
	(0.297)	(0.317)	(0.682)	
NTR gap$_j$				1.608
				(0.016)
Observations	11,130	11,130	11,130	11,130
Controls	N	Y	Y	Y

Notes: This table reports estimates of labor supply elasticities. Columns 1–3 present Poisson and Poisson-IV estimates of the migration share equations, while Column 4 reports first-stage OLS estimates for the real wage instruments. All specifications include origin-city fixed effects. Standard errors are reported between parentheses and clustered at the level of destination cities. The regressions are weighted by the 2000 population of each city. Workers are classified as skilled if they hold a junior college degree or higher, and unskilled otherwise (high school or below). Migration shares are calculated using data from the 2015 population mini-census; wage data are drawn from the CMDS; city-level wage bill data from the China Urban Construction Statistical Yearbook (CUCSY); housing price data from Xitai (2015); and housing expenditure shares from Chen et al. (2023). The real wage of skilled labor is instrumented using the AI augmentation shock, with construction details provided in Appendix B.2. The real wage of unskilled labor is instrumented using a trade shock, computed following the methodology of Facchini et al. (2019).

Table 2 reports our estimates of the lower-nest labor supply elasticities. We find that $\eta_s > \eta_u$, indicating that skilled workers are more responsive to cross-city differences in real wages when making migration decisions.

Upper Nest Taking the logarithm of the first component of equation (14) yields:

$$\ln\left(\frac{1-\pi_{ii}^e}{\pi_{ii}^e}\right) = \gamma_e \ln\left(\frac{V_{J,i}^e}{V_{ii}^e}\right) \tag{29}$$

where π_{ii}^e denotes the probability that a worker of type e stays in their origin city i. The indirect utility of staying, V_{ii}^e, is given by $(\ln w_i - \alpha_h^e \ln r_i)$. The indirect utility of migrating, $V_{J,i}^e$, can be recovered from the origin-city fixed effects in equation (28).

To identify the upper-nest labor supply elasticities, we exploit exogenous variation in labor demand driven by industrial productivity shocks. Using annual manufacturing survey data, we construct a balanced panel of firms for the period 2011–2015 and estimate Total Factor Productivity (TFP) at the two-digit industry level following the methodology of Levinsohn and Petrin (2003). We then construct a city-level TFP instrument Z_j^{TFP}, by interacting these industry-specific productivity measures with the local industrial composition (weighted by the number of firms in each sector within the city). The intuition behind this identification strategy is that industry-specific productivity shocks are orthogonal to unobserved factors that influence local workers' out-migration decisions.

Table 3: Estimates of upper nest parameters (Selected Specifications)

Emigration	Skilled Labor		Unskilled Labor	
	(1)	(2)	(3)	(4)
$\ln\left(\frac{V_{J,i}^e}{V_{ii}^e}\right)$	0.687	1.499	0.560	1.254
	(0.154)	(0.492)	(0.119)	(0.484)
Method	OLS	2SLS	OLS	2SLS
Observations	106	106	106	106
First Stage F-stat		7.714		20.789

Notes: Standard errors are reported between parentheses and clustered at the level of origin cities. Columns (1) and (3) report OLS estimates; Columns (2) and (4) report the 2SLS estimates. The regressions are weighted by the 2000 population of each city. The Kleibergen-Paap F statistic is reported for IV estimates. The instruments are constructed by interacting firm-level total factor productivity (TFP), estimated using the Levinsohn–Petrin (LP) method, with local firm composition, proxied by the output of firms within each city cell.

Table 3 reports our estimates of the upper-nest labor supply elasticities. We find that $\gamma_s > \gamma_u$, indicating that unskilled workers are less responsive to real wage differentials across cities and therefore less likely to outmigrate compared to skilled workers.

3.3 Other Parameters

Migration Cost τ_{ij}^e Migration costs are determined to rationalize observed migration shares and real income differentials across cities. The details of this procedure are provided in the Appendix C.2.

Hukou Strigency β^e_{hukou} We parameterize the stringency of the hukou system as part of the migration cost function:

$$\tau_{ij} = \beta_0 + \beta^e_{\text{hukou}} \times \text{HukouStringency}^e_j + \epsilon_{ij} \tag{30}$$

where β^e_{hukou} denotes the elasticity of migration costs with respect to hukou stringency, and $\text{HukouStringency}^e_j$ is the measure of hukou strictness for city j. Details on the institutional context of China's hukou system and the estimation of β^e_{hukou} are provided in Appendix D.2.

Housing Supply Elasticities σ_r We estimate the housing supply elasticity to be 0.554. To address endogeneity in housing expenditure, we construct a migration-based instrument using historical migration shares and out-migration flows. Estimation details are provided in the Appendix D.1.

Agglomeration $\lambda_s, \lambda_u, \phi_s, \phi_u$ We calibrate the agglomeration force parameters using estimates from Combes et al. (2020), obtaining $\{\lambda_s, \lambda_u, \phi_s, \phi_u\} = \{0.0835, 0.0566, 0.144, 0.0728\}$.

Fundamentals ξ_{sj}, ξ_{uj}, A_j After calibrating the agglomeration parameters, we recover the fundamental productivity terms by inverting the model.

We first back out the city-specific distribution parameters using the first-order conditions of the representative firm:

$$\eta_{sj} = \frac{\left(\frac{\text{Cost on AI-Skilled Sector}}{\text{Cost on Unskilled Labor Sector}}\right)^{1-\sigma}}{\left(\frac{w^u_j}{P^{es}_j}\right)^{\sigma} + \left(\frac{\text{Cost on AI-Skilled Labor Sector}}{\text{Cost on Unskilled Labor Sector}}\right)^{1-\sigma}}$$

$$\eta_{uj} = \frac{\left(\frac{w^u_j}{P^{es}_j}\right)^{\sigma}}{\left(\frac{w^u_j}{P^{es}_j}\right)^{\sigma} + \left(\frac{\text{Cost on AI-Skilled Labor Sector}}{\text{Cost on Unskilled Labor Sector}}\right)^{1-\sigma}}$$

where P^{es}_j denotes the unit cost of the AI–skilled labor composite. Note that these expressions result from the normalization $\eta_{sj} + \eta_{uj} = 1$.

The corresponding unit cost for AI–skilled labor is given by:

$$P^{es}_j = \left(\beta_{kj}\left(\frac{1}{A_k}\right)^{\frac{\rho}{\rho-1}} + (1-\beta_{kj})\left(\frac{w^s_j}{A_s}\right)^{\frac{\rho}{\rho-1}}\right)^{\frac{\rho-1}{\rho}} \tag{31}$$

Finally, the fundamental productivity term A_j is recovered as the residual that reconciles the observed total cost with the production function.

Task share of IT-embedded AI β_{kj} We assume that AI represents an IT-embedded form of technical change and adopt a uniform automation threshold across cities. Based on external estimates of the IT task share from Eden and Gaggl (2018) and David et al. (2022), which range from 0.04 to 0.09, we calibrate the automation threshold at $\beta_{kj} = \beta = 0.05$ for all cities.

Table 4: Summary of Estimated Model Parameters

Variable	Description	Value	Source
α_h^s	Skilled housing share	0.24	(Chen et al., 2023)
α_h^u	Unskilled housing share	0.26	(Chen et al., 2023)
λ^s	Agglomeration force of skilled labor	0.0835	(Combes et al., 2020)
λ^u	Agglomeration force of unskilled labor	0.0566	(Combes et al., 2020)
ϕ^s	Agglomeration force of skilled labor	0.144	(Combes et al., 2020)
ϕ^u	Agglomeration force of unskilled labor	0.0728	(Combes et al., 2020)
β_{kj}	Task share of IT-embedded AI	0.05	(Eden and Gaggl, 2018)
A_j	Fundamental Productivity	[0.316, 2.807]	Estimation
ξ_j^s	Skilled fundamental productivity	[0.074, 0.210]	Estimation
ξ_j^u	Unskilled fundamental productivity	[0.228, 0.680]	Estimation
ρ	Elas. of sub. between K and S	0.797	Estimation
σ	Elas. of sub. between K or S and U	0.346	Estimation
η_s	Skilled lower nest parameter	4.515	Estimation
η_u	Unskilled lower nest parameter	2.673	Estimation
γ_s	Skilled upper nest parameter	1.499	Estimation
γ_u	Unskilled upper nest parameter	1.254	Estimation
τ_{ij}^s	Skilled Migration Cost	$[0.761, +\infty]$	Estimation
τ_{ij}^u	Unskilled Migration Cost	$[0.194, +\infty]$	Estimation
σ_r	Housing Supply Elasticity	0.554	Estimation

Notes: The table reports the summary of calibrated and estimated parameters.

3.4 Model Fit

Our spatial equilibrium model covers 106 cities, initialized with skill share and population data from the 2010 Census. We validate the model fit against the 2015 economy using data on GDP, AI patents, skill premiums, skill ratios, rents, and population.[5] As shown in Figure A1, although the fit is not perfectly linear, the model moments are positively correlated with the data.

[5]The skill premium is constructed using Mincerian regressions (CMDS 2015), and rent data is from Xitai.

4 Counterfactual Analysis

In this section, we conduct a series of counterfactual experiments within our spatial equilibrium framework to quantify the aggregate and distributional impacts of AI. Following the task-based framework of Acemoglu and Restrepo (2018, 2022), we decompose the AI into following technological shocks:

1. **The Extensive Margin of AI Automation:** represented as an expansion of the set of tasks performed by AI capital, denoted by an increase in the automation frontier ($d \ln \beta_{kj}$). Through this channel, AI takes over tasks previously performed by labor, effectively displacing workers to reduce production costs.

2. **The Intensive Margin of AI Automation:** represented as a uniform increase in the productivity of AI capital in tasks that are already automated ($d \ln A_k$). This indicates existing AI becomes more efficient (eg. through improved algorithms or lower compute costs) without expanding the range of automated tasks.

By simulating these technological changes, we aim to: (1) understand how AI automation affects the spatial allocation of heterogeneous labor; (2) quantify the aggregate implications of this AI-induced spatial sorting; (3) disentangle the distinct mechanisms through which AI affects spatial sorting and aggregate productivity; and (4) evaluate the optimal design of the hukou system in the era of AI.

4.1 Measuring AI Automation

The Extensive Margin of AI Automation: We measure the extensive margin of AI automation by exploiting the AI automation–induced component of changes in the AI capital share and skilled wages. The construction of exogenous AI automation exposure is detailed in Appendix B.2.

In Appendix C.1, we derive that the city-level task displacement is given by:

$$d \ln \Gamma_{kj} = d \ln \beta_{kj} = \frac{d \ln s_j^k + \left(\frac{\rho}{\rho-1}\right) \cdot s_j^l \cdot d \ln w_j^s}{1 - \left(\frac{\rho}{\rho-1}\right) \cdot s_j^k \cdot \pi_{kj}} \tag{32}$$

where $d \ln s_j^k$ denotes the change in the AI capital share induced by AI automation, while $d \ln w_j^s$ represents the change in skilled wages induced by AI automation. s_j^l and s_j^k are the initial skilled labor and AI capital shares, respectively. π_{kj} represents the average cost-saving effect, which we assume to be uniform across cities. The derivation of (32) and the definition of the average cost-saving effect π_{kj} can be found in the Appendix C.1.

Equation (32) allows us to recover the city-level task displacement effects of AI automation, $d \ln \Gamma_{kj}$, using observable data on AI-induced changes in factor shares and factor prices, initial factor shares, and average cost-saving effects. We calibrate the average cost-saving effect by combining evidence from three recent experiments (Noy and Zhang, 2023; Peng et al., 2023; Brynjolfsson et al., 2025), taking the average across these studies to be $\pi_{kj}(i) = \pi_k(i) = -36\%$. Figure A2 presents our estimation of the city-level task displacement effects of AI automation.

The Intensive Margin of AI Automation: We measure the progress in AI capital productivity between 2013 and 2018 by comparing performance on three core tasks common to AI and humans: image recognition, speech recognition, and handwriting recognition. The task performance data is sourced from Kiela et al. (2023).

Table 5: Task Performance in 2013 and 2018

Task	2018	2013
Image recognition	111.69	55.84
Speech recognition	101.66	47.3
Handwriting recognition	102.67	98.67
Mean	105.34	67.27

Notes: The table reports benchmark scores across different AI tasks in 2013 and 2018. The bottom row presents the average across tasks.

Table 5 presents the relative task performance between AI and humans. From this comparison, we recover that AI capital productivity (A_k) increased by approximately $+56.6\%$ over this period.

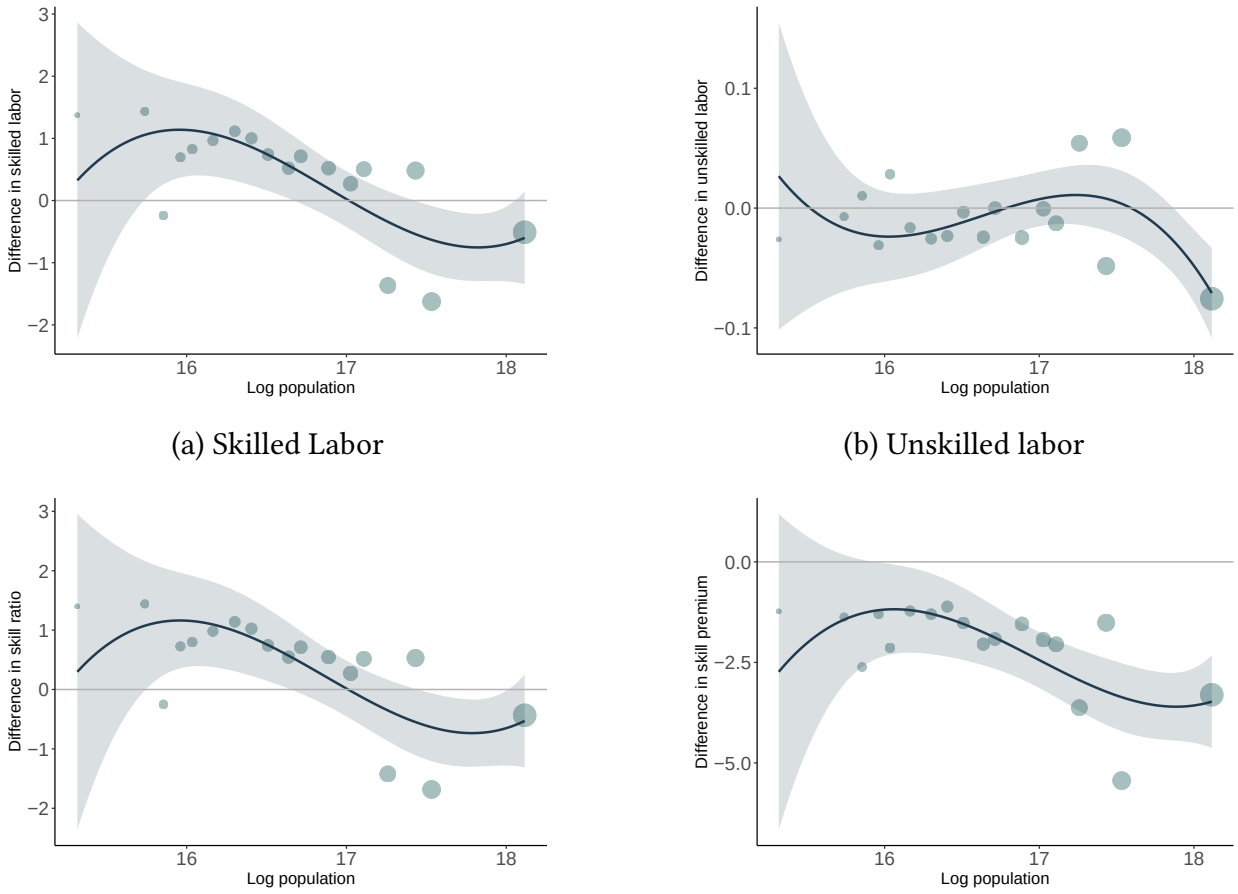
4.2 The Distributional Effects of AI

We begin by analyzing the distributional consequences of AI automation. The main objectives of this analysis are to: (1) document the patterns of AI-induced spatial sorting; and (2) quantify the unequal welfare gains resulting from this sorting. Understanding these distributional effects is crucial for identifying the mechanisms through which AI automation affects aggregate productivity.

4.2.1 Spatial Sorting

In this section, we analyze how AI automation affects the spatial allocation of labor and the skill premium across cities. Figure 1 presents the distributional effects of the extensive margin of AI

automation. We find that AI automation reverses the spatial sorting of skilled labor into large cities. AI automation also lowers the skill premium in large cities with high concentrations of cognitive labor.



(a) Skilled Labor

(b) Unskilled labor

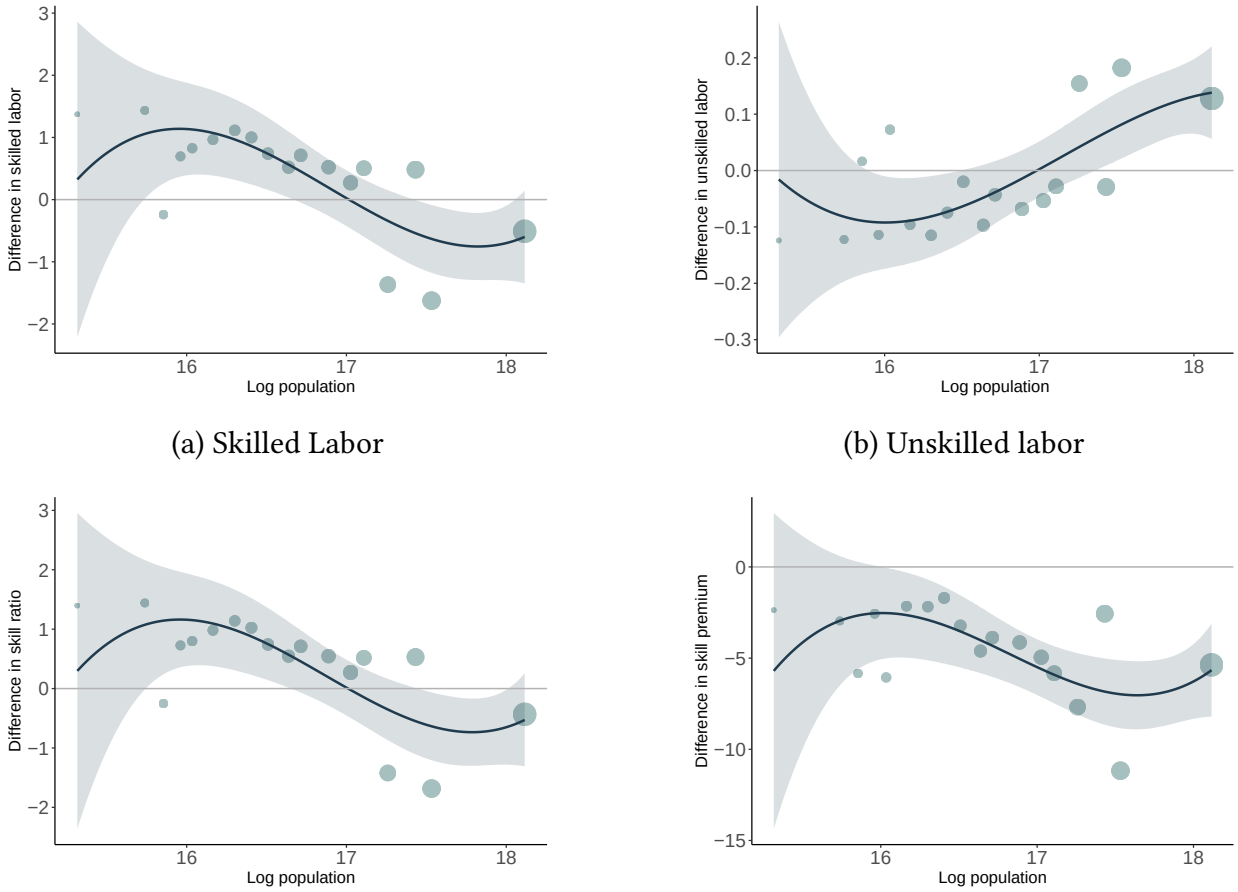
(c) Skill Ratio

(d) Skill Premium

Figure 1: The Spatial Sorting Effects of the Extensive Margin of AI Automation

Notes: The 106 cities are classified into 18 bins based on the 2000 population of each city. The top two bins contain 5 cities each, with the remaining bins containing 6 cities each. A polynomial regression is then fitted to the data.

Figure 2 presents the distributional effects of the intensive margin of AI automation, which are consistent with the task displacement effects in Figure 1 in most cases.



(a) Skilled Labor

(b) Unskilled labor

(c) Skill Ratio

(d) Skill Premium

Figure 2: The Spatial Sorting Effects of the Intensive Margin of AI Automation

Notes: The 106 cities are classified into 18 bins based on the 2000 population of each city. The top two bins contain 5 cities each, with the remaining bins containing 6 cities each. A polynomial regression is then fitted to the data.

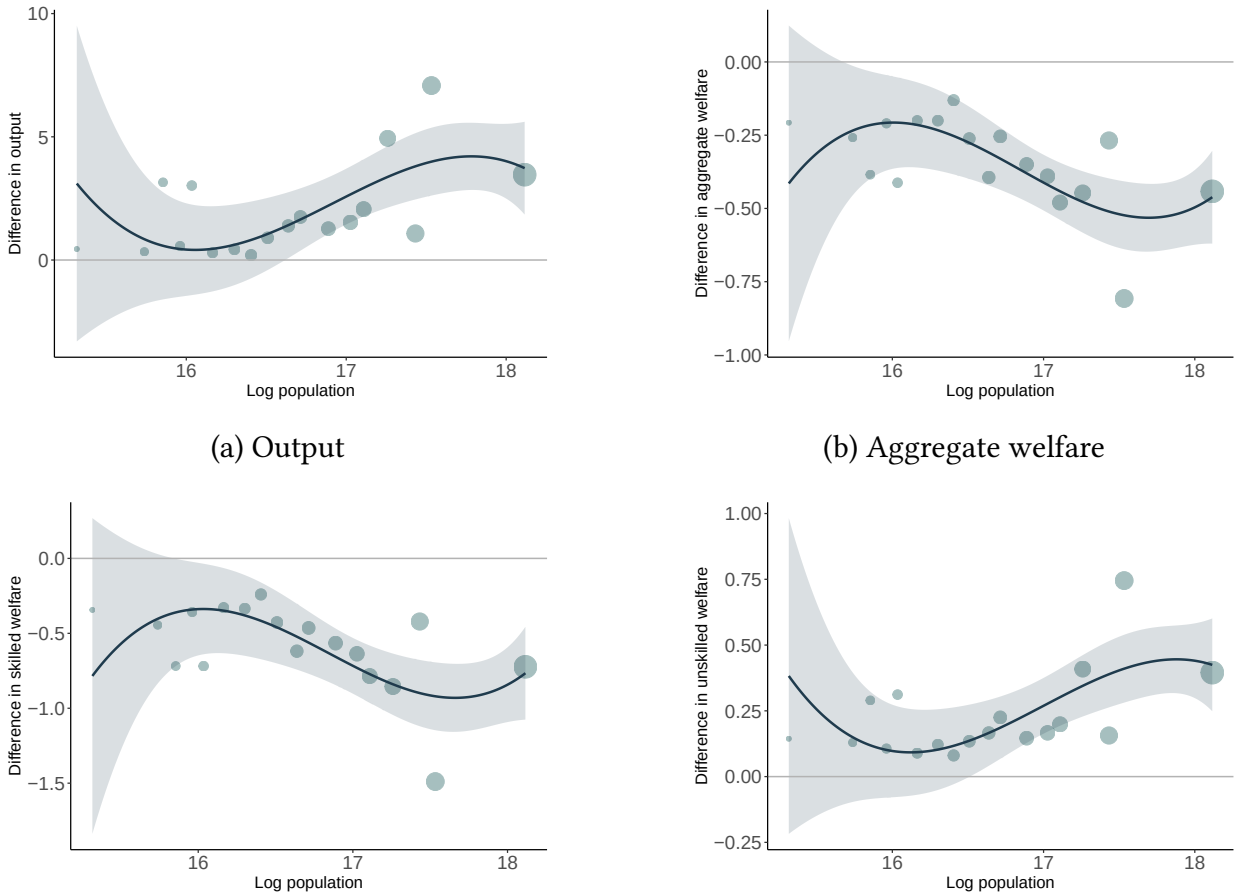
A prominent distinction between the two types of AI automation is that the extensive margin tends to drive unskilled labor away from the largest cities, whereas the intensive margin incentivizes unskilled labor to resort to them. This difference is driven by the interplay between substitution and productivity effects of AI automation. At the extensive margin, as skilled labor is displaced by AI in the largest cities and returns to medium-sized and small cities, the relative agglomeration premium declines, lowering the marginal product of unskilled labor in the largest cities. Since AI at the extensive margin fails to generate sufficient compensating productivity gains, the substitution effect dominates, causing unskilled labor to also resort to smaller cities. Conversely, at the intensive margin, the productivity effects of AI automation dominate, attracting unskilled labor to the largest cities.

One may argue that our results can be interpreted through the lens of extreme-skill complementarity (Eeckhout et al., 2014). As AI displaces routine cognitive tasks performed by middle-

skill workers, unskilled workers may increasingly concentrate in large cities. This interpretation is plausible to some extent, as the observed variation is largely driven by AI primarily replacing routine cognitive tasks during our sample period. However, our findings differ from theirs in an important way. While their framework emphasizes complementarities between extreme skills, our model captures the expansion of the AI automation frontier from routine cognitive tasks to non-routine cognitive tasks. For instance, in China, machine learning algorithms have been widely used in Fintech firms to detect fraud and assess creditworthiness. The advancement of Large Language Models (LLMs) may also fundamentally expand the automation frontier of AI to include non-routine cognitive tasks. Accordingly, our conclusions remain robust to the continued advancement of AI automation.

4.2.2 Unequal Gains from AI

We then investigate the aggregate productivity and welfare gains across cities. Figure 3 presents the aggregate gains from the extensive margin of AI automation. We find that large cities experience the greatest aggregate productivity effects, though this comes at the cost of skilled worker welfare. Combining this result with the distributional effects of AI automation, we conclude that the primary channel driving this productivity growth is the reallocation of unskilled labor towards large cities. Although the out-migration of skilled labor reduces agglomeration economies, these cities remain highly productive. Therefore, the inflow of unskilled labor contributes significantly to aggregate output. As shown in Figure A3, the aggregate gains from the intensive margin of AI automation yield similar implications.



(a) Output

(b) Aggregate welfare

(c) Skilled welfare

(d) Unskilled welfare

Figure 3: The Unequal Gains from the Extensive Margin of AI Automation

Notes: The 106 cities are classified into 18 bins based on the 2000 population of each city. The top two bins contain 5 cities each, with the remaining bins containing 6 cities each. A polynomial regression is then fitted to the data.

4.3 The Aggregate Implications of AI

Table 6 presents the aggregate implications of AI in the spatial equilibrium. We find that both the extensive and intensive margins of AI automation increase aggregate productivity, reduce skilled welfare and lower the skill premium. The aggregate effects of the intensive margin of AI automation are quantitatively two to four times larger than those of the extensive margin.

Table 6: The Aggregate Implications of AI

	Y	$\mathcal{W}$	$\mathcal{W}_s$	$\mathcal{W}_u$	w_s	w_u	$\ln\left(\frac{w_s}{w_u}\right)$
Panel A. Baseline Estimates							
AI Aut (Extensive)	1.102%	−0.438%	−3.033%	0.583%	−3.352%	0.637%	−3.964%
AI Aut (Intensive)	3.943%	−0.186%	−5.516%	1.910%	−5.954%	2.183%	−7.963%
Panel B. No Agglomeration							
AI Aut (Extensive)	−0.231%	−0.203%	−0.701%	−0.079%	−0.745%	−0.102%	−0.644%
AI Aut (Intensive)	0.014%	−0.003%	−0.045%	0.008%	−0.049%	0.008%	−0.058%
Panel C. Migration Autarky							
AI Aut (Extensive)	1.152%	−0.296%	−2.774%	0.616%	−2.935%	0.714%	−3.624%
AI Aut (Intensive)	4.037%	0.235%	−4.793%	2.086%	−4.956%	2.473%	−7.249%

To disentangle the underlying mechanisms, we conduct counterfactual simulations under two scenarios. First, we shut down endogenous agglomeration forces by setting $\{\lambda_s, \lambda_u, \phi_s, \phi_u\}$ to zero. Panel B of Table 6 shows that both types of AI automation have only negligible aggregate effects. This result indicates that knowledge spillovers play a vital role in amplifying the aggregate productivity effects of AI-induced spatial sorting.

Second, we simulate a scenario of migration autarky by setting migration costs $\tau_{ij} = \infty$ for all city pairs. Panel C of Table 6 shows that the aggregate productivity effects of AI automation are larger under migration autarky. This suggests that the AI-induced spatial allocation of labor is inefficient and fails to fully convert technological progress into productivity gains. One source of this inefficiency is the insufficient concentration of skilled labor in large cities. As skilled labor displaced by AI in urban cognitive hubs resorts to smaller, less productive cities, the resulting dispersion undermines agglomeration economies and leads to aggregate productivity losses. Therefore, restricting migration can partially mitigate this inefficiency by preventing such dispersion.

In addition, we find that the skill premium declines less under migration autarky than in the baseline model. This result contrasts with the intuition that preventing factor mobility exacerbates welfare losses from negative local shocks. The underlying reason lies in spatial ripple effects: when displaced skilled workers relocate to smaller cities, they generate a supply shock that depresses local skilled wages. Combined with the loss of agglomeration benefits in large cities, these forces cause a greater decline in the aggregate skill premium in the baseline relative to migration autarky.

4.4 The Effects of Reducing Migration Cost

Having documented the pattern of AI-induced spatial sorting, we then examine the role of migration costs and evaluate how hukou policy can be adjusted in response. Specifically, we simulate two counterfactual policy reforms: (1) a 50% reduction in hukou stringency for skilled labor in the 16 largest cities, and (2) a 50% reduction in hukou stringency for unskilled labor in the same cities. Details on the hukou stringency index and the elasticity between migration costs and hukou stringency are provided in Appendix D.2.

Following these policy changes, we recalculate migration costs using the following equation:

$$\tau_{ij}^{e'} = \tau_{ij}^{e} + \beta_{\text{hukou}}^{e} \times \Delta\text{HukouStringency}_{j}^{e} \tag{33}$$

We then re-run the counterfactual simulations to assess the aggregate implications of AI automation.

Table 7: The Aggregate Implications of Reducing Migration Cost

	Y	$\mathcal{W}$	$\mathcal{W}_s$	$\mathcal{W}_u$	w_s	w_u	$\ln\left(\frac{w_s}{w_u}\right)$
Panel A. Baseline Estimates							
AI Aut (Extensive)	1.103%	−0.438%	−3.035%	0.584%	−3.353%	0.637%	−3.965%
AI Aut (Intensive)	3.943%	−0.185%	−5.521%	1.913%	−5.956%	2.183%	−7.965%
Panel B: Relax Skilled Hukou Restrictions							
AI Aut (Extensive)	3.455%	3.091%	−2.238%	5.112%	−2.136%	5.951%	−7.632%
AI Aut (Intensive)	9.151%	3.567%	−7.045%	7.593%	−7.018%	8.891%	−14.610%
Panel C: Relax Unskilled Hukou Restrictions							
AI Aut (Extensive)	3.939%	3.184%	−2.471%	5.266%	−2.409%	6.135%	−8.050%
AI Aut (Intensive)	10.068%	3.704%	−7.537%	7.841%	−7.538%	9.196%	−15.325%

Table 7 presents the aggregate effects of reducing hukou restrictions for skilled and unskilled labor, respectively. This analysis yields two key implications. First, reducing either type of hukou restriction triples the gains from AI automation. Second, unskilled labor benefits disproportionately from these policy experiments. In particular, relaxing hukou constraints reinforces the inequality-reducing effects of AI in terms of both welfare and wages.

5 Conclusion

In this paper, we document AI-induced spatial sorting and quantify its aggregate and distributional effects in China. We propose a theoretical framework by integrating a task-based model of AI into a spatial equilibrium model. We first estimate the elasticity of substitution between AI and heterogeneous labor, finding that AI is a relative substitute for skilled labor and a relative complement for unskilled labor. Our paper provides a new perspective on spatial sorting and agglomeration. While most of the existing literature focuses on skill-biased agglomeration that drives the disproportionate concentration of skilled labor in large cities, we identify a novel channel through which AI could reverse this pattern and reduce the skill premium most significantly in large cities. We also find that AI automation increases aggregate productivity via the knowledge spillover channel, driven by the reallocation of unskilled labor to large cities.

The main mechanism we documented in this study can be applied to other form of AI, such as the large language models. Although we currently quantify the spatial equilibrium effects of AI that mainly automates routine cognitive tasks and shows only preliminary signs of non-routine displacement, our framework is robust enough to account for the expansion of the automation frontier into non-routine cognitive tasks. Thus, our core finding that AI can reverse the spatial sorting of skilled labor is likely to remain valid.

A limitation of our study is that we do not directly observe task-level substitution elasticities between AI and labor, nor do we differentiate between different occupations and AI. Future research leveraging more granular, task-level data would provide a deeper understanding of the nature of AI technologies and yield more precise estimates of the elasticity of substitution between AI and labor at the occupational or task level.

References

Acemoglu, Daron and Pascual Restrepo, "The race between man and machine: Implications of technology for growth, factor shares, and employment," *American economic review*, 2018, *108* (6), 1488–1542.

_ and _, "Tasks, automation, and the rise in US wage inequality," *Econometrica*, 2022, *90* (5), 1973–2016.

_, David Autor, Jonathon Hazell, and Pascual Restrepo, "Artificial intelligence and jobs: Evidence from online vacancies," *Journal of Labor Economics*, 2022, *40* (S1), S293–S340.

Adamopoulos, Tasso, Loren Brandt, Chaoran Chen, Diego Restuccia, and Xiaoyun Wei, "Land security and mobility frictions," *The Quarterly Journal of Economics*, 2024, *139* (3), 1941–1987.

Autor, David, Caroline Chin, Anna Salomons, and Bryan Seegmiller, "New frontiers: The origins and content of new work, 1940–2018," *The Quarterly Journal of Economics*, 2024, p. qjae008.

Autor, David H and David Dorn, "The growth of low-skill service jobs and the polarization of the US labor market," *American economic review*, 2013, *103* (5), 1553–1597.

Baum-Snow, Nathaniel, Matthew Freedman, and Ronni Pavan, "Why has urban inequality increased?," *American Economic Journal: Applied Economics*, 2018, *10* (4), 1–42.

Beaudry, Paul, Mark Doms, and Ethan Lewis, "Should the personal computer be considered a technological revolution? Evidence from US metropolitan areas," *Journal of political Economy*, 2010, *118* (5), 988–1036.

Behrens, Kristian, Gilles Duranton, and Frédéric Robert-Nicoud, "Productive cities: Sorting, selection, and agglomeration," *Journal of Political Economy*, 2014, *122* (3), 507–553.

Bloom, Nicholas, Mark Schankerman, and John Van Reenen, "Identifying technology spillovers and product market rivalry," *Econometrica*, 2013, *81* (4), 1347–1393.

Borusyak, Kirill, Peter Hull, and Xavier Jaravel, "Quasi-experimental shift-share research designs," *The Review of economic studies*, 2022, *89* (1), 181–213.

Bryan, Gharad and Melanie Morten, "The aggregate productivity effects of internal migration: Evidence from Indonesia," *Journal of Political Economy*, 2019, *127* (5), 2229–2268.

Brynjolfsson, Erik, Danielle Li, and Lindsey R Raymond, "Generative AI at work," Technical Report, National Bureau of Economic Research 2023.

_ , _ , and Lindsey Raymond, "Generative AI at work," *The Quarterly Journal of Economics*, 2025, *140* (2), 889–942.

Chen, Ziyang, Pierre-Philippe Combes, Sylvie Démurger, and Xiuyan Liu, "Housing Costs and Real Income Differences across Chinese Cities," 2023.

Combes, Pierre-Philippe, Gilles Duranton, and Laurent Gobillon, "Spatial wage disparities: Sorting matters!," *Journal of urban economics*, 2008, *63* (2), 723–742.

_ , _ , _ , Diego Puga, and Sébastien Roux, "The productivity advantages of large cities: Distinguishing agglomeration from firm selection," *Econometrica*, 2012, *80* (6), 2543–2594.

_ , Sylvie Démurger, Shi Li, and Jianguo Wang, "Unequal migration and urbanisation gains in China," *Journal of Development Economics*, 2020, *142*, 102328.

David, Tao Liang, X Wu Harry, and Kyoji Fukao, "Estimation of China's investment in ICT assets and accumulated ICT capital stock," *IDE Discussion Paper*, 2022, *833.*

Davis, Donald R and Jonathan I Dingel, "A spatial knowledge economy," *American Economic Review*, 2019, *109* (1), 153–170.

Diamond, Rebecca, "The determinants and welfare implications of US workers' diverging location choices by skill: 1980–2000," *American Economic Review*, 2016, *106* (3), 479–524.

_ and Cecile Gaubert, "Spatial sorting and inequality," *Annual Review of Economics*, 2022, *14* (1), 795–819.

Dyevre, Arnaud, "Public R&D spillovers and productivity growth," Technical Report, Working paper. Eaton, J., & Kortum, S.(1999). International technology ⋯2023.

Eden, Maya and Paul Gaggl, "On the welfare implications of automation," *Review of Economic Dynamics*, 2018, *29*, 15–43.

Eeckhout, Jan, Christoph Hedtrich, and Roberto Pinheiro, "IT and urban polarization," 2021.

_ , Roberto Pinheiro, and Kurt Schmidheiny, "Spatial sorting," *Journal of Political Economy*, 2014, *122* (3), 554–620.

Eloundou, Tyna, Sam Manning, Pamela Mishkin, and Daniel Rock, "GPTs are GPTs: Labor market impact potential of LLMs," *Science*, 2024, *384* (6702), 1306–1308.

Facchini, Giovanni, Maggie Y Liu, Anna Maria Mayda, and Minghai Zhou, "China's "Great Migration" : The impact of the reduction in trade policy uncertainty," *Journal of International Economics*, 2019, *120*, 126–144.

Fan, Jingting, "Internal geography, labor mobility, and the distributional impacts of trade," *American Economic Journal: Macroeconomics*, 2019, *11* (3), 252–288.

Gaubert, Cecile, "Firm sorting and agglomeration," *American Economic Review*, 2018, *108* (11), 3117–3153.

Giannone, Elisa, "Skilled-biased technical change and regional convergence," *University of Chicago. Unpublished manuscript*, 2017.

Imbert, Clément, Joan Monras, Marlon Seror, and Yanos Zylberberg, "Floating population: migration with (out) family and the spatial distribution of economic activity," Technical Report 2023.

_ , Marlon Seror, Yifan Zhang, and Yanos Zylberberg, "Migrants and firms: Evidence from china," *American Economic Review*, 2022, *112* (6), 1885–1914.

Jaffe, Adam B, Manuel Trajtenberg, and Rebecca Henderson, "Geographic localization of knowledge spillovers as evidenced by patent citations," *the Quarterly journal of Economics*, 1993, *108* (3), 577–598.

Khanna, Gaurav, Wenquan Liang, Ahmed Mushfiq Mobarak, and Ran Song, "The productivity consequences of pollution-induced migration in China," *American Economic Journal: Applied Economics*, 2025, *17* (2), 184–224.

Kiela, Douwe, Max Bartolo, Yixin Nie, Divyansh Kaushik, Atticus Geiger, Zhengxuan Wu, Bertie Vidgen, Grusha Prasad, Amanpreet Singh, Pratik Ringshia et al., "Dynabench: Rethinking benchmarking in NLP," *arXiv preprint arXiv:2104.14337*, 2021.

_ , Tristan Thrush, Kawin Ethayarajh, and Amanpreet Singh, "Plotting Progress in AI," *Contextual AI Blog*, 2023. https://contextual.ai/blog/plotting-progress.

Kogan, Leonid, Dimitris Papanikolaou, Lawrence DW Schmidt, and Bryan Seegmiller, *Technology-skill complementarity and labor displacement: Evidence from linking two centuries of patents with occupations*, National Bureau of Economic Research, 2021.

Krusell, Per, Lee E Ohanian, José-Víctor Ríos-Rull, and Giovanni L Violante, "Capital-skill complementarity and inequality: A macroeconomic analysis," *Econometrica*, 2000, *68* (5), 1029–1053.

Levinsohn, James and Amil Petrin, "Estimating production functions using inputs to control for unobservables," *The review of economic studies*, 2003, *70* (2), 317–341.

Morten, Melanie and Jaqueline Oliveira, "The effects of roads on trade and migration: Evidence from a planned capital city," *American Economic Journal: Applied Economics*, 2024, *16* (2), 389–421.

Munshi, Kaivan and Mark Rosenzweig, "Networks and misallocation: Insurance, migration, and the rural-urban wage gap," *American Economic Review*, 2016, *106* (01), 46–98.

Noy, Shakked and Whitney Zhang, "Experimental evidence on the productivity effects of generative artificial intelligence," *Science*, 2023, *381* (6654), 187–192.

Peng, Sida, Eirini Kalliamvakou, Peter Cihon, and Mert Demirer, "The impact of ai on developer productivity: Evidence from github copilot," *arXiv preprint arXiv:2302.06590*, 2023.

Pennington, Jeffrey, Richard Socher, and Christopher D Manning, "Glove: Global vectors for word representation," in "Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)" 2014, pp. 1532–1543.

Restrepo, Pascual, "Automation: Theory, evidence, and outlook," *Annual review of economics*, 2024, *16* (1), 1–25.

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

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

Tombe, Trevor and Xiaodong Zhu, "Trade, migration, and productivity: A quantitative analysis of China," *American Economic Review*, 2019, *109* (5), 1843–1872.

Webb, Michael, "The impact of artificial intelligence on the labor market," *Available at SSRN 3482150*, 2019.

Zhang, Jipeng, Ru Wang, and Chong Lu, "A quantitative analysis of Hukou reform in Chinese cities: 2000–2016," *Growth and Change*, 2019, *50* (1), 201–221.

Appendices

Appendix A Figures and Tables

Table A1: Summary Statistics of AI Shock

	(1)	(2)
Mean	2.92	8.72
Standard deviation	4.28	2.02
Interquartile range	6.98	3.19
Specification		
Excluding Farmer and Salesperson		Yes
Effective sample size (1/HHI of s_o weights)		
Across Occupations	2.75	56.27
Largest s_o weight		
Across Occupations	0.599	0.055
Observation counts		
No. of Occupation shocks	363	360

Notes: This table report the summary statistics of AI shock at occupation level.

Table A2: First-Stage Estimates: Labor Demand

Dependent Variable	$\Delta \ln(\frac{S_j}{U_j})$		
	(1)	(2)	(3)
ΔZ_s	-0.119	-0.171	
	(0.062)	(0.078)	
ΔZ_u	0.236		0.338
	(0.122)		(0.134)
Period FE	Yes	Yes	Yes
Observations	122	122	122
Adj. R^2	0.071	0.054	0.054

Notes: This table reports the first-stage results of the labor demand estimation. The instrument for skilled labor is constructed using changes in thermal inversion days, following Khanna et al. (2025), while the instrument for unskilled labor is based on migration flows induced by agricultural shocks, following Imbert et al. (2022). Standard errors, clustered at the destination city level, are reported in parentheses. The regressions are weighted by the 2000 population of each city.

Table A3: Migration Shares by Skill Group

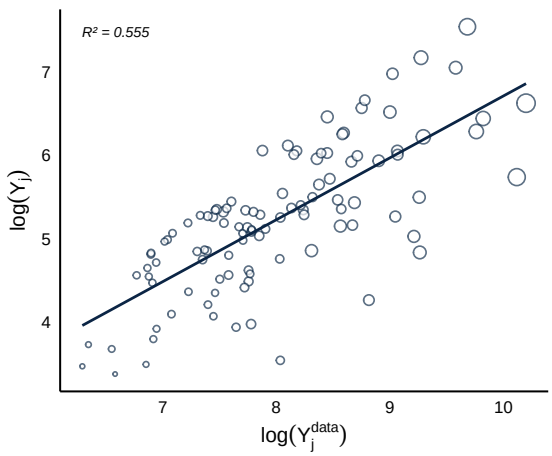
	Skilled		Unskilled	
Category	N	Share (%)	N	Share (%)
Panel A. Duration since hukou departure				
$\leq$ 1 year	2485	12.86	10461	15.62
1–5 years	7920	40.99	26121	39.01
5–10 years	4779	24.74	14003	20.91
$>$ 10 years	4136	21.41	16382	24.46
Panel B. Migration purpose				
Work/Job	12234	63.32	54594	81.52
Other	7086	36.68	12373	18.48
Panel C. Hukou origin				
Rural	4222	21.85	39096	58.38
Urban	15098	78.15	27871	41.62
Total	19320	100	66967	100

Notes: Migration types for skilled and unskilled workers are calculated using the 2015 Mini Census.

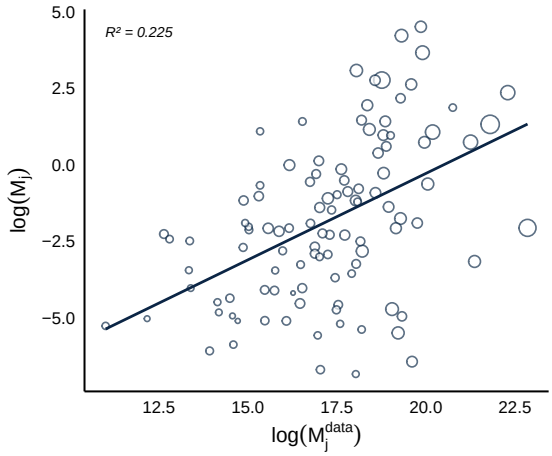
Table A4: Top 15 Industries for Migrants

Industry	Frequency	Share (%)
Wholesale and Retail Trade	15573	19.68
Accommodation and Catering	10568	13.35
Residential and Other Services	8010	10.12
Construction	6565	8.30
Miscellaneous Manufacturing	6518	8.24
Textiles and Garments	3882	4.91
Food Processing	2903	3.67
Transportation and Postal Services	2604	3.29
Computer and Electronics Mfg.	2442	3.09
Real Estate	2220	2.81
Electrical Machinery Mfg.	1767	2.23
Education	1639	2.07
Wood and Furniture	1393	1.76
IT and Software Services	1342	1.70
Agriculture, Forestry, and Fishery	1338	1.69
Total	79135	100

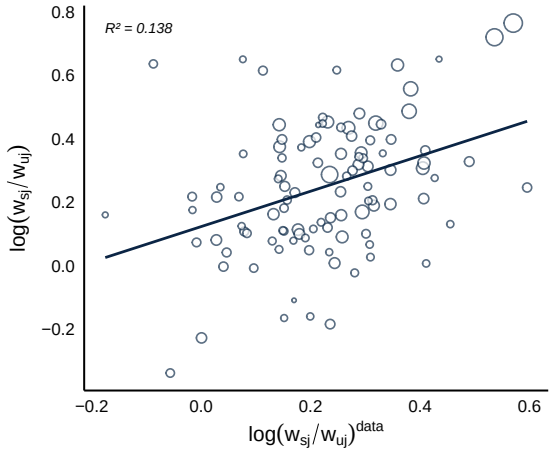
Note: This is calculated using data from the CMDS 2017. We retain all intercity migrants in the sample, restricting our analysis to those who migrated after 2010. The table presents the top 15 industries employing these migrants.



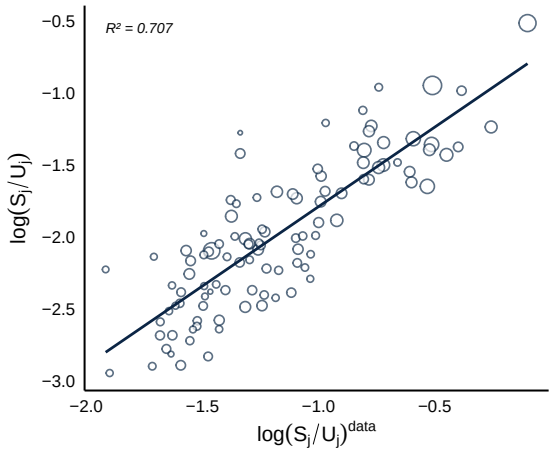
(a) Output



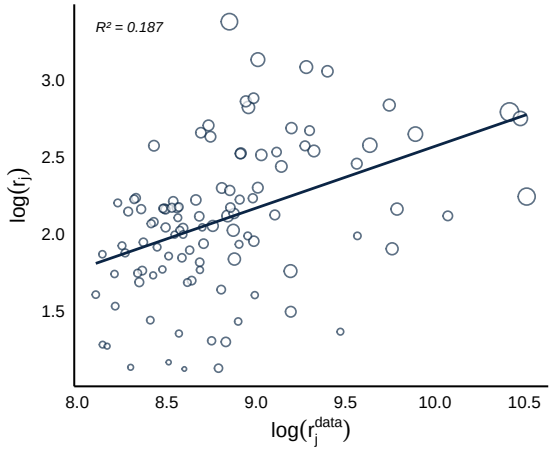
(b) AI Capital



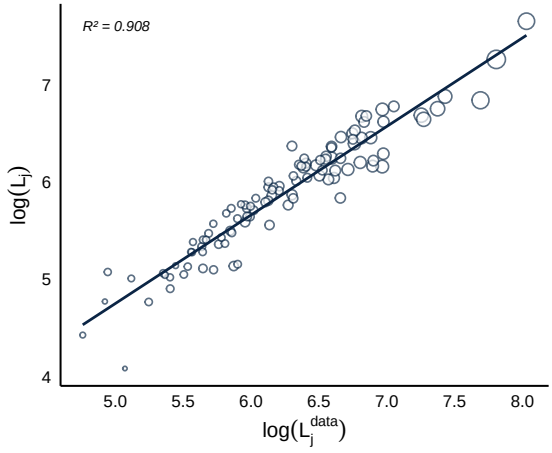
(c) Skill Premium



(d) Skill Ratio



(e) Rent



(f) Population

Figure A1: Model Fit

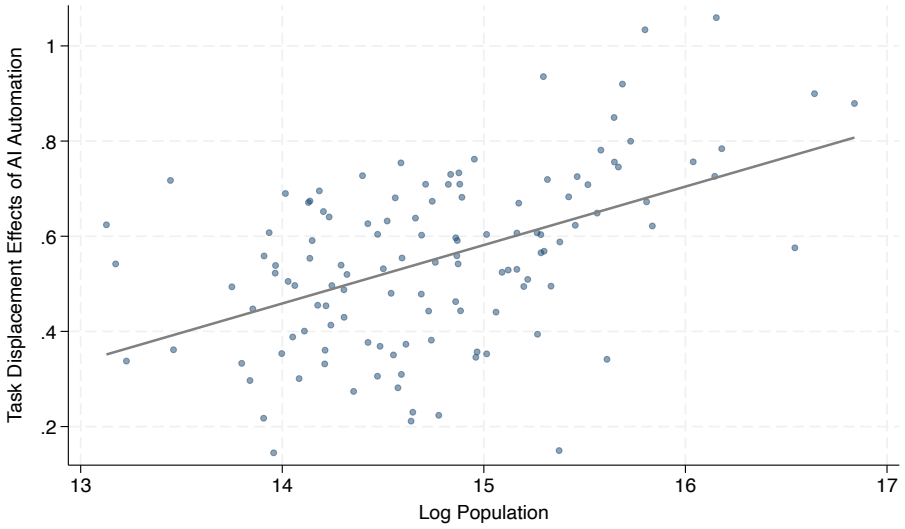
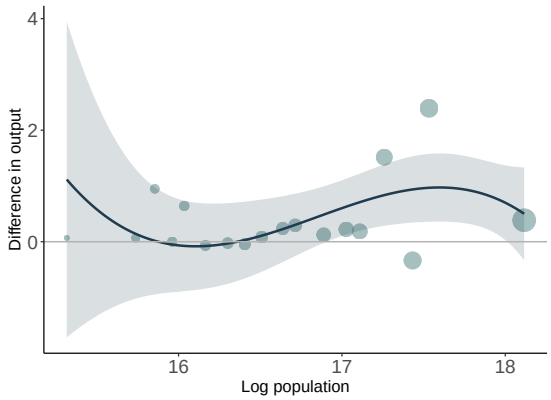
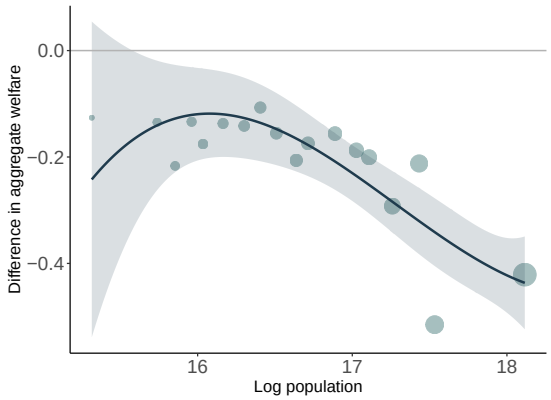


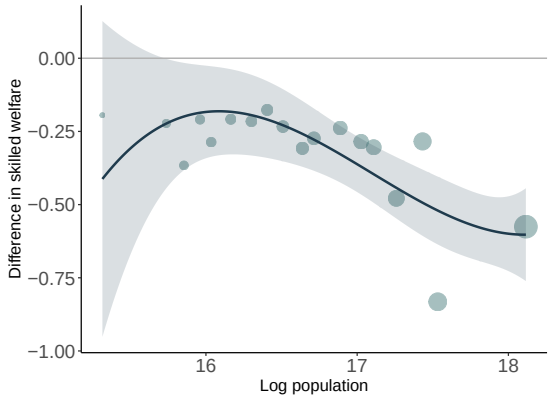
Figure A2: Task Displacement Effects of AI Automation



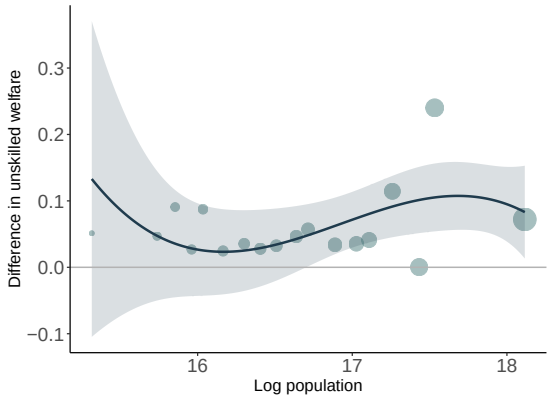
(a) Output



(b) Aggregate welfare



(c) Skilled welfare



(d) Unskilled welfare

Figure A3: The Unequal Gains from the Intensive Margin of AI Automation

Notes: The 106 cities are classified into 18 bins based on the 2000 population of each city. The top two bins contain 5 cities each, with the remaining bins containing 6 cities each. A polynomial regression is then fitted to the data.

Appendix B Data and Measurement

B.1 Data

B.1.1 Census Data

We use data from the 2000, 2010, and 2015 population censuses in China. The 2000 Census data is sourced from IPUMS.[6] The 2010 Census and the 2015 "mini" Census are obtained from the National Bureau of Statistics (NBS).

B.1.2 Firm Data

Data on the software assets of publicly listed firms are sourced from CSMAR.[7] The Annual Surveys of Manufacturing Firms (2011–2015) are sourced from EPS Data.

B.1.3 Population and Wage Data

Data on city-level population and wages are obtained from the China Urban Construction Statistical Yearbooks (CUCSY, 2010–2020), maintained by the Ministry of Housing and Urban-Rural Development.

B.1.4 China Migrants Dynamic Survey (CMDS) Data

Data on the city-level skill premium are sourced from the China Migrants Dynamic Survey (CMDS) for the years 2012, 2013, 2015, 2017, and 2018.[8]

B.1.5 Public Research Grants Data

We obtain public research grant data from the National Natural Science Foundation of China (NSFC), including details on grant timing, funding amounts, university affiliations, abstracts, and publications.

B.1.6 Occupation Data

We measure AI exposure by calculating the textual similarity between AI-related patents and occupational task descriptions. To address the discrepancy between the O*NET database and the Chinese Standard Classification of Occupations (CSCO), we manually map 363 3-digit CSCO

[6] This sample represents a 1% extract of the national population, comprising 11,804,344 individuals. We use the demographic information from this sample to calculate the occupational shares within each city.

[7] CSMAR is a research database that provides comprehensive data on publicly listed firms in China.

[8] We employ Mincer regressions to estimate skill premiums for each city.

occupations to one or more O∗NET counterparts. This crosswalk enables us to assign detailed task descriptions to each occupation in China. For the textual analysis, we use two distinct data sources: occupational task-based descriptions from the Dictionary of Occupational Titles (DOT) and output-related descriptions from the Census Alphabetical Index of Occupations and Industries (CAI).

B.1.7 AI Patent Data

Data on US patents are from the Artificial Intelligence Patent Dataset (AIPD) from Identifying Artificial Intelligence (AI) Invention. This dataset was constructed using machine learning models to classify all patents published by the USPTO into eight component technologies, including Knowledge Processing, Speech, AI Hardware, Evolutionary Computation, Natural Language Processing, Machine Learning, Computer Vision, and Planning/Control. We also construct AI exposure using the stock of Chinese AI-related patents sourced from Patsnap.[9] We select the following keywords and classify patents with these keywords occurring in the title, abstract, claims, or description as AI patents: Machine Learning, Supervised Learning, Deep Learning, Reinforcement Learning, Neural Networks, Natural Language Processing, Computer Vision, Speech Recognition, and Knowledge Graphs. Since the text used in task descriptions for occupations is in English, we use the English translations of Chinese patent text from Patsnap.

[9]Patsnap is a large data provide of R&D and patent relevant information.

B.2 Measurement of AI Exposure

To measure AI exposure at the occupation level, we adopt the framework developed by Webb (2019), Kogan et al. (2021), and Autor et al. (2024). The main intuition is to identify technology-driven labor market changes by computing the textual similarity between patent documents and occupational descriptions.

Specifically, we combine US and Chinese AI patent texts with occupational data, using task (input) descriptions from the Dictionary of Occupational Titles (DOT) and output descriptions from the Census Alphabetical Index (CAI). We measure AI automation using the overlap between AI patents and occupational inputs, and augmentation using the overlap with outputs. However, the output measure is confounded, as it may reflect either augmentation or the direct automation of final products. Therefore, to ensure precise identification, we focus primarily on the input-based measure.

We construct numerical representations for both the textual content of each AI patent and the occupational descriptions derived from the DOT (task-based) and CAI (output-based). To quantify textual similarity, we represent each document as a weighted average of word embeddings. This approach allows us to compute cosine similarity scores between AI patents and occupation-specific texts, capturing the degree of semantic overlap.

B.2.1 Linking CSCO to SOC

The Chinese Standard Classification of Occupations (CSCO) does not provide the granular task descriptions required for our study. Therefore, we map Chinese occupations to the US Standard Occupational Classification (SOC) system to exploit the comprehensive task content data in O*NET.

We construct this crosswalk in three steps. First, we use the mapping provided by the China Family Panel Studies (CFPS) to link CSCO codes to the International Standard Classification of Occupations (ISCO-88). Second, we convert ISCO-88 codes to the ISCO-08 standard. Finally, we apply existing crosswalks to map ISCO-08 to US SOC-2010 codes, allowing us to assign detailed O*NET task descriptions to each Chinese occupation.

B.2.2 Construction of AI Exposure

The step of processing each text document is as follows:

(1) Exclude common "stop words" that are not informative, such as "is", "the", "a", "and", "in", "to", and similar terms, from the list of stop words provided by spaCy.

(2) Perform part-of-speech tagging using spaCy to classify each word as a noun or verb. Keep only nouns and verbs in the text, and lemmatize the words by transforming verbs to their base

form and nouns to their singular form.

(3) Extract word embeddings for each term in the cleaned document using the pre-trained word embeddings GloVe from Pennington et al. (2014). Compute their average weighted by TF-IDF scores to obtain a vector representing the document's semantic content.[10]

(4) Calculate the cosine similarity score for all patent-occupation pairs within a $\mathrm{T}(t-t_0)$ period, based on the results of TF-IDF-weighted document vectors. The cosine similarity $S^i_{p,o}$ between patent p and occupation o is as follows.

$$S_{p,o} = \frac{V_{\text{AIPatent}} \cdot V_{\text{DOT}}}{\|V_{\text{AIPatent}}\| \cdot \|V_{\text{DOT}}\|}$$

(5) Normalize the scores by subtracting the median score for each occupation for a given patent. Set different thresholds of similarity score σ_t at the 10%, 15%, and 20% percentiles, and select the patent-occupation pairs with adjusted textual similarity scores larger than σ_t (Autor et al., 2024).

$$\mathrm{N}_{p,o} = \begin{cases} 1 & \text{if} \quad S_{p,o} \geq \sigma_t \\ 0 & \text{otherwise} \end{cases}$$

Finally, we aggregate patent-occupation pairs to construct an occupation-level measure of AI exposure. To account for the fact that the development of industrial AI often builds on existing technologies, we use a citation-weighted patent count, averaged across patent cohort-years based on the year of issuance.

We construct the long-difference AI exposure between baseline year t_0 and year t, capturing AI shocks over the period $\mathrm{T}(t-t_0)$.

$$\Delta\text{PatentStock}^{\text{US}}_{o,t-t_0} = \sum_{p \in \mathrm{T}(t-t_0)} w_p \times \mathrm{N}^{\text{US}}_{p,o} \quad \text{with} \quad w_p = \frac{\text{NCited}_p}{\text{AvgNcited}_{y(p)}}$$

B.2.3 Validity of AI Exposure

To assess the validity of the AI exposure measure constructed in this section, we follow the approach of Borusyak et al. (2022) and provide summary statistics of AI shocks at the occupation level. Table A1 reports the corresponding shock-level statistics. In Column (1), we find that the effective sample size is only 2.75 across 363 three-digit occupations from the Chinese Standard Classification of Occupations (CSCO). This insufficient effective sample size is because the num-

[10] We calculate the document vectors for patent and occupation texts separately, as there are significant differences in word frequency between the two datasets. This approach allows for more accurate TF-IDF weighted document vectors.

ber of farmers is so high in the Chinese census, accounting for approximately 60% of the sample.

To mitigate concerns about identification, we exclude the three occupations with the largest employment shares : two types of farmers and salespersons. Column (2) presents the summary statistics after these exclusions. The effective sample size increases to 56.27, and the maximum occupation weight falls to 0.055, alleviating concerns about concentration bias.

B.2.4 City AI Exposure

To measure AI exposure at the city level, we utilize occupational composition data from the 2000 China Population Census. We construct a shift-share instrument by interacting initial local occupational shares with a plausibly exogenous shock derived from U.S. AI patenting.

The instrument is defined as:

$$\Delta AI_j^{\text{IV}} = \left(\frac{S_{jo,2000}}{S_{j,2000}} \right) \times \Delta \text{PatentStock}_{o,t-t_0}^{\text{US}}$$

where $\Delta \ln \text{PatentStock}_{o,2020\text{–}2016}^{\text{US}}$ represents the change in U.S. AI patent stock by occupation o over the same period, $S_{jo,2000}$ denotes the skilled employment in occupation o in city j in the year 2000, and $S_{j,2000}$ is the total skilled employment in city j in the same year.

B.3 Measurement of AI Capital

In this section, we introduce our measure of AI capital. We conceptualize AI as IT capital-embodied technical change. This assumption is well grounded in the context of pre–Large Language Model (LLM) AI technologies, which primarily include machine learning, natural language processing, and other software-based applications.

The main idea of our measurement strategy is to isolate the component of firms' software assets that is driven by AI-related shocks. To do so, we combine data on software assets of publicly listed firms with AI-relevant public R&D grants from China's National Natural Science Foundation, which serve as a credible proxy for AI-specific investment. These grant records contain rich information, including project titles, abstracts, and funding amounts, which allow us to clearly identify and classify AI-specific R&D investments. Using this information, we construct a firm-level proxy for AI capital, which is then aggregated to the city level for use in the structural estimation.

We assume that a firm's software asset is composed of AI-related and other software, and model it as follows:

$$\text{Software}_f = \left(\text{AI Software}_{ft}\right)^{\alpha}$$

where Software_{ft} denotes the total software assets of firm f in year t, AI Software_{ft} captures the AI-relevant software component, and α represents the share of AI-related software in the overall software capital.

The production of AI software at the firm level is modeled as a Cobb–Douglas function of public AI grants and private R&D investment:

$$\text{AI Software}_{ft} = \left[\prod_{u \in \mathcal{U}} \left(G_{ut}\right)^{s_{fu}}\right]^{\gamma} I_{ft}$$

where AI Software_{ft} denotes the AI-specific software of firm f in year t, I_{ft} represents the firm's private R&D investment on AI, and G_{ut} is the amount of public AI-related grant funding received by university u in year t. The term s_{fu} captures the technological distance between firm f and university u, and γ denotes the elasticity of AI capital production with respect to public R&D spillovers.

To create a pre-determined link between universities and firms, we construct a measure of technological distance as follows:

$$s_{fu} = \frac{V_f \cdot V_u}{\|V_f\| \cdot \|V_u\|} \cdot \frac{1}{d_{l(f)l(u)}}$$

where V_f and V_u denote the technological stock vectors of public firm f and university u, respec-

tively. Following the approach of Bloom et al. (2013) and Dyevre (2023), we assume that entities with greater technological similarity are more likely to benefit from knowledge spillovers.[11] The technological vector for each firm is constructed using its patent portfolio prior to 2010, while the university vector is based on historical public research funding received before 2010. We map Chinese academic subject classifications to IPC classes using standardized definitions to ensure consistency in the technological space. In addition, $d_{l(f)l(u)}$ denotes the geographic distance between the cities in which firm f and university u are located. This term captures the localized nature of knowledge spillovers (Jaffe et al., 1993).

We estimate the elasticity of firm software assets with respect to public AI grants using the following specification:

$$\ln \text{Software}_{ft} = \alpha \cdot \gamma \sum_{u \in \mathcal{U}} s_{fu} \ln G_{u,t-1} + \eta_{dj} + \delta_t + \epsilon_{ft}, \tag{A1}$$

where $\ln \text{Software}_{ft}$ denotes the log of software assets for firm f in year t, while $\sum_{u \in \mathcal{U}} s_{fu} \ln G_{u,t-1}$ captures the firm's exposure to lagged AI-related public grants, weighted by the technological and geographic proximity s_{fu} to universities. We include sector-city fixed effects η_{dj} and year fixed effects δ_t. Finally, ϵ_{ft} denotes the idiosyncratic error term.

We then extract the predicted AI software capital as follows:

$$\widehat{\text{AI Software}}_{fjt} = \underbrace{\exp\left(\ln \widehat{\text{Software}}_{fjt}\right)}_{\text{Predicted Total Software}} - \underbrace{\exp\left(\ln \widehat{\text{Software}}_{fjt} - \alpha \cdot \gamma \sum_{u \in \mathcal{U}} s_{fu} \ln G_{u,t-1}\right)}_{\text{Counterfactual Software}} \tag{A2}$$

Next, we aggregate AI capital to the city level in two steps. First, we define two periods: Period 1 covers the years $\{2012, 2013\}$ and Period 2 covers $\{2017, 2018\}$. Using our public firm–level sample, we compute the AI-to-labor ratio (R_{jt}) for each city j in period t by aggregating the predicted AI capital and firm-level wage bills:

$$R_{jt} = \frac{\sum_{f \in \mathcal{F}_j} \widehat{\text{AI Software}}_{ft}}{\sum_{f \in \mathcal{F}_j} \text{Wage Bill}_{ft}},$$

where $\mathcal{F}_j$ denotes the set of firms in city j observed in our sample.

Finally, we scale the estimated AI-to-labor ratio by each city's total wage bill to recover the city-level AI capital:

$$K_{jt} = R_{jt} \times \text{Total Wage Bill}_{jt}.$$

We acknowledge that our measure of AI capital is an imperfect proxy for actual AI investment

[11] We assume that these spillovers do not generate business-stealing effects across regions.

or adoption. It likely underestimates the true level of AI capital because it does not capture important components such as hardware investment, unobserved private investment spillovers, and government-led industrial policies, all of which may significantly influence AI adoption.

Nonetheless, our estimates remain informative. We find that approximately 1%–2% of the variation in public firms' software assets can be explained by exposure to public AI funding shocks. More importantly, we argue that public grants are positively correlated with private investment in AI, $\text{Cov}\left(\left[\prod_{u\in\mathcal{U}}\left(G_{ut}\right)^{s_{fu}}\right]^{\gamma}, I_{ft}\right) > 0$). Therefore, our measure also captures part of the information on endogenous private investment response to public AI funding.

B.4 Other Measurement

Wage Estimation (Mincer Regression) To obtain skill premium at city level, we run the following Mincer regression:

$$\ln w_{ij}^{e} = \beta_0 + \delta_j^{e} + \beta_1 \,\text{Age}_i + \beta_2 \,\text{Age}_i^2 + \mathbf{X}_i\gamma' + \theta_d + \varepsilon_i^{e}. \tag{A3}$$

where $\ln w_{ij}^{e}$ denotes the log wage of individual in city j, while Age_i represents the age of individual i. δ_j^{e} denotes the city j skill type e fixed effects (with unskilled labor in Shanghai as the omitted group), and θ_d represents industry fixed effects. The vector of controls $\mathbf{X}_i$ includes hukou dummies (urban, rural), work type dummies (employer, employee, self-employed), and gender indicators. The estimated average wage fixed effect for labor of type e in city j is given by $\widehat{\beta_0} + \widehat{\delta_j^{e}}$. We then could recover the skill premium at city level.

Appendix C Model Derivation

C.1 Task Displacement Effects of AI

In this section, we derive the task displacement effects, expressing them as a function of the AI automation-induced component of changes in AI capital share and skilled wages.

We begin with the AI capital share in the AI-skilled labor composite. In our nested CES framework, the capital share s_j^k is determined by the relative expenditure on automated tasks versus non-automated skilled labor:

$$s_j^k = \frac{\Gamma_{kj}(i)}{\left(\Gamma_{kj}(i)\left(\frac{1}{A_k}\right)^{\frac{\rho}{\rho-1}} + \left(1-\beta_{kj}\right)\left(\frac{w_j^s}{A_s}\right)^{\frac{\rho}{\rho-1}}\right)} = \frac{\Gamma_{kj}(i)}{I_j(i)} \tag{A4}$$

where $I_j(i)$ represents the price index (unit cost) of the AI-skilled sector. This cost index is derived from the CES aggregation of automated and non-automated tasks:

$$C_j(i) = \left(\Gamma_{kj}(i)\left(\frac{1}{A_k}\right)^{\frac{\rho}{\rho-1}} + \left(1-\beta_{kj}\right)\left(\frac{w_j^s}{A_s}\right)^{\frac{\rho}{\rho-1}}\right)^{\frac{\rho-1}{\rho}} = I_j(i)^{\frac{\rho-1}{\rho}} \tag{A5}$$

We now decompose the total change in the price index of the AI–skilled sector. This change reflects both the direct effect of an expansion in the automation frontier and the indirect effect of rising skilled wages, which affect the sector' s unit cost.

Applying the Envelope Theorem, we isolate the contribution of skilled wage changes to the cost index:

$$\begin{aligned}\frac{\partial \ln I_j(i)}{\partial \ln w_j^s} d\ln w_j^s &= \left(\frac{\rho}{\rho-1}\right)\frac{\left(1-\beta_{kj}\right)\left(\frac{w_j^s}{A_s}\right)^{\frac{\rho}{\rho-1}}}{I_j(i)} \cdot d\ln w_j^s \\ &= \left(\frac{\rho}{\rho-1}\right)\cdot s_j^l \cdot d\ln w_j^s\end{aligned}$$

We then decompose the total change in the cost index $I_j(i)$ into two components: one driven by the expansion of the automation frontier (Γ_{kj}) and the other by task displacement due to changes in skilled labor costs. The derivation proceeds as follows:

$$d\ln I_j(i) = \frac{\Gamma_{kj}(i)}{I_j(i)} \cdot d\ln \Gamma_{kj} + \frac{\left(1-\beta_{kj}\right)\left(\frac{w_j^s}{A_s}\right)^{\frac{\rho}{\rho-1}}}{I_j(i)} \cdot d\ln\left(1-\beta_{sj}\right)$$

$$
\begin{aligned}
&= \frac{\Gamma_{kj}(i)}{I_j(i)} \cdot \frac{\int_{\mathcal{A}_j} \psi_k(i)^{\frac{\rho}{1-\rho}} di}{\Gamma_{kj}(i)} - \frac{\left(\frac{w_j^s}{A_s}\right)^{\frac{\rho}{\rho-1}}}{I_j(i)} \cdot \mathcal{A}_j \\
&= \frac{1}{I_j(i)} \cdot \int_{\mathcal{A}_j} \left[\psi_k(i)^{\frac{\rho}{1-\rho}} - \left(\frac{A_k w_j^s}{A_s}\right)^{\frac{\rho}{\rho-1}} \right] di \\
&= \frac{\Gamma_{kj}}{I_j(i)} \cdot \frac{1}{\Gamma_{kj}} \cdot \int_{\mathcal{A}_j} \psi_k(i)^{\frac{\rho}{1-\rho}} \pi_{kj}(i) di \\
&= s_j^k \cdot \frac{\int_{\mathcal{A}_j} \psi_k(i)^{\frac{\rho}{1-\rho}} di}{\int_{\Gamma_j} \psi_k(i)^{\frac{\rho}{1-\rho}} di} \cdot \frac{\int_{\mathcal{A}_j} \psi_k(i)^{\frac{\rho}{1-\rho}} \pi_{kj}(i) di}{\int_{\mathcal{A}_j} \psi_k(i)^{\frac{\rho}{1-\rho}} di} \\
&= \left(\frac{\rho}{\rho-1}\right) \cdot s_j^k \cdot d\ln \Gamma_{kj} \cdot \pi_{kj}
\end{aligned}
$$

We define the cost saving effects π_{kj} as:

$$
\pi_{kj}(i) = \frac{\rho-1}{\rho} \left[1 - \left(\frac{A_k w_j^s \psi_k(i)}{A_s} \right)^{\frac{\rho}{\rho-1}} \right] \tag{A6}
$$

We can now express the evolution of the capital share as a function of the automation shock, the cost savings, and the change in wages. The task displacement effect, holding wages constant, is given by:

$$
\begin{aligned}
d\ln s_j^k &= d\ln \Gamma_{kj} - d\ln I_j(i) \\
&= d\ln \Gamma_{kj} - \left(\frac{\rho}{\rho-1}\right) \cdot s_j^k \cdot \pi_{kj} \cdot d\ln \Gamma_{kj} - \left(\frac{\rho}{\rho-1}\right) \cdot s_j^l \cdot d\ln w_j^s
\end{aligned}
$$

Rearranging this relationship yields the final expression in the main text.

$$
d\ln \Gamma_{kj} = d\ln \beta_{kj} = \frac{d\ln s_j^k + \left(\frac{\rho}{\rho-1}\right) \cdot s_j^l \cdot d\ln w_j^s}{1 - \left(\frac{\rho}{\rho-1}\right) \cdot s_j^k \cdot \pi_{kj}} \tag{A7}
$$

C.2 Migration cost

To back out the model-implied migration cost, we first compute the relative migration cost between cities, conditional on selecting a reference destination city k. The ratio of migration probabilities is given by:

$$
\frac{\pi_{ij}^e}{\pi_{ik}^e} = \left[\frac{\exp(-\tau_{ij}^e) V_{ij}^e}{\exp(-\tau_{ik}^e) V_{ik}^e} \right]^{\eta_e} \tag{A8}
$$

Rearranging the expression, we obtain the relative migration cost:

$$\frac{\exp(-\tau_{ij}^e)}{\exp(-\tau_{ik}^e)} = \left(\frac{\pi_{ij}^e}{\pi_{ik}^e}\right)^{\frac{1}{\eta_e}} \left(\frac{V_{ik}^e}{V_{ij}^e}\right) \tag{A9}$$

We then substitute this expression for relative migration cost into the upper nest equation:

$$\begin{aligned}
\ln\left(\frac{1-\pi_{ii}^e}{\pi_{ii}^e}\right) &= \gamma_e \ln\left(\frac{V_{J,i}^e}{V_{ii}^e}\right) \\
&= \frac{\gamma_e}{\eta_e} \ln\left[\left(\exp\left(-\tau_{ik}^e\right) V_{ik}^e\right)^{\eta_e} + \sum_{j \in J \setminus \{k\}} \left(\left(\frac{\pi_{ij}^e}{\pi_{ik}^e}\right)^{\frac{1}{\eta_e}} \left(\frac{V_{ik}^e}{V_{ij}^e}\right) V_{ij}^e\right)^{\eta_e}\right] - \gamma_e \ln V_{ii}^e
\end{aligned}$$

Finally, the migration cost between city i and city k for a worker of type e can be expressed as:

$$\begin{aligned}
\tau_{ij}^e &= \ln V_{ij}^e - \ln V_{ii}^e + \frac{1}{\eta_e} \ln\left(\frac{1-\pi_{ii}^e}{\pi_{ij}^e}\right) - \frac{1}{\gamma_e} \ln\left(\frac{1-\pi_{ii}^e}{\pi_{ii}^e}\right) \\
&= \ln\left(\frac{w_j^e}{w_i^e}\right) - \alpha_h^e \ln\left(\frac{r_j}{r_i}\right) + \frac{1}{\eta_e} \ln\left(\frac{1-\pi_{ii}^e}{\pi_{ij}^e}\right) - \frac{1}{\gamma_e} \ln\left(\frac{1-\pi_{ii}^e}{\pi_{ii}^e}\right)
\end{aligned} \tag{A10}$$

Appendix D Other Estimation

D.1 Housing Supply Elasticities

In this section, we estimate the housing supply elasticity. The equilibrium housing price is given by:

$$\ln r_j = \frac{\sigma_r}{1+\sigma_r} \ln\left(\alpha_h^s w_j^s L_j^s + \alpha_h^u w_j^u L_j^u\right) \tag{A11}$$

To isolate variation affecting housing prices, we use migration data from the 2015 mini-census to extract total out-migration from city i over the period 2005–2015.

Based on this, we construct a population-migration instrument for total housing expenditure:

$$Z_j^M = \sum_{i \in \mathcal{O}} s_{ij} \times \ln M_{i,2015-2005}^{-j} \tag{A12}$$

where $\ln M_{i,2015-2005}^{-j}$ denotes the log number of out-migrants from origin i between 2005 and 2015, and s_{ij} represents the migration share from i to j before 2000. When constructing the out-migration measure, we employ a leave-one-out strategy to exclude flows to city j to avoid mechanical correlation with the dependent variable.

Table A5: Estimation of Housing Supply Elasticity

Dependent Variable	$\ln r_j$	
	(1)	(2)
Log Housing Expenditure	0.444	0.554
	(0.040)	(0.129)
Estimation	OLS	2SLS
Observations	163	163

Notes: The table reports estimates of the housing supply elasticity. Standard errors clustered at the city level are shown in parentheses. The regressions are weighted by the 2000 population of each city.

D.2 Migration Cost and Hukou System

In this section, we estimate migration costs in the spatial equilibrium model and quantify their elasticities with respect to hukou stringency for different types of workers.

In Appendix C.2, We show that migration costs can be expressed as a function of observables and estimated parameters. After recovering labor supply elasticities, we recover model-implied

migration costs τ_{ij} under the assumption that stayers face zero migration costs ($\tau_{ii} = 0$). Compared with the approaches in Fan (2019) and Morten and Oliveira (2024), our method does not require imposing additional functional form assumptions on migration costs (Imbert et al., 2023).

After recovering the migration cost, we focus on the hukou system as the most significant component of destination-specific migration costs in China. We estimate the elasticity of migration costs with respect to hukou stringency. The following paragraph introduces the institutional context of the hukou system in China and presents the estimation procedure.

Hukou System The hukou system in China restricts access to local public goods for non-local migrants. Every individual is assigned a place-based hukou at birth, determined by the hukou status of their parents in the locality of residence. The primary purpose of this population management policy is to regulate urban population size by limiting internal migration to large cities, thereby mitigating excessive congestion and pressure on local resources. Therefore, hukou restrictions are typically more stringent in larger cities.

The elasticity between migration cost and hukou stringency is estimated using the following specification:

$$\tau_{ij} = \beta_0 + \beta^e_{\text{hukou}} \times \text{HukouStringency}^e_j + \epsilon_{ij} \tag{A13}$$

where $\text{HukouStringency}^e_j$ denotes the hukou stringency index for workers of type e in destination city j. This index is sourced from Zhang et al. (2019), who use textual analysis of local hukou policy documents to construct a city-level measure of hukou stringency. We estimate this equation by applying an exponential transformation to the dependent variable and using a Poisson regression model.

However, hukou stringency is likely endogenous to migration costs. To address this concern, we use historical per capita grain output in 2008 as an instrument to estimate the impact of hukou stringency (Imbert et al., 2023).

The intuition behind this instrument stems from the policy objective underlying China's hukou system. As a low-income country with a large agricultural population, China faced the challenge of ensuring adequate food supply for its rapidly urbanizing cities. In response, the hukou system was designed in part to restrict rural-to-urban labor migration and alleviate pressure on urban food provision. Based on this logic, we argue that cities with more limited historical food supply adopted stricter hukou system.

Specifically:

$$Z^s_j = \ln\left(\frac{\text{Grain Output}_j}{\text{Pop}^{\text{Urban}}_j}\right), \quad Z^u_j = \ln\left(\frac{\text{Grain Output}_j}{\text{Pop}^{\text{Rural}}_j}\right).$$

Table A6 reports the estimates of migration cost elasticities with respect to hukou stringency. We find that $\beta^u_{\text{hukou}} > \beta^s_{\text{hukou}}$, indicating that hukou stringency contributes more to the migration costs faced by unskilled labor. This is consistent with the fact that skilled workers often face lower migration barriers and are more likely to obtain local hukou, as many Chinese cities seek to attract skilled migrants.

Table A6: Migration cost and hukou stringency

Migration Cost	Skilled Labor		Unskilled Labor	
	(1)	(2)	(3)	(4)
$\text{HukouStringency}^s_j$	1.510	1.236		
	(0.464)	(0.596)		
$\text{HukouStringency}^u_j$			0.732	4.424
			(0.220)	(1.528)
Method	OLS	2SLS	OLS	2SLS
Observations	1,216	1,216	1,914	1,844

Notes: This table reports estimates of the elasticity of migration costs with respect to hukou stringency. Columns 1–3 present OLS estimates of the migration share equations, while Columns 2–4 report 2SLS estimates using historical per capita grain output as an instrument. Standard errors, clustered at the destination-city level, are reported in parentheses. The regressions are weighted by the 2000 population of each city.

Appendix E Robustness

E.1 The Definition of Migrants

In our main analysis, we classify individuals as migrants if their current residence differs from their hukou (household registration) city. We exclude long-term migrants who left their hukou location more than 10 years ago, as their initial migration decision may no longer reflect current economic conditions. In addition, we include all migrants regardless of their stated motive (work vs. other reasons), and we pool rural and urban hukou holders into a single category.[12]

A potential concern with this measurement is that sample selection may systematically vary by skill group. We examine this in Table A3 by reporting migration duration, stated motives, and hukou type within each skill group.

First, Panel A of Table A3 reveals no systematic heterogeneity in migration duration across skill types, with approximately 20–25 percent of all migrants having left their hukou city more than a decade ago. Second, Panel B demonstrates that the share of unskilled workers migrating for employment is 20 percentage points higher than that of skilled workers. This difference stems mainly from migration for education, improved living conditions, and marriage. To assess the sensitivity of our labor supply elasticity estimates, we restrict the sample to include only those migrating for work, ensuring consistency with our model where migration is driven by wages. We find a quantitatively similar, albeit slightly higher, labor supply elasticity.

A further consideration is whether aggregating rural and urban migrants obscures important heterogeneity. Panel C reveals that 58 percent of unskilled migrants hold rural hukou, compared to only 22 percent of skilled migrants. A potential concern is that these individuals might still be engaged in agriculture, rendering them incompatible with an urban production framework.

To address this, Table A7 shows that less than 5 percent of unskilled laborers are employed in the agricultural sector. When we disaggregate by hukou type, we find no systematic difference between rural and urban unskilled labor: both groups show an agricultural employment share of around 4 percent. As documented by Imbert et al. (2023), these rural migrants move to large cities to integrate into the urban workforce, often facing poor living conditions. Consequently, omitting them from the analysis would lead to a mismeasurement of the effective urban labor supply.

[12] Since the 2015 Census does not explicitly record agricultural hukou status, we define an individual as a rural migrant if they hold land in a rural area.

Table A7: Distribution of Employment by Industry

	Full Sample		Rural		Urban	
Industry Type	Freq.	Share (%)	Freq.	Share (%)	Freq.	Share (%)
Non-Agricultural	64,093	95.71	37,366	95.57	26,727	95.90
Agricultural	2,874	4.29	1,730	4.43	1,144	4.10
Total	66,967	100	39,096	100	27,871	100

Notes: The full sample is restricted to unskilled workers.

E.2 Rural and Urban Unskilled Workers

Another concern regarding the pooling of unskilled workers with rural versus urban hukou is that this aggregation may mask important heterogeneity. For instance, rural unskilled migrants may sort into different labor markets (e.g., disproportionately working in the construction sector) or exhibit different housing consumption behaviors (eg. living in domitory).

We then verify the industrial shares of rural unskilled and urban unskilled labor. Figure A4 shows that the industrial distributions of rural unskilled and urban unskilled labor are highly correlated, with an R^2 of 0.945. This finding helps alleviate concerns regarding potential heterogeneity between the two groups.

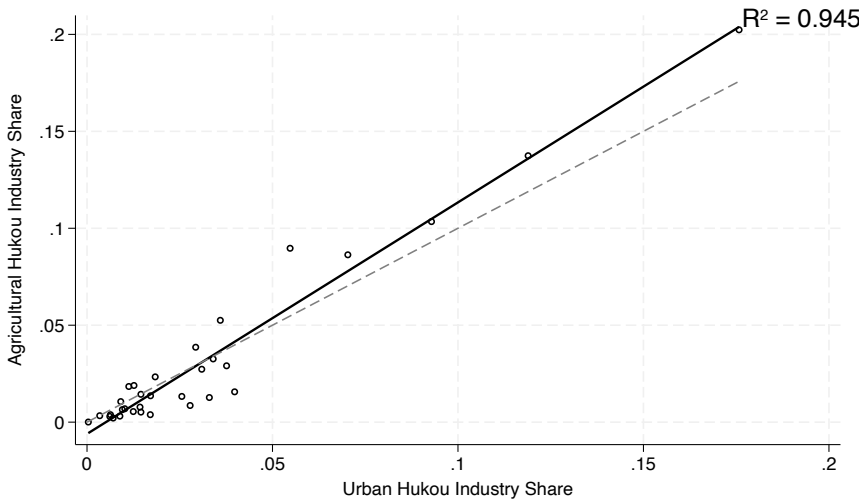


Figure A4: Sectoral Sorting of Labor by Hukou Status

Notes: This figure plots the industrial distribution of agricultural and urban hukou unskilled labor, using data from the 2017 CMDS. The sample is restricted to migrants who moved after 2010.

Regarding housing expenditure heterogeneity, we calibrate the housing shares for skilled and unskilled labor to 0.24 and 0.26, respectively, following Chen et al. (2023). These values are quantitatively similar, suggesting that differences in housing consumption patterns are minimal.

E.3 Model Extension

In our theoretical model, we assume that both skilled and unskilled workers are employed in the tradable sector. However, Table A4 shows reveals that the majority of migrants are actually employed in the consumer service sector. Motivated by this evidence, we extend our framework in this section to incorporate a non-tradable sector. We assume this sector hires only local labor and produces goods that are consumed locally.

The production of non-tradable sectors is given by:

$$Y_{jCS} = A_{jCS}\left[(U_j)^{\zeta} + A_s(S_j)^{\zeta}\right]^{\frac{1}{\zeta}} \tag{A14}$$

Consumers consume tradable goods (the numeraire), local non-tradable services, and housing. We assume housing is a necessity. We assume consumers have Price-Independent Generalized Linearity (PIGL) preferences. The indirect utility of a consumer in city j with income w_j is defined as:

$$V(e_r, P_r) = \frac{1}{\varepsilon}\left(\frac{w_j}{r_j^{\psi} p_{jCS}^{1-\psi}}\right)^{\varepsilon} - \nu \ln\left(\frac{r_j}{p_{jCS}}\right) \tag{A15}$$

with $\varepsilon > 0$ and PIGL shifters $\nu > 0$. p_{jCS} and r_j denote the prices of consumer services and housing, respectively.

Applying Roy's identity, the expenditure shares on housing and consumer services are given by:

$$s_j^h = \psi \; + \; \nu \left(\frac{r^{\psi} p^{1-\psi}}{w_j} \right)^{\varepsilon}, s_j^{CS} = (1-\psi) \; - \; \nu \left(\frac{r^{\psi} p^{1-\psi}}{w_j} \right)^{\varepsilon} \tag{A16}$$

where $s_j^h + s_j^{CS} = 1$.