DISCUSSION PAPER SERIES

DP20201
(v. 2)

AGGLOMERATION AND HUMAN CAPITAL

Yujiang River Chen and Coen Teulings

INTERNATIONAL TRADE AND REGIONAL ECONOMICS, LABOUR ECONOMICS AND MACROECONOMICS AND GROWTH

ISSN 0265-8003

AGGLOMERATION AND HUMAN CAPITAL

Yujiang River Chen and Coen Teulings

Discussion Paper DP20201
First Published 05 May 2025
This Revision 05 September 2025

Centre for Economic Policy Research
187 boulevard Saint-Germain, 75007 Paris, France
2 Coldbath Square, London EC1R 5HL
Tel: +44 (0)20 7183 8801
www.cepr.org

AGGLOMERATION AND HUMAN CAPITAL

Abstract

The high return to human capital on GDP per capita, reaching up to 50% across countries or regions, is puzzling. We develop a spatial model in which regions are classified as either rural areas or monocentric cities. In this framework, both human capital and the concentration of employment in city centers drive knowledge spillovers. Regional land prices clear the market for inter-regional labor mobility, leading to joint predictions for the public return to human capital and for land prices. We test the model using data on wages and real-estate prices for 47 U.S. rural areas and 34 CMSAs from 1979 to 2015. Our findings indicate that the public return on wages is 25% of the private return in rural areas, rising to 136% in cities. The return to human capital on GDP per capita is 24% in the cross section. The rise in urban house prices is largely accounted for by knowledge spillovers.

JEL Classification: J24, J31, I26, R12, R13

Keywords: Agglomeration externalities, Cities, Regional house prices

Yujiang River Chen - yc349@cam.ac.uk
University of Cambridge

Coen Teulings - c.n.teulings@outlook.com
Utrecht University, CEPR and CEPR

Acknowledgements
We thank Daron Acemoglu, Gabriel Ahlfeldt, David Albouy, Michael Amior, David Autor, Nicholas Bloom, Jan Eeckhout, Pieter Gautier, Edward Glaeser, Esteban Rossi-Hansberg, William Janeway, Koen Jochmans, Alan Manning and Guy Michaels, for helpful comments. We also benefited from feedback from seminar audiences at Cambridge, CREST, CSIC, IFS, MIT, NBER, Utrecht and UEA.

Agglomeration and Human Capital

Yujiang Chen and Coen Teulings*

August 2025

Abstract

The high return to human capital on GDP per capita, reaching up to 50% across countries or regions, is puzzling. We develop a spatial model in which regions are classified as either rural areas or monocentric cities. In this framework, both human capital and the concentration of employment in city centers drive knowledge spillovers. Regional land prices clear the market for inter-regional labor mobility, leading to joint predictions for the public return to human capital and for land prices. We test the model using data on wages and real-estate prices for 47 U.S. rural areas and 34 CMSAs from 1979 to 2015. Our findings indicate that the public return on wages is 25% of the private return in rural areas, rising to 136% in cities. The return to human capital on GDP per capita is 24% in the cross section. The rise in urban house prices is largely accounted for by knowledge spillovers.

JEL classification: J24, J31, I26, R12, R13

Keywords: Spatial Sorting, Agglomeration externalities, Cities, Regional wage differentials, Regional house prices

*Yujiang Chen, University of Cambridge, yc349@cam.ac.uk; Coen Teulings, Utrecht University and CEPR, c.n.teulings@outlook.com. We thank Daron Acemoglu, Gabriel Ahlfeldt, David Albouy, Michael Amior, David Autor, Nicholas Bloom, Jan Eeckhout, Pieter Gautier, Edward Glaeser, Esteban Rossi-Hansberg, William Janeway, Koen Jochmans, Alan Manning, and Guy Michaels for helpful comments. We also benefited from useful feedback from seminar audiences at Cambridge, CREST, CSIC, IFS, MIT, NBER, SAET, Utrecht and UEA. All mistakes are our own.

1 Introduction

Throughout history, cities have played a crucial role in driving economic progress. Urbanization increased during periods of exceptional prosperity and declined during subsequent downturns. This pattern is aptly illustrated by Bairoch (1988) (Tables 11.2, 13.1, and 13.4) in the case of Belgium, the world's most urbanized country during the peak of its textile industry between 1300 and 1500. In the 16$^{\text{th}}$ century, the Dutch Republic surpassed Belgium and the latter's urbanization rate declined. Over the following two centuries, the Dutch economy flourished, and the Netherlands emerged as the world's most urbanized country, with Amsterdam serving as the leading global port and financial center. Although the Netherlands remained the world's most prosperous country well into the 18$^{\text{th}}$ century, its dominance—along with its urbanization rate—gradually waned. Only around 1850, with the industrial revolution in full swing, did the United Kingdom overtake the Netherlands as the world's most urbanized nation.

This is just one example of how the concentration of economic activity drove prosperity during the early stages of capitalist development. In more recent phases of growth, Lucas (1988) emphasizes the role of human capital formation as the driving force behind agglomeration, contributing to the sharp divergence in GDP per capita between "the West" and "the Rest." Similarly, Barro and Lee (1996) find a high cross-country correlation between average years of education and GDP per capita. If this correlation were interpreted as causal, the total return to education in terms of GDP per capita would be around 50%—far higher than standard estimates of the private return to wages (roughly 10%). Teulings and van Rens (2008) report the same order of magnitude. Gennaioli et al. (2013) find that, within countries, human capital tends to cluster in specific regions. Exploiting this regional variation in education and GDP per capita, they, too, estimate returns to education that align closely with those reported by Barro and Lee (1996).

Whether this correlation is truly causal remains a subject of debate. Based on changes in compulsory schooling laws, Acemoglu and Angrist (2000) found no evidence of human capital externalities. Acemoglu et al. (2014) argue that institutional quality, rather than human capital, is a more plausible explanation for the large cross-country differences in GDP per capita observed since the beginning of the capitalist era. Nonetheless, even if one questions the causal interpretation of a simple regression of GDP per capita on average years of education, the consistently large coefficient of this variable remains an empirical fact that calls for explanation.

This paper addresses the issue using two complementary approaches. First, we combine predictions about wages and housing prices. Second, we examine the interaction between human capital and urbanization. If human capital generates a public return in excess of the private return, then regions with higher average human capital should exhibit higher wages, even after accounting for individual workers' own human capital. However, inter-regional labor mobility tends to equalize utility across regions, implying that differences in house prices must offset differences in wages. Furthermore, the spatial structure of monocentric cities is more favorable to knowledge spillovers than that of rural areas. As a result, house-price differentials should be more pronounced in cities than in rural areas. Combining both approaches allows us to test these predictions.

We develop a simple general equilibrium model in which workers with varying levels of human capital are perfectly mobile across regions. The concentration of high-human-capital workers in specific regions generates stronger agglomeration externalities in those areas. Regional house prices (or, more precisely, land prices) adjust to equalize utility across regions. Following Lucas and Rossi-Hansberg (2002), we assume that residential land use constrains the degree of spatial agglomeration. The model contrasts two idealized forms of spatial organization: rural areas and cities. In rural areas, workers live and work at the same location, and their ideas must travel between these locations to generate knowledge spillovers. In cities, workers commute to a central business district (CBD) where employment is concentrated, allowing ideas to be exchanged without transmission costs. When the intensity of knowledge spillovers is low—as with low-human-capital workers—commuting costs outweigh the benefits of idea transmission, making the rural structure more efficient. Conversely, when spillover intensity is high—as with high-human-capital workers—the benefits of frictionless idea exchange in cities outweigh commuting costs, making the urban structure with a CBD more efficient.

We emphasize the critical importance of the elasticity of substitution between land and other forms of consumption, which implies that the commonly assumed Cobb-Douglas utility function is too restrictive for our purposes. In addition, we introduce two types of regional amenities: one endogenous to the model (regional consumption spillovers) and one exogenous (an agreeable climate, measured by average January temperature). The role of consumption spillovers in urban areas is well documented (Ahlfeldt et al. (2015); Jacobs (2016); Gautier et al. (2010); de Groot et al. (2015); Teulings et al. (2018); Diamond (2016); Almagro and Domínguez-Iino (2024)). For our model to account for the observed data, the interaction between spillovers on both the production and consumption sides of the economy is essential. Incorporating January temperature introduces additional testable predictions: its direct effect is to push up land prices, but not wages, since climate does not influence productivity. However, there is also an indirect effect: higher land prices reduce land consumption per capita, leading to greater population density and, in turn, stronger agglomeration externalities.

The model generates joint predictions for regional fixed effects in log wages and log land expenditure across rural areas and cities. We test these predictions for the United States using data on 34 cities and 47 rural areas spanning the period from 1979 to 2015. By analyzing wages and land expenditure jointly, the model produces a set of overidentifying restrictions that can be empirically tested. These restrictions are largely supported by the data, and the model's predictions are broadly confirmed. We find a public return to human capital—over and above the private return—of 0.25 log points in rural areas and as much as 1.36 log points in cities. Consistent with the model's predictions, these differentials are magnified still further in terms of land prices: 1% higher wages in a city due to an increase in its average human capital raises land prices there by 7.8%. This mechanism explains the house price increases in cities in many countries over the past decades. Counterfactual calculations show that regional sorting of human capital and the spatial organisation of knowledge spillovers in cities each account for 10% of GDP. We show the public return to be much higher in the cross section than in the time series. When human capital increases in one region holding its level in other regions constant (= cross section), knowledge spillovers in that region are much higher than when human capital increases in all regions simultaneously (= time series). Knowledge spillovers are only 13% in the time series of what they are in the cross section.

Our paper connects three strands in the literature. The first relates to the public return to human capital (e.g., Acemoglu and Angrist (2000); Bils and Klenow (2000); Moretti (2004); Moretti (2011)). This strand focuses on identifying suitable instruments for the unobserved component of human capital and understanding regional variation in human capital. We address the first issue by means of a single index model for human capital, providing new evidence in favor of this approach by showing that observed and unobserved human capital are close substitutes. To address the second issue, we use Bartik (or shift-share) instruments to identify demand shocks. While these instruments are strong, they do not pass overidentification tests when disaggregated into separate instruments by industry. As a result, the credibility of our findings rests on the combined strength of a reliable method for identifying unobserved human capital and a structural model that captures the joint implications for wages and land rents.

The second strand of the literature comes from macroeconomics, which has sought to reconcile the standard Solow growth model with the pronounced divergence in GDP per capita across countries over the past two centuries (e.g., Lucas (1988); Romer (1990); Gennaioli et al. (2013); Duranton and Puga (2023)).

The final strand concerns the literature on the internal structure of cities, notably the theoretical model by Lucas and Rossi-Hansberg (2002), the corresponding welfare analysis in Rossi-Hansberg (2004), and subsequent applications on the impact of the Berlin Wall Ahlfeldt et al. (2015) and the rise of London after 1850 as the first modern metropolis Heblich et al. (2020). By combining the approaches in Lucas and Rossi-Hansberg (2002) and Gennaioli et al. (2013), we address a technical issue in the specification of agglomeration externalities. Gennaioli et al. (2013) adopt a log-linear (or supermodular) specification based on the wage bill and the average level of human capital. In this setup, the marginal effect of a higher mean level of human capital on log wages is independent of the regional wage bill. Consequently, concentrating more high-human-capital workers in one location does not amplify agglomeration externalities. Achieving a Ricardian comparative advantage for high-density regions in terms of knowledge spillovers requires log supermodularity rather than simple supermodularity, as shown in Sattinger (1975) and Teulings (1995).

In Section 2, we describe our theoretical model, while Section 3 presents the empirical evidence. We illustrate the model's implications through a series of counterfactuals: How much of the economy's growth between 1979 and 2015 can be attributed to the rise in human capital? What if cities did not exist? What if human capital was evenly distributed across regions? Finally, Section 4 discusses the broader implications of our analysis, including the increase in urbanization, the growing segregation of worker types between major cities and rural areas, and the resulting political consequences—notably, the widening political divide between urban and rural regions and the rise of populism.

2 Spatial equilibrium model

2.1 Regional labor and land markets

We consider an economy composed of regions indexed r. Each region produces a tradable commodity characterized by an index H_r (a real number; its interpretation is discussed below). Consumers combine these tradable commodities into a single composite commodity used for private consumption. Because there are no inter-regional trade costs and markets for tradable goods are perfectly competitive, the price of this composite commodity is the same across all regions. Let p denote the log of this price. Without loss of generality, we define the units of the composite commodity such that its price is equal to unity, hence $p = 0$.

Workers share identical homothetic utility functions over this composite commodity, land use for residential purposes, and two amenities. This utility function first combines the composite commodity and land use into private consumption using a CES aggregator with an elasticity of substitution no greater than one. It then combines private consumption with the two amenities using a Cobb-Douglas aggregator with an elasticity of substitution equal to unity. Because the amenities are not priced, it is convenient to adopt a mixed representation of utility: an indirect utility function with respect to the prices of the composite commodity and land, and a direct utility function for the quantities of the amenities. In log-log form, and using $p = 0$ as our numeraire, utility is given by

$$\underbrace{u\left(y, p_r, a_r\right)}_{\text{utility}} = \underbrace{y}_{\text{log income}} - \underbrace{\overline{\eta}^{-1} \ln\left(\overline{\lambda} + \lambda \mathrm{e}^{\overline{\eta} p_r}\right)}_{\text{price index of private consumption}} + \underbrace{\alpha' a_r}_{\text{(dis)amenities}}, \tag{1}$$

where p_r is the log price of land and a_r is the vector of (dis)amenities, both in region r; λ denotes the land share in private consumption when the land price is unity (i.e., $p_r = 0$); η is the elasticity of substitution between tradables commodities and land; and the vector α collects the value shares of the (dis)amenities a_r in total utility relative to private consumption. Throughout, Greek letters denote parameters and Latin letters denote variables. A bar above a parameter indicates its complement to unity, for example $\overline{\lambda} \equiv 1 - \lambda$. These assumptions imply $0 < \eta \leq 1$, $0 < \lambda < 1$. Because the utility function is homothetic, utility is composed of two additive terms: log income and a term capturing all other arguments. Without loss of generality, we scale the utility function such that the coefficient on log income is equal to one.

Workers supply one unit of labor and are endowed with a level of human capital h (a real number). They live and work in the same region. Human capital raises their log wage in region r, denoted $w_r(h)$, $w_r'(h) > 0$. A worker's wage is her sole source of income, hence $y = w_r(h)$. Workers choose the region r in which they live and work so as to maximize their utility. Since there are no costs to moving between regions, worker mobility equalizes the utility of each h-type worker across regions to a nationwide benchmark $u^*(h)$:

$$u\left[w_r(h), p_r, a_r\right] = u^*(h), \quad \forall r, \ \forall h. \tag{2}$$

Since $u\left[w_r(h), p_r, a_r\right]$ depends on h only through $w_r(h)$, and since $u\left[w_r(h), p_r, a_r\right]$ is additive in $w_r(h)$ and its other arguments p_r and a_r, equations (1) and (2) imply that $w_r(h)$ must be equal to a region-specific intercept w_r and a nationwide function $w(h)$ that captures the return to human capital:

$$w_r(h) = w_r + w(h), \quad w'(h) > 0.$$

We have not yet specified the units of measurement for the human capital index h. Without loss of generality,[1] we can define this index such that $w(h)$ is linear in h with a unit slope of

$$w_r(h) = w_r + h. \tag{3}$$

The fixed effect w_r is therefore the log nominal income of an individual with a human capital index $h = 0$ in region r, while the index itself serves as a measure of log human capital evaluated at market prices. The region-specific intercepts w_r in the wage function $w_r(h)$ will be determined endogenously. Since utility is invariant to any increasing transformation, the intercept of $u^*(h)$ can be normalized to zero without loss of generality; hence

$$u^*(h) = h. \tag{4}$$

Alternatively, a worker's human capital index h can be interpreted as her log real income mea-

[1] Consider an alternative human capital index h^*, such that $y = w^*(h^*) + w_r$ with $w^{*\prime} > 0$. We can use $h = w^*(h^*)$ rather than h^* as the human capital index without altering the implications of the model.

sured in terms of utility. Analogously, H_r represents the average log real income in region r.

Regions are characterized by two types of consumption amenities. The first is an exogenous attribute, such as January temperature, a common proxy in the literature for quality of life. People tend to prefer regions with more agreeable January temperatures (e.g., Glaeser and Gottlieb (2009)). In keeping with this interpretation, we denote the exogenous amenity as T_r. The second amenity is endogenous to the model. Following Ahlfeldt et al. (2015), we allow for regional consumption spillovers. A higher log density of purchasing power per unit of land, denoted z_r, supports a denser network of services, such as shops, restaurants, and cultural performances (Jacobs (2016); Gautier et al. (2010); de Groot et al. (2015); Diamond (2016); Almagro and Domínguez-Iino (2024)). Thus, the vector of local amenities a_r and their corresponding utility weights α in equation (1) read

$$a_r = [T_r, z_r]', \quad \alpha = [\alpha_T, \alpha_z]'. \tag{5}$$

Land rents are collected by a class of absentee landlords. Since workers always receive their outside utility $u^*(h) = h$, landlords serve as the residual claimants in the model.

Due to the homotheticity of consumption, land demand is proportional to log real income h. Analogous to equation (3), the land demand $l_r(h)$ of an individual with human capital index h in region r can be expressed as a region-specific fixed effect plus log real income h:

$$l_r(h) = l_r + h. \tag{6}$$

Analogous to w_r, l_r denotes the log land demand of an individual with a human capital index $h = 0$ in region r. For our empirical analysis, we use data not on log land prices but on log land expenditure, given by $l_r + p_r$. We therefore define v_r as the log land expenditure for a worker with human capital index $h = 0$ in region r:

$$v_r \equiv l_r + p_r.$$

Proposition 1 characterizes l_r and v_r as a function of log wages and January temperature.

Proposition 1: Regional land prices and land expenditure

Assume:

$$\lambda - \overline{\eta}_\lambda < \overline{\eta}_\lambda \alpha_z < \lambda. \tag{7}$$

A first-order Taylor approximation of log land use and log land expenditure around the point $p_r = 0$ satisfies:

$$\begin{aligned}
l_r &= \lambda_l + \overline{\lambda}_z w_r - \lambda_z \alpha_T T_r, \qquad (8)\\
v_r &= \lambda_0 + \lambda_w w_r + \lambda_T T_r,\\
\eta_\lambda &\equiv \overline{\lambda}\overline{\eta}, \quad 0 \leq \eta_\lambda < 1,\\
\lambda_z &\equiv \frac{\overline{\eta}_\lambda}{\lambda - \overline{\eta}_\lambda \alpha_z} > 1, \quad \lambda_l \equiv \frac{\lambda}{\overline{\eta}_\lambda}\lambda_z \ln \lambda\\
\lambda_0 &\equiv \frac{\overline{\eta}_\lambda - \eta_\lambda \lambda \lambda_z}{\overline{\eta}_\lambda^2} \ln \lambda, \quad \lambda_w \equiv 1 + \frac{\eta_\lambda}{\overline{\eta}_\lambda}\lambda_z > 1, \quad \lambda_T \equiv \frac{\eta_\lambda}{\overline{\eta}_\lambda}\lambda_z \alpha_T > 0.
\end{aligned}$$

Proof: See Appendix.■

The special case $\eta = 1$ (Cobb-Douglas preferences) helps to clarify Proposition 1. Regions with higher nominal wages (higher w_r) compensate for higher land prices. When $\eta = 1$, we have $\eta_\lambda = 0$, which implies $\lambda_w = 1$ and $\lambda_T = 0$; as a result, land expenditure varies proportionally with nominal income, reflecting a constant share of consumption. When $\eta < 1$, land prices increase more than proportionally to wages because the log land share in consumption, $v_r - w_r$, rises with wages. This effect is further amplified by consumption spillovers $\alpha_z > 0$: higher land prices lead to greater population density, which increases consumption spillovers. The same logic applies to the effect of January temperature.

Condition (7) is necessary for a bounded equilibrium. If this condition were violated, an increase in local land prices would cause workers to substitute away from land use toward the composite commodity, leading to a higher population density. This, in turn, would increase local purchasing power and consumption spillovers, further driving up land prices. As a result, land prices would become unbounded, and all workers would cluster at a single location. Condition (7) holds when the consumption spillover α_z and the elasticity of substitution η are sufficiently low.

2.2 Agglomeration externalities on the production side

Production in each region follows a Leontief technology that requires inputs of all levels of human capital h in fixed proportions. These proportions are determined by the density function of a normal distribution with common variance σ^2 across regions and a region-specific mean $H_r - \frac{1}{2}\sigma^2$.[2] In combination with equation (3), this assumption implies that the wage distribution within each region is log normal, which aligns with observed data. H_r is exogenously determined by the composition of industries in a region. For example, the concentration of the IT industry in San Francisco and San Jose, and R&D activities in Boston, results in much higher values of H_r in these areas compared to rural states like Louisiana, Mississippi, and Georgia. Production does not require land or physical capital as inputs.

Regional TFP depends on intra-regional agglomeration externalities or knowledge spillovers: [3]

$$w_r = \psi \left(1 + \chi H_r\right) \left(\chi_0 + m_r + H_r\right) - \psi, \tag{9}$$

where m_r is the log number of workers contributing to the agglomeration externality at a given point in space, and $0 < \psi < 1$, $\chi > 0$, and χ_0 are parameters. Later, we will choose a convenient parameterization for χ_0. The factor $\psi \left(1 + \chi H_r\right)$ measures the strength of knowledge spillover. The magnitude of these spillovers equals ψ in a region where the human capital index $H_r = 0$. However, because $\chi > 0$, the spillovers increase with the human capital index. We assume that agglomeration externalities remain positive even in the region with the lowest human capital

[2] By defining the mean level of human capital as $H_r - \frac{1}{2}\sigma^2$, the log of mean worker income in region r is equal to $w_r + H_r$.

[3] We view the term $(\chi_0 + m_r + H_r)$ in equation (9) as a linear approximation of the correct non-linear specification $\ln\left[1 + \exp\left(\chi_0 + m_r + H_r\right)\right]$. This approximation holds for large $m_r \to \infty$, since

$$\lim_{m_r \to \infty} \left(\ln\left[1 + \exp\left(\chi_0 + m_r + H_r\right)\right] - \left(\chi_0 + m_r + H_r\right)\right) = 0.$$

Normally, this approximation is appropriate. However, at one point in the paper, we apply the opposite approximation, which holds for small $m_r \to -\infty$ and hence $M_r \equiv \mathrm{e}^{m_r} \to 0$, yielding $\exp\left(\chi_0 + m_r + H_r\right)$ since

$$\lim_{M_r \to 0} \frac{\ln\left(1 + M_r \mathrm{e}^{\chi_0 + H_r}\right) - M_r \mathrm{e}^{\chi_0 + H_r}}{M_r} = 0.$$

The non-linearity of $\ln\left[1 + \exp\left(\chi_0 + m_r + H_r\right)\right]$ means that w_r for a worker who benefits from no spillovers is equal to zero:

$$\lim_{m_r \to \infty} w_r = \lim_{m_r \to \infty} \left(\psi \left(1 + \chi H_r\right) \ln\left[1 + \exp\left(\chi_0 + m_r + H_r\right)\right]\right) = 0,$$

yet equal to minus infinity if equation (9) also held for small m_r.

index:

$$H_r > -\chi^{-1}, \forall r \Rightarrow 1 + \chi H_r > 0. \tag{10}$$

Knowledge spillovers for a given worker are proportional to the log of the total wage bill of her coworkers in the neighborhood, $m_r + H_r$.

Our specification differs slightly from that in Gennaioli et al. (2013), who use a log-linear (or supermodular) specification of the total log wage bill and the human capital index

$$w_r = \psi\left(\chi_0 + m_r + \theta H_r\right),$$

where $\theta > 1$. For $\theta = 1$ and $\chi = 0$, their specification and ours are identical: knowledge spillovers are proportional to the total log wage bill. For $\theta > 1$, knowledge spillovers increase more than proportionally with the human capital index H_r. Gennaioli et al. (2013) provide evidence that this is indeed the case.

Their log-linear specification, however, imposes an important constraint: an increase in H_r has the same effect on log wages regardless of the size of the total wage bill $m_r + H_r$. In other words, the knowledge spillover from a rise in a region's average human capital per worker H_r affects a worker's log wage equally, whether that higher average comes from a single colleague or a legion of coworkers. Only the average level of human capital—not the number of people brought together—matters. This is not what we mean by agglomeration externalities. Ricardian comparative advantage of high human capital in knowledge spillovers requires these factors to enter in a log-supermodular way rather than simply supermodular, similar to productivity effects of additional human capital in complex jobs, as in the labor market models by Sattinger (1975) and Teulings (1995). The specification in equation (9) allows for log supermodularity.

Our modeling of the intra-regional spatial structure builds on ideas from Lucas and Rossi-Hansberg (2002), Rossi-Hansberg and Wright (2007a), and Ahlfeldt et al. (2015). Residential land use means workers cannot cluster at a single point in space. Following Lucas and Rossi-Hansberg (2002) and Ahlfeldt et al. (2015), knowledge spillovers diminish with distance: at distance s, only a fraction $\mathrm{e}^{-\delta s}$ of spillovers remains. Similarly, commuting incurs a cost proportional to distance, where a share κ of labor supply is lost per unit of distance. Thus, at distance s, only $\mathrm{e}^{-\kappa s}$ of a worker's labor supply is effective, impacting both her output and the knowledge spillovers she provides to nearby colleagues. We assume $\delta > \kappa$. This is confirmed by empirical evidence from Berlin in Ahlfeldt et al. (2015) (Table 5), where δ exceeds κ by a factor 5 or more.

We consider two contrasting archetypal spatial structures: rural areas and cities. In a rural areas, people work where they live. Since workers do not commute, ideas must travel between workplaces. In cities, the opposite holds. Following Rossi-Hansberg and Wright (2007a), all employment is concentrated in the CBD, which uses no land since land is not an input in production. Thus, all jobs are located at a single point in space, eliminating any loss in the transmission of ideas. In this case, ideas do not travel; workers do. We treat a region's spatial structure as exogenous. Consequently, each region in our model is characterized by three exogenous factors: its mean level of human capital H_r, its January temperature T_r, and whether its spatial structure is that of a rural area or a city.

For both rural areas and cities, we assume that all h-type workers are distributed uniformly across space. Workers choose their land use (or, equivalently, their lot size) to maximize utility. Regional log land prices p_r adjust to clear the land market. Rural areas are scale-free and can extend infinitely. There, land prices p_r are determined solely by local amenities and TFP. The supply of land affects only the population size, not land prices. In contrast, cities have a limited scale because workers must commute to the CBD; beyond a certain size, commuting costs become prohibitive. Without loss of generality, we normalize the nationwide means across all cities to zero:

$$\mathrm{E}\left[H_r | r = \text{city}\right] = \mathrm{E}\left[T_r | r = \text{city}\right] = 0, \tag{11}$$

2.2.1 Rural areas

As noted, in rural areas, workers live and work at the same location. According to equation (9), the knowledge spillover w_r^r in a rural region (denoted by the superscript r) is equal to

$$\begin{aligned} w_r^r &= \psi\left(1 + \chi H_r\right)\left(\ln\left[\int_0^\infty 2\pi s \mathrm{e}^{-\delta s - l_r - H_r} \mathrm{d}s\right] + \chi_0 + H_r\right) - \psi \\ &= \psi \chi H_r + \psi\left(1 + \chi H_r\right)\left(\lambda_z \alpha_T T_r - \overline{\lambda}_z w_r^r\right), \\ \chi_0 &\equiv 1 + 2\ln\delta - \ln 2 - \ln\pi + \lambda_l. \end{aligned} \tag{12}$$

In the first equality, the integral represents the log number of workers contributing to agglomeration benefits m_r. Consider a worker at a particular location and let s be the distance from this point to another location in the rural area; $2\pi s$ is the circumference of the circle at distance s; $\mathrm{e}^{-\delta s}$ is the fraction of knowledge spillovers that survive over this distance; and $\mathrm{e}^{-l_r - H_r}$ is the population density.[4] Solving the integral and substituting the expression for χ_0 yields the second line. We normalize χ_0 such that $\psi\chi$ measures the agglomeration effect of human capital in rural areas.

Solving in equation (12) for w_r^r and applying a first-order Taylor expansion in H_r and T_r yields:[5]

$$
\begin{aligned}
w_r^r &\cong \omega_H H_r + \omega_T T_r, \\
v_r^r &\cong \lambda_0 + \lambda_w \left(\omega_H H_r + \omega_T T_r\right) + \lambda_T T_r, \\
\psi_1 &\equiv \frac{\psi}{1 + \overline{\lambda}_z \psi}, \quad \omega_H \equiv \psi_1 \chi, \quad \omega_T \equiv \psi_1 \lambda_z \alpha_T,
\end{aligned}
\tag{13}
$$

where $\cong$ implies that we use a first-order Taylor approximation around the point $H_r = T_r = 0$.

A bounded equilibrium requires the following assumption:

$$
\psi \left(1 + \chi H_r\right) < -\overline{\lambda}_z^{-1}, \forall r \Rightarrow \psi_1 > \psi, \quad \omega_H > 0, \quad \omega_T > 0. \tag{14}
$$

This is the equivalent of assumption (7), but for the production rather than the consumption side of the economy. It holds when the elasticity of substitution η, the consumption spillover λ_z, the agglomeration parameter ψ, and the human capital effect χH_r on knowledge spillovers are all sufficiently low. If this assumption fails, agglomeration externalities would cause wages to rise, pushing up land prices and reducing land use per capita. This would amplify agglomeration externalities further, leading to unbounded productivity growth and the concentration of all economic activity at a single point in space. This is one reason why allowing $\eta < 1$ is crucial: a Cobb-Douglas utility function makes it more difficult to satisfy assumptions (7) and (14). A second reason is discussed in the empirical section; see Table VII.

[4] Population density is the inverse of average land use; l_r is log land use for a worker with $h = 0$, and $l_r + H_r$ is log mean land use for all workers; see also footnote 2.

[5] The exact solution for w_r^r reads

$$
w_r^r = \frac{\psi\chi H_r + \psi\lambda_z\alpha_T T_r + \psi\chi\lambda_z\alpha_T H_r T_r}{1 + \overline{\lambda}_z \psi \left(1 + \chi H_r\right)}.
$$

Equation (13) implies that the regional fixed effect in log wages w_r^r increases with the mean level of human capital H_r: a higher mean boosts knowledge spillovers, adding to the private return to human capital. Since the private return has been normalized to unity (see equation (3)), ω_H can be interpreted as the public return to human capital—over and above the private return—expressed as a share of that private return.

The fixed effect also increases with January temperature. While the labor supply curve suggests that, holding land prices constant, wages should decrease with the January temperature since a pleasant climate serves as a compensating differential for workers, knowledge spillovers reverse the sign of this effect in general equilibrium. The reason for this counterintuitive implication is that the initial negative effect of an agreeable climate on wages is offset by higher land prices. In the absence of spillovers, wages would be determined solely by labor demand. The positive utility effect of January temperature would be exactly offset by higher land prices, leaving wages unaffected. However, once knowledge spillovers come into play, a higher January temperature raises land prices relative to the composite commodity. Workers respond by reducing their land use, increasing regional population density and thereby boosting knowledge spillovers. This may help to explain why the IT industry clusters in Californian cities: the pleasant climate supports high land prices, leading to high population density conducive to knowledge spillovers. This counterintuitive effect offers a clear empirical test of the model.

Like the fixed effect in log wages, log land expenditure increases with both human capital and January temperature. For January temperature, this occurs by two mechanisms: directly, because labor supply dictates that the amenity of a more pleasant climate is offset by higher land prices; and indirectly, because these higher land prices reduce land use, increasing population density and thereby boosting knowledge spillovers. Since $\lambda_w > 1$, land prices are more sensitive than wages to changes in the mean level of human capital. The higher the consumption spillover α_z, the larger this excess sensitivity.

2.2.2 Cities

The previous section derived the magnitude of knowledge spillovers in rural areas. This section repeats the exercise for cities. As noted earlier, these two types of regions have contrasting spatial structures: in rural areas, ideas travel but workers do not; in cities, workers commute but ideas do not. Therefore, while κ is irrelevant for rural areas, δ is irrelevant for cities.

Since commuting costs reduce a worker's labor supply by a fraction $1 - \mathrm{e}^{-\kappa s}$ at distance s from the CBD, the log nominal income $w_r^c(s)$ of a worker with human capital index $h = 0$ living at distance s from the CBD in city r reads:

$$w_r^c(s) = w_r^c - \kappa s,$$

where the superfix c denotes cities, w_r^c is the log wage in the CBD, and κs is the share of the labor supply lost to commuting. Substituting $w_r^c(s)$ in equation (8) yields the land demand and land expenditure at distance s from the CBD:

$$\begin{aligned} l_r^c(s) &= l_r - \overline{\lambda}_z \kappa s, \\ v_r^c(s) &= v_r - \lambda_w \kappa s. \end{aligned} \tag{15}$$

Since $\overline{\lambda}_z < 0$, demand increases with distance s. Because locations further from the CBD necessitate longer commuting times, land is cheaper in those locations. As a result, workers substitute away from the composite commodity toward greater land use. Since land use is the inverse of population density, this density decreases with s.

At the city boundary, S_r a worker is indifferent between working in the CBD and working at their home location while benefiting from the knowledge spillovers of workers in the surrounding rural area. For simplicity, we assume that the area surrounding the city shares the same human capital index H_r as the city itself. At first glance, one might expect the log wage in the surrounding area to be the same as the log wage w_r^r that would prevail if the entire region were a rural area (see equation (13)). However, this is not the case.

Consider a worker living just outside the city boundary S_r. She benefits from the knowledge spillovers of neighboring workers who also live outside the city—roughly half of her neighbors. The other half live inside the city and commute to the CBD, contributing to the knowledge spillovers there instead. As a result, in the limiting case where $\delta/\kappa \to \infty$, the effective number of contributing neighbors to the local knowledge spillover is just half that in a fully rural area.[6] The outside log wage is therefore $\frac{1}{2}w_r^r$.[7] Hence, w_r^c relates to the city's boundary S_r by

$$w_r^c(S_r) = w_r^c - \kappa S_r = \frac{1}{2}w_r^r \Rightarrow \Delta_r \equiv w_r^c - \frac{1}{2}w_r^r = \kappa S_r \geq 0. \tag{16}$$

Thus, Δ_r equals the log income differential between a worker living directly beside the CBD and a worker living at the city's periphery, who loses a share $1-\mathrm{e}^{-\kappa S_r}$ of her labor supply to commuting costs. Equations (15) and (16) yield Proposition 2.

Proposition 2: City population size and labor supply

The city's log population n_r^c, its log labor supply in the CBD m_r^c, the log average land expenditure v_r^c, and the log wage surplus Δ_r satisfy:

$$\begin{aligned}
n_r^c &= \ln\left(\int_0^{S_r} 2\pi s \mathrm{e}^{-(l_r-\overline{\lambda}_z\kappa s)-H_r}\mathrm{d}s\right) \cong \ln\pi + 2\ln\Delta_r + \frac{2}{3}\overline{\lambda}_z\Delta_r - 2\ln\kappa - l_r - H_r, \qquad (17)\\
m_r^c &= \ln\left(\int_0^{S_r} 2\pi s \mathrm{e}^{-\kappa s-(l_r-\overline{\lambda}_z\kappa s)-H_r}\mathrm{d}s\right) \cong \ln\pi + 2\ln\Delta_r - \frac{2}{3}\lambda_z\Delta_r - 2\ln\kappa - l_r - H_r,\\
v_r^c &= \ln\left(\int_0^{S_r} 2\pi s \mathrm{e}^{(v_r-\lambda_w\kappa s)-(l_r-\overline{\lambda}_z\kappa s)-H_r}\mathrm{d}s\right) - n_r^c \cong v_r^c(0) - \frac{2}{3}\lambda_w\Delta_r,\\
\Delta_r &\cong 2\psi_2\left(1+\frac{\psi_2}{\psi_1\psi}\omega_H H_r\right)\left[\rho + \ln\Delta_r + \frac{1}{4\psi_1}\left(\omega_T T_r + \omega_H H_r\right)\right],
\end{aligned}$$

where

$$\psi_2 \equiv \frac{\psi}{1+\psi\left(1-\frac{1}{3}\lambda_z\right)}, \qquad \rho \equiv \ln(\delta/\kappa) - \frac{1}{2}\ln 2 \tag{18}$$

[6] This can be understood as follows: consider all coworkers located within a distance of, e.g., $3/\delta$ from a worker living and working at location $s = S_r$ just outside the city. These coworkers reside within a circle of radius $3/\delta$ centered on the worker. They account for a fraction $\int_0^{3/\delta}\mathrm{e}^{-\delta s}\mathrm{d}s = (1-\mathrm{e}^{-3})\delta^{-1} = 0.95\delta^{-1}$ of the worker's knowledge spillovers. Since this worker is positioned just beyond the city boundary, which lies at radius S_r, the city boundary intersects the midpoint of her spillover circle. If S_r is much larger, e.g., $10/\delta$, the arc of the city boundary within the spillover circle can be approximated by a straight line through its center, effectively bisecting the circle. Only those coworkers residing in the half that lies outside the city boundary contribute to the worker's local spillovers, since the other half commute to the CBD and contribute there instead. Consequently, the knowledge spillovers at the boundary are only half as large as they would be in a rural area without the presence of a neighboring city.

[7] We use the approximation of w_r^r, which holds for small $M_r \equiv \mathrm{e}^{m_r} \to 1$; see footnote 3:

$$w_r^r = \psi(1+\chi H_r) M_r^r \mathrm{e}^{\theta_0 + H_r}.$$

Since $M_r = \frac{1}{2}M_r^r$ for a worker living just outside the city, the outside log wage is $\frac{1}{2}w_r^r$.

and using

$$\ln\left(\int_0^S s\mathrm{e}^{-\kappa s}\mathrm{d}s\right) = \ln\left(\int_0^{\kappa S} x\mathrm{e}^{-x}\mathrm{d}x\right) - 2\ln\kappa \cong 2\ln S - \ln 2 - \frac{2}{3}\kappa S, \tag{19}$$

where $\cong$ implies that we are using a Taylor approximation of the log lower incomplete gamma function in equation (19).

A solution for Δ_r exists if

$$1 - \ln 2 < \rho + \frac{1}{4\psi_1}\left(\omega_T T_r + \omega_H H_r\right) + \ln\psi_2 + \ln\left(1 + \frac{\psi_2}{\psi_1\psi}\omega_H H_r\right), \tag{20}$$

specifying a lower bound for H_r and T_r for this condition to be satisfied. Let $H^{\min}$ and $\Delta^{\min}$ be the lowest values of H_r and Δ_r for which the condition is satisfied when $T_r = 0$. $H^{\min}$ and $\Delta^{\min}$ satisfy

$$1 - \ln 2 = \rho + \frac{\omega_H}{4\psi_1}H^{\min} + \ln\psi_2 + \ln\left(1 + \frac{\psi_2}{\psi_1\psi}\omega_H H^{\min}\right) \tag{21}$$

$$\Delta^{\min} = 2\psi_2\left(1 + \frac{\psi_2}{\psi_1\psi}\omega_H H^{\min}\right). \tag{22}$$

Proof: See Appendix. ■

The integrals in equation (19) are incomplete gamma functions. We use a Taylor approximation to evaluate them, which is almost exact for the relevant values of $\kappa S_r = \Delta_r < 1$. For this reason, we use the equality sign "$=$" instead of the approximate equality sign "$\cong$" when applying this approximation. The result has a clear interpretation: the first two terms, $2\ln S - \ln 2$, are equal to $\ln\int_0^{\kappa S} x\mathrm{d}x - 2\ln\kappa$ (which omits the factor e^{-x} from the integrand) and capture the city's log surface area. The third term, $\frac{2}{3}\kappa S$, reflects the differential effect of the factor e^{-x}.

The integrals have a similar structure to that in equation (12). The city's population is computed by integrating over all distances from the city center at $s = 0$ to the edge at $s = S_r$ over the circumference $2\pi s$ at location s times the density $\mathrm{e}^{\overline{\lambda}_z\kappa s - l_r - H_r}$ at that location, as given by equation (15). To calculate the labor supply m_r^c, the population is multiplied by a factor $\mathrm{e}^{-\kappa s}$ to reflect the commuting cost, which reduces the effective labor supply of workers living at distance s from the CBD.

To calculate the log average land expenditure per person, v_r^c, we must account for the decreasing land prices and increasing lot sizes farther from the CBD, again using equation (15).[8] Log land expenditure at the city's edge is just $v_r^c(S_r) = v_r^c(0) - \lambda_w \Delta_r$ (see equations (8) and (16)). Average log land expenditure is a weighted average of $v_r^c(0)$ and $v_r^c(S_r)$, with weights $1/3$ and $2/3$, respectively; since available land at distance s from the CBD is proportional to s, land prices near the city's edge have a greater influence on the average than those near the CBD. Substituting the expression for m_r^c into equation (12) and performing further derivations yields the expression for Δ_r^c as an (implicit) function of H_r and T_r, see the final line of equation (17).

A solution for Δ_r must be greater than ψ_2, because when $\Delta_r = \psi_2$, the city's radius shrinks to zero and a city with a CBD cannot exist. This requires ρ, and thus the log ratio of the travel cost of ideas to commuting cost, $\ln(\delta/\kappa)$, to be sufficiently high. In fact, δ must be greater than κ. This condition is intuitive: if the decline in labor supply caused by commuting exceeds the decline in knowledge spillovers, commuting to the CBD is not worthwhile. The negative effect of the lost labor supply due to commuting time outweighs the benefits of concentrating all jobs in the CBD to facilitate the transfer of ideas.

Proposition 3 uses Taylor approximations to express the solution to the model as four linear equations for log wages and log land expenditure in both rural areas and cities.

Proposition 3: City wages, existence, and efficiency

Assume that a solution for Δ_r exists for $H_r = T_r = 0$:

$$\begin{aligned} \rho + \ln \psi_2 &> 1 - \ln 2, \\ \frac{\psi_2}{\psi_1} \frac{\Delta_0}{\Delta_0 - 2\psi_2} &> 1. \end{aligned} \tag{23}$$

[8] Since v_r^c is the average of $v_r^c(s)$ weighted by the number of people at each location s, the integral must be divided by the total number of people, $e^{n_r^c}$. Taking logs then implies that we must subtract n_r^c.

A Taylor approximation of log wages, log land prices, and log city size reads:

$$\begin{aligned} w_r^r &\cong \omega_H H_r + \omega_T T_r, \\ w_r^c &\cong \Delta_0 + \Delta_H \omega_H H_r + \Delta_T \omega_T T_r, \\ v_r^r &= \lambda_0 + \lambda_w \omega_H H_r + (\lambda_w \omega_T + \lambda_T)\, T_r, \\ v_r^c(0) &\cong \lambda_0 + \lambda_w \Delta_0 + \Delta_H \lambda_w \omega_H H_r + (\Delta_T \lambda_w \omega_T + \lambda_T)\, T_r, \end{aligned} \tag{24}$$

where $\Delta_H > \Delta_T > 1$.

Proof: See Appendix. ■

Under assumption (23), the effects of H_r and T_r are smallest for log wages in rural areas and largest for log land expenditure in the CBDs of cities. For H_r, there are two common multipliers greater than one: (i) $\lambda_w > 1$, the *land expenditure multiplier*, which captures the additional effect of human capital on log land expenditure relative to log wages; and (ii) Δ_H, the *city multiplier*, which reflects the additional agglomeration effect of human capital in cities. The higher the mean level of human capital, the larger the wage differential between cities and rural areas, as knowledge spillovers increase more than proportionally with the level of human capital: $\chi > 0$.

For T_r, the effect on log land expenditure consists of two components. The reduced-form parameter λ_T captures the direct effect of a more pleasant climate on land expenditure, which is the same in both rural areas and cities. The reduced-form parameter ω_T reflects the indirect effect of higher land expenditure due to a more pleasant climate on log wages in rural areas. This indirect effect is subject to the same land expenditure multiplier λ_w as human capital but involves a smaller city multiplier, $\Delta_T, 1 < \Delta_T < \Delta_H$.

The equations for w_r^c and $v_r^c(0)$ include a city dummy, $\Delta_0 > 0$ and $\lambda_w \Delta_0 > \Delta_0$, respectively, but no coefficient on the log population size n_r^c. This is because the spatial organization of a region is exogenous in our model, whereas city size is not.

Figure I illustrates the relationship between H_r (on the x-axis) and w_r^r and w_r^c (on the y-axis) for $T_r = 0$, using the parameter estimates discussed in the next section. It shows both the full non-linear relationships from equations (12) and (17) (solid lines) and their linear approximations around the nationwide city mean, $w_r^r = 0$ and $w_r^c = \Delta_0$, from equations (8) and (24) (dotted lines).

The red crosses and blue circles represent the estimated values for U.S. regions (see below). Within the relevant range of H_r, the linear approximations are reasonably accurate, except for the values $H_r < H^{\min}$, where the city form does not exist. At that point, $\Delta_r = \Delta^{\min}$ and $\mathrm{d}w_r^c/\mathrm{d}H_r = \infty$. Equation (24) serves as the starting point for our empirical analysis.

3 Empirical evidence

3.1 Data

Estimating this model requires data on wages, human capital, land expenditure, and January temperature across a set of regions. Data on the human capital index h are derived from individual characteristics, while data on land expenditure are based on house prices. For the latter, we make two auxiliary assumptions. First, the rental rate of housing, including depreciation and maintenance, is constant over time. This implies that housing costs are a fixed fraction of house prices for both owners and renters. Second, housing services are produced using a Cobb-Douglas production function with land and the composite commodity as its inputs. This implies that land costs represent a fixed share of overall housing costs. Under these assumptions, and given the earlier assumptions that all income is consumed and utility is homothetic, the log land expenditure for a worker with human capital index $h = 0$ in region r equals the log mean house price in that region minus the regional mean human capital index H_r. Subtracting this index corrects for regional differences in purchasing power that arise from differences in average human capital. The construction of the H_r index is discussed in the next section.

We draw data from four different sources. Individual-level data come from the Current Population Survey, Merged Outgoing Rotation Groups (CPS-MORG), covering the years 1979 to 2015. From this dataset, we use information on hourly wages, years of education, occupation, industry, and standard personal characteristics such as gender, age, marital status, and race. Our sample includes all workers aged 16 to 64.

For our regional classifications, we select 34 MSAs for which both individual- and regional-level data are available. In each state, the remaining area is treated as a single rural region. For the years 1979 to 1985, we identify MSAs using the 1970 Census ranking; from 1986 to 1988, we use the CMSA and PMSA identifiers; from 1989 to 2003, we rely on MSAFIPS; and for the remaining years, we use CBSAFIPS. Hawaii and Alaska are excluded. Additionally, New Jersey is dropped from rural areas because its entire area falls within NY-NJ MSA. This results in a final sample of 34 MSAs and 47 rural areas.

Regional population and employment data are sourced from the U.S. Census Bureau. The Housing Price Index (HPI) is obtained from the Federal Housing Finance Agency's All-Transaction Indexes. To make housing prices comparable across regions, we also calculated a housing value index using additional data from the Zillow Home Value Index (Zillow Research). We use the estimated median home value for all homes within a region. As state-level data for Maine and Louisiana are missing from this dataset, we use county-level average home values for those states instead. All data span the period from 1979 to 2015 at annual frequency. Average January temperatures are drawn from the U.S. National Oceanic and Atmospheric Administration's 1981–2010 U.S. Climate Normals (see Glaeser et al. (2004)).

3.2 Construction of human capital index h

Estimating the model requires data on the average level of human capital, H_r, and the fixed effect, w_r, in each region. Measuring both variables presents a challenge, as human capital is only partially observable. The observed component of mean human capital H_r^o in region r is likely to be a biased estimate of the actual level, H_r, because the observed and the unobserved components are expected to be positively correlated. When a region's workforce is positively selected on observed human capital, it is also likely to be positively selected on unobserved human capital. As a result, inter-regional variation in observed human capital underestimates the true variation. In this case, the regional fixed effect w_r captures the influence of the omitted unobserved human capital. Consequently, the coefficient on H_r^o in a wage regression is likely to be an upwardly biased estimate of the public return to H_r, as it incorporates the private return to unobserved human capital. We propose a simple correction, based on the notion that observed and unobserved human capital are substitutes in production. We provide empirical support for this assumption below.

An index of the observed human capital for individual i is constructed using a simple log-linear wage equation:[9]

$$
\begin{aligned}
w_i &= w_s + h_i + e_i, \qquad (25)\\
h_i &\equiv h_i^o + \underline{h}_i,\\
h_i^o &\equiv \beta' x_i.
\end{aligned}
$$

For convenience, the index s denotes both the region and the time at which the region is observed; thus, there are $\mathrm{R} \times \mathrm{T}$ values of w_s, where $\mathrm{R} = 81$ and $\mathrm{T} = 37$ represent the number of regions and years, respectively. Since we use cross-sectional data, each individual i is observed in only one region r at a single point in time t. Therefore, each individual i corresponds to a unique s. We refer to each region–time combination as an economy. In equation (25), w_i is the observed log hourly wage for worker i; w_s is an economy fixed effect; h_i^o and $\underline{h}_i$ are the observed and unobserved components of human capital, respectively; and e_i is measurement error in the observed log wage. The vector x_i represents the standard set of personal characteristics for individual i, and β is a vector of parameters, common to all economies, of the same dimension as x_i. This parameter vector aggregates the components x_i into a single index of observed human capital, h_i^o.

In our empirical specification, we account for a number of well-known non-linearities in x_i, such as the experience profile and the interaction between years of education and experience, using dummies for each level of educational attainment. Additionally, we slightly relax the assumption that β is constant across economies by including interaction terms for marital status, gender, and a time trend to capture evolving attitudes toward working women. We also account for the differential impact of being Black in Southern states. Note that inflation and nationwide productivity growth are captured by the time fixed effects.

Without loss of generality, we assume that the components h_i^o and $\underline{h}_i$ are orthogonal across all regions. We note that referring to h_i^o as the observed human capital of a worker glosses over a range of complex issues, such as whether the effects of gender or race might be attributed to differences in human capital or whether these variables are proxies for other factors, such as the

[9]The linearity of this equation is not an important restriction on the generality of the analysis (see Gautier and Teulings, 2008). Suppose wages are an increasing but non-linear function of some human capital index h^*: $w = w(h^*) = w(\beta' x)$, with $w'(h^*) > 0$. By defining a transformed human capital index $h = w(h^*)$, linearity can be imposed without loss of generality. Any non-linearity in $w(\beta' x)$ can then be addressed by applying a polynomial in x, as we do in our empirical specification.

interrupted careers of women or discrimination against women and people of color. However, since our goal is simply to aggregate all observable determinants of a worker's earning capacity into a single index, we set these issues aside.

We define the mean level of observed human capital in an economy s as

$$H_s^o \equiv \mathrm{E}\left[h_i^o|s\right].$$

Similarly, H_s is defined as the mean of h_i for all workers in economy s, and w_s^o is defined as the estimated fixed effect, uncorrected for unobserved human capital:

$$\widehat{w}_s^o \equiv \widehat{w}_s + \mathrm{E}\left[\underline{h}_i|s\right]. \tag{26}$$

Estimating equation (25) using standard techniques yields an unbiased estimate of $\widehat{w}_s^o$, but not of $\widehat{w}_s$, because $\underline{h}_i$, and therefore $\mathrm{E}[\underline{h}_i|s]$, is unobserved. As a result, $\widehat{w}_s^o$ includes the private return to unobserved human capital. Using the observed mean level of human capital H_s^o as an estimate of total human capital H_s requires us to correct for the omitted effect of unobserved human capital. This correction simultaneously yields an adjusted estimate of the regional fixed w_s. Both corrections are essential for accurately estimating our model.

We use the single-index assumption Teulings (1995), which states that workers' human capital can be meaningfully summarized in a single index $h_i = h_i^o + \underline{h}_i$, as in equation (25). The critical assumption here is not the unit coefficient on the components h_i^o and $\underline{h}_i$[10] or the linearity of those components,[11] but their additivity, implying that observed and unobserved human capital are perfect substitutes: each unit of observed human capital can be replaced by one unit of unobserved human capital at a fixed rate of transformation. As a logical consequence of this assumption, a region that selects workers with high human capital h_i does so randomly across both the observed and unobserved components. This reasoning underpins the following Proportionality Assumption.

Proportionality Assumption

[10]Deviations could be eliminated by simply redefining the unit of measurement for h_i.

[11]Consider an alternative index h_i^* that enters as $h_i = h\left(h_i^*\right) + \underline{h}_i$, where $h\left(\cdot\right)$ is an increasing function. In this case, we can replace $h\left(h_i^*\right)$ with $h_i^o = h\left(h_i^*\right)$.

When a worker has human capital $h_i = h_i^o + \underline{h}_i$, the index h_i serves as a sufficient statistic for the expected values of its observed and unobserved components. In particular, the following relations apply:

$$\begin{aligned} \mathrm{E}\left[h_i^o | h_i, s\right] &= R_h^2 h_i \\ \mathrm{E}\left[\underline{h}_i | h_i, s\right] &= \left(1 - R_h^2\right) h_i, \end{aligned} \tag{27}$$

where

$$R_h^2 \equiv \frac{\mathrm{Var}\left[h_i^o\right]}{\mathrm{Var}\left[h_i^o\right] + \mathrm{Var}\left[\underline{h}_i\right]}, \tag{28}$$

where Var$[h_i^o]$ and Var$[\underline{h}_i]$ are the variance of h_i^o and $\underline{h}_i$ in the nationwide populations.

This assumption is a natural extension of the idea that observed and unobserved human capital are perfect substitutes, making the decomposition of h_i into its components irrelevant for workers' overall human capital. Taking expectations over h_i for economy s in equation (27) yields a simple expression for H_s and $\widehat{w}_s$ as functions of H_s^o and $\widehat{w}_s^o$:

$$\begin{aligned} H_s &= R_h^{-2} H_s^o, \\ \mathrm{E}\left[\underline{h}_i | s\right] &= \frac{1 - R_h^2}{R_h^2} H_s^o, \\ \widehat{w}_s &= \widehat{w}_s^o - \frac{1 - R_h^2}{R_h^2} H_s^o, \end{aligned} \tag{29}$$

where we apply equation (26) in the final step.

We use equation (25) to estimate the parameter vector β, the region fixed effects $\widehat{w}_s^o$, and observed human capital, both at the individual level h_i^o and as a regional mean H_s^o. The composition of the vector x_i and the estimation results for the parameter vector β are provided in the Appendix.

Three examples illustrate the distribution of the index h_i^o. The 10th percentile is -0.426; a typical worker in this group is a Black married female with 10 years of education and 26 years of work experience. The median value is -0.012, corresponding to a white married male with 12 years of education and 8 years of experience. The 90th percentile is 0.429, corresponding to a white married female with 18 years of education and 21 years of experience.

The intra-regional variance in log wages can be decomposed into three orthogonal components: observed and unobserved human capital, market imperfections (e.g., due to search frictions), and measurement error in log wages. Bound and Krueger (1991) report that measurement error accounts for 30% of the variance in log wages. Gottfries and Teulings (2023) find that market imperfections account for 8%. Therefore, human capital h_i explains the remaining $R_w^2 = 1 - 0.38 = 0.62$ of the variance in observed log wages w_i. Since we observe the variance of the observed component of human capital, the variance of unobserved human capital can be taken as the residual. Applying equation (29), we obtain

$$
\begin{aligned}
&R_h^2 = 0.60, \\
&H_s = 1.67 H_s^o, \\
&\widehat{w}_s = \widehat{w}_s^o - 0.67 H_s^o.
\end{aligned}
$$

Log land expenditure for a worker with human capital $h_i = 0$, denoted v_s, can be calculated as the log of house prices in region s minus H_s.

We use these estimates of H_s, w_s, and v_s to estimate the model in Section 3.3. Their reliability depends on the Proportionality Assumption, which can be tested empirically in several ways. Single index models are a special case of Rosen (1974) hedonic pricing model (see, e.g., Sattinger (1975); Teulings (1995); Teulings (2005); Card and Lemieux (1996); Gabaix and Landier (2008)). In these models, workers with high human capital have a comparative advantage in complex jobs and are, in a market equilibrium, assigned accordingly. Since wages increase with human capital, workers in complex jobs have both higher human capital and higher wages. Thus, job complexity can serve as an instrument for human capital.

We use three-digit occupational dummies as proxies for job complexity. In a first-stage regression, we regress w_i on these occupational dummies and construct an occupational complexity index, $o_i = \beta_o' d_i$, representing the explained portion of this regression. Letting R_o^2 be the R^2 of this regression, and after adjusting for measurement error in w_i, it turns out that R_o^2 is exactly equal to R_h^2, both at 0.60. This index can serve as an instrument for both the observed and unobserved components of human capital.

We use log wages minus observed human capital as a proxy for unobserved human capital:

$w_i - h_i^o = \underline{h}_i + \varepsilon_i$ (unobserved human capital plus measurement error in wages; see equation (25)). We therefore regress both h_i^o (observed human capital) and $w_i - h_i^o$ (unobserved human capital) on the occupational complexity index o_i. If the Proportionality Assumption holds perfectly, R^2 of the regression using h_i^o should be $R_o^2 = 0.60$, while the R^2 of the regression $w_i - h_i^o$ should be $R_w^2 R_o^2 = 0.25$.

We can extend this idea by estimating separate β_o-vectors for h_i^o and $w_i - h_i^o$, denoted as $\widehat{\beta}_o$ and $\underline{\beta}_o$, respectively. If h_i^o and $\underline{h}_i$ capture different aspects of human capital that are imperfect substitutes—such as intellectual ability and social intelligence in a double index model of occupational production—then these β_o-vectors span a two- rather than one-dimensional space. Different combinations of h_i^o and $\underline{h}_i$ can make a worker suitable for different occupations, even if their total h_i is the same.

Given that the three-digit occupational classification includes over 300 categories, it can be expected to span this two-dimensional space. Therefore, the linear combination that correlates most strongly with the observed component h_i^o should differ from the linear combination that best correlates with the unobserved component $\underline{h}_i$ in a double index model.

The results for both tests are reported in Table I. Clearly, the data do not perfectly pass these tests, as the R^2 statistic are lower than predicted, and the correlation between $\underline{\beta}_o' d_i$ and $\widehat{\beta}_o' d_i$, though well above 0.50, is less than one. Nonetheless, the single index model provides a reasonable description of the data. The correction of H_s^o and $\widehat{w}_s^o$ for unobserved human capital can be considered a limiting case where observed and unobserved human capital are perfect substitutes. To the extent that these components are imperfect substitutes, the Proportionality Assumption likely overestimates the correlation between them, and therefore underestimates the correlation between the region fixed effect and human capital H_s. Consequently, the inferred inter-economy variation of H_s can be viewed as an upper bound for the true variation, while the inferred estimates of human capital externalities w_r are a lower bound for the true human capital externalities.

Table II summarizes our results for the mean regional values of H_s over the sample period (see Table A.3 in the Appendix for H_s values for all regions), normalizing the overall mean of H_s for cities to zero, as specified in equation (11). Consistent with this normalization, the mean of H_s for the city sample is zero. On average, rural areas have lower values of H_s than cities. Accordingly, the top three regions are cities (Boston, San Jose, and San Francisco), while the bottom three are rural areas (Louisiana, Mississippi, and Georgia). The average difference between these groups over the observation period is 0.200. Since the human capital index H_s is scaled such that its private return is normalized to unity, this implies that log wages are 0.200 higher in the top three regions compared to the bottom three based on private returns alone. As we shall see, public returns further increase the wage differential between these groups. Note that the overall increase in H_s during the period covered by our data exceeds this inter-regional difference. Furthermore, the gap between the top and bottom three regions has widened over time.

3.3 Bartik instruments for regional human capital

The model treats average human capital and January temperature as exogenous characteristics of economy s. While most economists would likely accept the exogeneity of January temperature for our purposes, they may be more skeptical about the exogeneity of a region's average human capital. One would expect this human capital to respond to various changes, such as improved amenities and increased housing supply, most of which are endogenous factors. We therefore need an instrument to capture the exogenous variation in regional human capital. One promising candidate is the regional industrial structure, which tends to be highly persistent over time Amior and Manning (2018). Nation wide differences in industrial structure provides exogenous variation in average human capital per economy s.

We construct two instruments. The first, the Between (B) instrument $\widetilde{H}_{rt}^{B}$, captures nationwide variation in employment growth rates across industries over the observation period. The second, the Within (W) instrument $\widetilde{H}_{rt}^{W}$, captures nationwide variation in the growth rates of the human capital requirements within each industry over the same period,

$$\widetilde{H}_{rt}^{B} = \sum_{j} \frac{E_{rj}}{E_r} g_{(-r)tj} H_{(-r)j}, \quad \widetilde{H}_{rt}^{W} = \sum_{j} \frac{E_{rj}}{E_r} H_{(-r)j} g_{(-r)tj}^{H}, \tag{30}$$

where E_{rtj} and H_{rtj} denote employment and the mean level of human capital in region r and time t for industry j, where the mean of a variable over one of the indexes is indicated by omitting that index (e.g., $E_{rj} = \mathrm{T}^{-1}\Sigma_t E_{rtj}$), where the subscript $(-r)$ refers to the nationwide mean excluding region r; where $g_{(-r)tj}$ represents the cumulative nationwide growth of employment in industry j at time t ($E_{(-r)tj}$) compared to its mean over the sample period ($E_{(-r)j}$) relative to the growth of total employment relative to its mean over the sample period ($E_{(-r)t}/E_{(-r)}$); and where $g^H_{(-r)tj}$ is the cumulative nationwide growth in the human capital requirement:

$$g_{(-r)tj} = \frac{E_{(-r)tj}}{E_{(-r)j}} - \frac{E_{(-r)t}}{E_{(-r)}}, \quad g^H_{(-r)tj} = \frac{H_{(-r)tj}}{H_{(-r)j}} - 1.$$

The B- and W-instruments rely on independent variation in nationwide industry shocks at time t, either in employment across industries $g_{(-r)tj}$ with differing levels of human capital requirements, or in human capital requirements within industries $g^H_{(-r)tj}$. The local employment shares E_{rj}/E_r and the mean level of human capital requirement $H_{(-r)j}$ are not required to be exogenous (see Borusyak et al. (2022)).

3.4 Estimation results

Equation (24) implies that log wages and log land expenditure for economy s satisfy

$$x_s = \beta_{xz0} + \beta_{xzH}\widetilde{H}_s + \beta_{xzT}T_r + \varepsilon_{xs}, \tag{31}$$

where $x \in \{w, v\}$ denotes log wages or log land expenditure, $z \in \{r, c\}$ refers to rural areas or cities, and ε_{xs} is an error term. Equation (24) includes an expression for log land expenditure in cities near the CBD, $v^c_s(0)$, while our data may be interpreted instead as reflecting the log of average land expenditure v^c_s, which is lower due to the negative effect of commuting costs on land prices in the outskirts of the city (see equation (17)). The data fit the model well when we apply equation (24) for $v^c_s(0)$, but poorly when using equation (17) for v^c_s. One possible explanation is that the city's boundary as implied by the model extends beyond the geographical limits defined in the data collection. Regardless of the reason, we use the expression for $v^c_s(0)$ from equation (24) to interpret the estimation results.

Table III presents estimates of equation (31) using the Bartik instruments $\widetilde{H}^W_s$ and $\widetilde{H}^B_s$, including both region and time fixed effects. Note that T_r varies across regions but remains constant over time, which is why we write T_r instead of T_s. As a result, T_r is absorbed by the region fixed effects.

Our model implies that the comparative advantage of regions in specific industries is entirely driven by agglomeration benefits, and that land is not a required input for production. Aside from January temperature, other location-specific factors are assumed to be irrelevant. While this assumption simplifies our theoretical analysis, it is at odds with the data for certain industries, particularly mining and agriculture. We therefore present estimates both in- and excluding these two industries. In addition, we report results using both the W- and B-instruments together and using each instrument individually, resulting in three combinations. In total, this yields six sets of regression results.

We have most confidence in the results using only the W-instrument and excluding mining and agriculture (panel E). Our concern with the B-instrument stems from its reliance on the interaction between inter-industry differences in employment growth and the average education level in each industry, which means it may capture regional differences in overall employment growth rather than in education level. The W-instrument does not suffer from this ambiguity.

Given our long time series ($T = 37$), the error terms may be correlated over time. As our focus is on long-term effects, we have opted to ignore transitional dynamics. However, to account for the intertemporal correlation that such dynamics might introduce, we report both the standard OLS t-values (in parentheses) and Driscoll-Kraay t-values (in square brackets), which adjust for potential correlation in the residuals over time.

The first-stage regression results are consistent across all six panels. The difference between the OLS and Driscoll-Kraay t-values is minimal. The instruments are strong in every panel. A unit coefficient on an instrument implies that the entire variation in education levels from other regions with similar industry compositions is reflected in the regional education level. The observed coefficients are remarkably close to one. Including or excluding mining and agriculture has little effect on these results.

Similar to the first-stage regressions, the second-stage results are consistent across all six panels. Here, the OLS and Driscoll-Kraay t-values differ substantially. As the latter remain reliable even when residuals are correlated, the subsequent discussion focuses on Driscoll-Kraay t-values. Across the six panels, the estimates for the coefficient β_{vcH} (log land expenditure in cities) are in the interval $(4.5, 8.5)$ and are highly significant, while the coefficients β_{wcH} (log wages in cities) are in the interval $(0.6, 0.85)$ and are weakly significant. The coefficients for rural areas are mostly positive but statistically insignificant. These findings align with the model's predictions that all coefficients β_{xzH} should be positive and that β_{vcH} is significantly larger than the other three coefficients, since both λ_w and Δ_H exceed one. However, the estimates for β_{wrH} are not sufficiently reliable to test the model's prediction that $\beta_{wcH} > \beta_{wrH}$.

Panels A and D report the Sargan overidentification test applied within and between instruments, both with and without agriculture and mining. Both instruments pass the test in some, but not all, cases. Various authors have noted that Bartik instruments based on industry variation can be decomposed into time series and industry dimensions (Goldsmith-Pinkham et al. (2020); Borusyak et al. (2022)). This allows $\widetilde{H}_s^W$ and $\widetilde{H}_s^B$ to be broken down into a potentially large set of instruments, enabling overidentification tests with greater power. Building on this idea, we decompose the standard Bartik instrument in equation (30) into separate instruments for each industry j. Equation (30) implicitly assumes that the instruments for all industries share the same coefficient, which would be reasonable in the absence of measurement errors. However, this assumption is unrealistic. The coefficients can be expected to vary across industries, and indeed, an F-test strongly rejects the equality of coefficients across industries in the first-stage regression. As a result, the Sargan test does not apply to the standard Bartik instrument with a single coefficient for all industries. Instead, we use the flexible instrument allowing for separate coefficients across industries in the first stage, and subsequently apply the Sargan overidentification test in the second stage on the instruments for each industry.[12]

This test strongly rejects the exogeneity of the instruments, even when we exclude industries other than agriculture and mining where land is expected to be an important input.[13] In the Appendix tables, we present regressions for individual two-digit industries, revealing an erratic pattern of coefficients in the second-stage regressions. We conclude that our Bartik instruments effectively capture demand shocks for human capital and measure an average impact on knowledge spillovers. However, these spillovers vary across industries, in line with the analysis in Desmet and Rossi-Hansberg (2014). Consequently, industry shocks are not exogenous: the coefficient in the second-stage regression depends on which industry experiences the shock.

[12]If the flexible Bartik instrument is valid, using the standard Bartik instrument and testing overidentifying restrictions on the coefficients for each industry will lead to rejection of the instrument, since it yields the same inequality in coefficients as found in the first-stage regression.

[13]See Desmet and Rossi-Hansberg (2014), who argue that manufacturing was prone to knowledge spillovers during the introduction of electrical power in the 1920s, which led to its agglomeration in cities. However, by the 1960s, as the potential of electrical power was fully realized, manufacturing began to leave the city, making room for the rise of professional services in the 1980s.

While these results are consistent with the structural model, the imprecision of the estimates for rural areas prevents us from making precise statements about the values of the structural parameters. One way to address this issue is to exploit the cross-sectional variation in the data by excluding region fixed effects. This approach also enables the estimation of the coefficients β_{xzT} for January temperature. However, it requires the additional assumption that the employment shares E_{rj}/E_r of each industry in region r are exogenous, which is debatable. Effectively, this estimation procedure replaces the actual mean level of human capital in region r at time t with its predicted value, based on the region's mean industry composition and nationwide shocks to human capital requirements in these regions. As in Table III, the results are very similar across the six panels. We therefore focus our discussion on the preferred Panel E (see Table IV). Results for the other panels are available in the online Appendix.

The results are qualitatively similar to those in Table IV, but with much greater precision. Again, the first-stage regression in column (1) confirms that the instrument is strong. In the second-stage regressions, all coefficients are highly significant except for β_{vrH}. Focusing on the two coefficients that remain significant even when region fixed effects are included, the estimate of β_{vcH} without region fixed effects (6.9) lies at the midpoint of the range obtained with fixed effects $(4.5, 8.5)$. This is encouraging for the validity of our estimates without region fixed effects. The point estimate of β_{wcH} is nearly twice as large without region fixed effects (1.3) compared to the range with fixed effects $(0.6, 0.85)$, though it still falls within the confidence interval reported in Panel E of Table III. As predicted by the model, all four coefficients β_{xzT} are significantly positive, with β_{vcT} being the largest.

Since the specification reported in Table IV includes separate time fixed effects for each of the four regressions for w_r^r, w_r^c, v_r^r, and v_r^c, the intercepts may vary between years. The intercepts reported in Table IV refer to the midpoint of our observation period, 1998.[14] The choice of this reference year matters, because these intercepts are used to estimate Δ_0. As the intercepts vary between years, so does Δ_0. In fact, the intercept of the equation for w_r^r and v_r^r trend up, and to lesser extent, that also applies for w_r^c, while the intercept for v_r^c trends down. We address this issue at the end of the next subsection.

[14]The actual midpoint is 1997. We use 1998 instead, as the data on house prices in cities deviate markedly between 1992 and 1997 from the rest of our observation period.

3.5 Reduced-form parameters and overidentifying restrictions

Equation (31) contains 12 coefficients β_{xzQ} for $x \in \{w, v\}$, $z \in \{r, c\}$, $Q \in \{H, T, 0\}$. The intercepts β_{xz0} have no structural interpretation on their own, as they depend on the units of measurement for wages and land expenditure. Only the differences between the intercepts for cities and rural areas carry structural meaning, defined as $\beta_{x\Delta 0} \equiv \beta_{xc0} - \beta_{xr0}$. Equation (24) specifies how these coefficients relate to seven reduced-form parameters: $\omega_H, \omega_T, \lambda_w, \lambda_T, \Delta_H, \Delta_T$, and Δ_0. For now, we ignore λ_T, which implies that β_{vrT} and β_{vcT} do not have a structural interpretation either; only the difference between them does: $\beta_{v\Delta T}$. This leaves nine coefficients ($4 \times \beta_{xzH}, 2 \times \beta_{wzT}, \beta_{v\Delta T}, 2 \times \beta_{x\Delta 0}$) to identify six reduced-form parameters ($\omega_H, \omega_T, \lambda_w, \Delta_H, \Delta_T$ and Δ_0). As a result, there are $9 - 6 = 3$ testable overidentifying restrictions.

First, we analyze the relationship between the coefficients β_{xzH} and the reduced-form parameters ω_H, Δ_H, and λ_w:

$$\begin{aligned}
&\beta_{wrH} = \omega_H, \;\; \beta_{wcH} = \Delta_H \omega_H, \;\; \beta_{vrH} = \lambda_w \omega_H, \;\; \beta_{vcH} = \Delta_H \lambda_w \omega_H, \\
&\omega_H > 0, \;\; \lambda_w > 1, \;\; \Delta_H > 1 \Rightarrow 0 < \beta_{wrH} < \{\beta_{wcH}, \beta_{vrH}\} < \beta_{vcH}, \\
&\frac{\beta_{wcH}\beta_{vrH}}{\beta_{wrH}\beta_{vcH}} = 1.
\end{aligned} \tag{32}$$

The second line lists the inequality restrictions, while the third line specifies the first overidentifying restriction. The estimators $\widehat{\beta}_{xzH}$ for these coefficients are mutually uncorrelated, as they are estimated using separate sets of observations: for cities and rural areas, and for wages and land expenditure.[15] This observation is used to derive the distribution of the test statistic; see the Appendix. The overidentifying restriction implies that there are two independent estimates for each of the three structural parameters. For example, both $\widehat{\beta}_{wrH}$ and $\widehat{\beta}_{wcH}\widehat{\beta}_{vrH}/\widehat{\beta}_{vcH}$ are valid and uncorrelated estimators of ω_H. The optimal estimator $\widehat{\omega}_H$ is a weighted geometric mean of the two, with weights based on their relative standard errors.

Next, we consider the relationship between the coefficients β_{xzT} and the reduced-form param-

[15]Strictly speaking, this statement is not entirely accurate. First, to the extent that the error terms ε_{ws} and ε_{vs} do not reflect measurement error in the data on w_s and v_s, they may be correlated. Second, the instrument $\widetilde{H}_s$ is subject to measurement error, which does not affect the consistency of $\widehat{\beta}_{xzH}$, but does contribute to the error term. We ignore this potential correlation.

eters ω_T and Δ_T:

$$\beta_{wrT} = \omega_T, \;\; \beta_{wcT} = \Delta_T \omega_T, \;\; \beta_{vrT} = \lambda_w \omega_T + \lambda_T, \;\; \beta_{vcT} = \Delta_T \lambda_w \omega_T + \lambda_T, \tag{33}$$
$$0 < \beta_{wrT} < \{\beta_{wcT}, \beta_{vrT}\} < \beta_{vcT},$$
$$\lambda_w \frac{\beta_{w\Delta T}}{\beta_{v\Delta T}} = 1,$$
$$\omega_T = \beta_{wrT}, \;\; \omega_T = \beta_{wcT}/\Delta_T, \;\; \Delta_T = \frac{\beta_{wcT}}{\beta_{wrT}}.$$

Again, the second line lists the inequality restrictions, while the third line specifies the second overidentifying restriction, using the estimator $\widehat{\lambda}_w$ from equation (32).[16] The fourth line specifies two expressions for ω_T. Again, the optimal estimator $\widehat{\omega}_T$ is a weighted geometric mean of two estimators based on the expressions.

Finally, the city dummies $\beta_{x\Delta 0}$ relate to the reduced-form parameter Δ_0:

$$\beta_{w\Delta 0} = \Delta_0, \quad \beta_{v\Delta 0} = \lambda_w \Delta_0, \tag{34}$$
$$0 < \beta_{w\Delta 0} < \beta_{v\Delta 0},$$
$$\lambda_w \frac{\beta_{w\Delta 0}}{\beta_{v\Delta 0}} = 1.$$

The first line gives two expressions for Δ_0. Once again, the optimal estimator $\widehat{\Delta}_0$ is a weighted geometric mean of the two. The relation in the third line represents the third overidentifying restriction, using the estimator $\widehat{\lambda}_w$ from equation (32). The expressions for these estimators, their standard errors, and the standard errors of the test statistics are provided in the Appendix.

The point estimates for β_{xzH}, β_{xzT}, and $\beta_{x\Delta 0}$ listed in Table IV confirm all predictions specified in the second line of equations (32), (33), and (34): (i) they are all positive; (ii) $\beta_{vzH} > \beta_{wzH}$ and $\beta_{vzT} > \beta_{wzT}$; (iii) $\beta_{xcH} > \beta_{xrH}$ and $\beta_{xcT} > \beta_{xrT}$; and (iv) $\beta_{v\Delta 0} > \beta_{w\Delta 0}$. Table V reports the t-values for the tests of the three overidentifying restrictions listed in equations (32), (33), and (34), along with the point estimates and corresponding t-values for the reduced-form parameters $\lambda_w, \omega_H, \omega_T, \Delta_H, \Delta_T$, and Δ_0.

The overidentifying restriction based on the coefficients β_{xzH} for the effect of human capital is

[16] It is possible to derive more efficient estimators for the semi-structural parameters ω_H, λ_w, and Δ_H by using information from all $\widehat{\beta}_{xzQ}$ rather than just from $\widehat{\beta}_{xzH}$. We prefer to use the less efficient estimator as it offers clearer insight into the model's fit.

supported by the data. At face value, the restriction based on $\beta_{x\Delta 0}$ is heavily rejected. However, this is largely due to the incredible precision of the estimates of the coefficient β_{xz0}. The ratio $\lambda_w \frac{\beta_{w\Delta 0}}{\beta_{v\Delta 0}}$ is estimated to be 0.82, which is reasonably close to unity. The restriction based on $\beta_{x\Delta T}$ is clearly rejected. The estimation results also confirm that $\Delta_H > \Delta_T$; see equation (39). With the exception of the overidentifying restrictions on the coefficients $\beta_{x\Delta T}$, the model fits the data well. We return to the issue of the coefficients β_{xzT} for January temperature when discussing Table VIII.

Table V provides direct estimates of the land expenditure and city multipliers, $\widehat{\Delta}_H = 5.49$ and $\widehat{\lambda}_w = 5.04$ respectively. As expected, both multipliers are larger than one, in both cases substantially larger. The combined effect of these parameters multiplied by the public return to human capital in rural areas, $\widehat{\omega}_H = 0.247$, implies that one percent higher wages in a city due higher human capital drives up house prices in that city by $\widehat{\omega}_H \widehat{\lambda}_w \widehat{\Delta}_H + 1 = 7.8\%$ (+1 accounts for the income effect of the private return). This magnitude of this effect is largely due to the market clearing condition for interregional labour mobility, which drives up land prices in regions with high wages more than proportionally.

The effect of H_s on w_s^r, w_s^c, v_s^r, and v_s^c might be different between cross sectional and time series variation. Knowledge spillovers are generated at the knowledge frontier. When the stock of knowledge increases over time, generating further knowledge requires a higher level of human capital as a starting point. Regional knowledge spillovers at time t might therefore depend on its mean level of human capital H_{rt} compare to its nation wide average $\overline{H}_t$.

If so, the increase in the average level of human capital by 0.275 over the period 1979-2015 might not yield a corresponding increase in knowledge spillovers. If regional knowledge spillovers are proportional to a region's absolute level of human capital H_{rt}, the increase in average human capital $\overline{H}_t$ should have driven up the difference in real wages and in particular land prices between cities and rural areas, since cities are more sensitive to knowledge spillovers. This is not true, however, if these spillovers are proportional to a region's human capital compared to its nation wide mean, $H_{rt} - \overline{H}_t$.

Our framework allows testing this hypothesis empirically, using the variation in the intercepts β_{xz0} between years. We estimate the following equation by OLS (adding a subscript t to β_{xz0} to account for its variation between years):

$$\beta_{xz0t} = \beta_{xz0} - \mu \overline{H}_{xzt} + \varepsilon_{xzt},$$

$$\overline{H}_{wrt} \equiv \overline{H}_t, \qquad \overline{H}_{wct} \equiv \Delta_H \overline{H}_t,, \qquad \overline{H}_{vrt} \equiv \lambda_w \overline{H}_t,, \qquad \overline{H}_{vct} \equiv \lambda_w \Delta_H \overline{H}_t.$$

In fact, $\overline{H}_{xzt} = \beta_{xzH} \overline{H}_t$, see equation (33); we use the estimates of the reduced form parameters reported in Table V for this regression. If $\mu = 0$, knowledge spillovers depend on the absolute level of human capital H_{rt}, while if $\mu = 1$, knowledge spillovers depend on the level of human capital compared to its nation wide mean, $H_{rt} - \overline{H}_t$. We find $\mu = 0.87$ (t-value: 17.0). For a large part, the knowledge spillovers depend on a region's human capital relative to other regions. Even then, a share $\overline{\mu} = 0.13$ of the increase in human capital persists as a factor driving up the importance of knowledge spillovers over the course of the period 1979-2015. We shall refer to $\Delta H_r - \mu \Delta \overline{H}$ as the public return relevant part of the increase in regional human capital. Note the full increase ΔH_r is relevant for the private return.

3.6 Structural parameters

Table VI provides an overview of the number of estimable coefficients, reduced-form parameters, and structural parameters for three versions of the model. The simplest version makes no distinction between rural areas and cities and excludes January temperature. The second version introduces the distinction between rural areas and cities, while the third also incorporates January temperature. The simplest model includes just two estimable coefficients, which exactly identify two reduced-form parameters but underidentify the five structural parameters. Full identification can only be achieved by relying on exogenous sources for structural parameters, as discussed below.

Introducing the distinction between rural areas and cities adds three additional reduced-form parameters and one structural parameter, along with four new estimable coefficients. As the total number of estimable coefficients ($2 + 4 = 6$) exceeds that of reduced-form parameters ($2 + 3 = 5$) by one, this version of the model contains one overidentifying restriction on the reduced-form parameters. The estimable coefficients and structural parameters are equal in number.[17]

[17]This suggests that the model is exactly identified, which is incorrect: λ, η, and α_z identify just two reduced-form

The further extension of the model to include January temperature introduces just two additional reduced-form parameters and one structural parameter, while adding four estimable coefficients. Together, these two extensions therefore make a considerable contribution to the model's empirical content.

Table VII provides an overview of the values of the structural parameters that both align with external sources and offer the best fit for the model for rural areas. The first two rows present the preferred values of these structural parameters. The following rows display the corresponding implied values of the reduced-form parameters. When available, these values are compared to their estimated counterparts from Table V, shown in the final column.

The agglomeration elasticity ψ is set at 3%. A broad range of empirical studies (e.g., Ciccone and Hall (1996); Gennaioli et al. (2013); Ahlfeldt et al. (2015)) find values of about 7%. Since we find a substantially higher effect for higher educated, we need not be surprised that the average effect is smaller. The land share in consumption λ is set at 10%, reflecting an average housing share in consumption of 30% times a land share in housing of 33%, in line with Ahlfeldt et al. (2015) and Duranton and Puga (2023). The remaining parameters are calibrated to match as closely as possible the estimates for the reduced form parameters.

Column 2 lists the formulas derived in Section 2 for calculating the reduced-form parameters shown in column 1, derived from Proposition 1, 2, and 3. The elasticity of substitution between land and other consumption, η, plays a crucial role here. Many studies assume a Cobb-Douglas utility function with $\eta = 1$ for analytical convenience (e.g., Lucas and Rossi-Hansberg (2002); Ahlfeldt et al. (2015)). This assumption is highly restrictive, as demonstrated in column 3, which presents the formulas for the special case $\eta = 1$. The row on the land-expenditure multiplier λ_w shows that this assumption is clearly rejected by the data, as it would imply that λ_w is equal to one, whereas its empirical value is much higher. We therefore conclude that η must be less than one.

Given the above, we have some flexibility in choosing the exact value of η, since a higher value of the parameter α_z for the endogenous consumption amenity has a similar effect as a lower value of η. The heterogeneity in lot sizes between rural areas and cities, and within cities between locations near the CBD and in the outskirts, shows that there is scope for substitution between

parameters, meaning this part of the model is underidentified, while the remaining parameters are overidentified.

land and other consumption. Beyond this observation, the reduced-form parameters offer little guidance as to the precise value of η.

Conditional on the choice of $\eta = 0.60$, the share α_z of the consumption amenity in the value of private consumption must be 1.7%. Our results thus support the presence of agglomeration spillovers in consumption, consistent with a wide body of evidence (Ahlfeldt et al. (2015); Jacobs (2016); Gautier et al. (2010); de Groot et al. (2015); Teulings et al. (2018); Diamond (2016); Almagro and Domínguez-Iino (2024)). The spillovers on the consumption and the production sides are mutually reinforcing: by raising land prices in high-density areas, they compel workers in regions with consumption spillovers to economize on land use. This allows for higher population density, which in turn strengthens knowledge spillovers on the production side. Without consumption spillovers, it would be difficult for the model to generate the land-expenditure multiplier of $\lambda_w = 5.04$ observed in the data.

The parameter χ, which captures the effect of human capital on the strength of agglomeration externalities, can be derived from the parameter ω_H, representing the public return to human capital in rural areas. The implied value of χ is 6.7, comparable to the value of 6.25 reported by Gennaioli et al. (2013).[18]

Table VIII outlines the implications for the reduced-form parameters for cities and January temperature. The parameter ρ, which represents the log ratio of the travel cost of ideas to the commuting cost, thus determining the efficiency of cities relative to rural areas, must align with parameter for the city wage premium for $H_r = 0$, $\Delta_0 = w_r^r - w_r^c$; see equation (24). This requires us, in turn, to set $\rho = 3.915$, higher than Ahlfeldt et al. (2015) estimate of $\delta/\kappa = 5$, which would imply $\rho = \ln 5 - 1/2 \ln 2 = 1.26$; see equation (18). The discrepancy between our concept of δ/κ and that of Ahlfeldt et al. (2015) can be attributed to the highly stylized nature of our spatial model, including the assumptions that production does not require land, all jobs are located in the CBD, and there is no job agglomeration in rural areas. We therefore take this deviation with a grain of salt.

The values of Δ_T and Δ_H are implied by other parameters: Δ_0, ψ, ψ_1, and ψ_2. For Δ_T, the predicted and estimated values are reasonably in line. The predicted value of Δ_H, however, is

[18]Their preferred value of θ is 7.25, which refers to the multiplier for the number of workers. Our χ, by contrast, refers to the multiplier for the total wage bill, with corresponds to their $\theta - 1$.

65% higher than its estimate. Moreover, it turns out that its predicted value is not very sensitive to the precise value of the underlying parameters.

Given the value of $\Delta_0 = 0.10$, our estimation results are roughly consistent with a share α_T of a pleasant climate in private consumption of 10%. This value of α_T yields predicted values for β_{wrT}, β_{wcT}, and β_{vcT} that are lower than their estimated values, while the predicted value of β_{vrt} is higher than its estimate. Where the estimated coefficients on January temperature follow the predicted qualitative pattern $0 < \beta_{wrT} < \beta_{wcT} < \beta_{vcT}$, they do not align with the prediction that $\beta_{vrT} > \beta_{wrT}$. For some reason, land prices in rural areas are less sensitive to January temperature than predicted by the model. A higher value of α_T would improve the fit for the estimated coefficients β_{wrT}, β_{wcT}, and β_{vcT}, but would worsen the fit for β_{vrT}.

Figure I plots the mean level of human capital H_r on the x-axis and the log wage fixed effects $\widehat{w}_r^r$ (red dots) and $\widehat{w}_r^c$ (blue circles) for 1998, on the y-axis. The shown values of w_r^z are corrected for the effect of January temperature using the coefficient β_{wzT}. Also shown are w_r^r and w_r^c as functions of H_r, with January temperature set at its mean value across cities, $T_r = 0$. Both the non-linear functions (solid lines) and their linear approximations (dotted lines) are shown, using the structural parameters from Tables VII and VIII. The difference in the slopes of the functions for rural areas and cities, captured by the city multiplier $\Delta_H > 1$, is clearly visible in both the functions and the data.

The figure shows that the model explains a substantial portion of the inter-regional variation in $\widehat{w}_r$. For some cities, the average level of human capital falls below the lower bound required for a city to exist: $T_r = 0$, $H_r < H^{\min} = -0.015$ (see Table VIII). In some cases, such as Los Angeles ($H_r = -0.076$) and Riverside ($H_r = -0.078$), their favorable climate allows them to overcome this constraint, as a higher January temperature reduces $H^{\min}$. For other cities, H_r remains below the theoretical lower bound.

3.7 Choice of spatial structure

Our model treats a region's spatial structure, be it rural or urban, as an exogenous characteristic. Clearly, the spatial structure of a city is essentially irreversible. While cities may decline, once a region becomes a city it cannot revert to a rural area. But what drives a rural area to become a city?

Since $\Delta_H > 1$, cities hold a comparative advantage in human capital–intensive activities. Regions with a high level of H_r can therefore be expected to transition to the city form.

However, the causality may also run in the opposite direction: activities requiring high human capital can be expected to cluster in cities because these activities benefit most from the high knowledge spillovers fostered by the spatial structure of a monocentric CBD. Indeed, Desmet and Rossi-Hansberg (2014) documented how manufacturing agglomerated in cities during the electricity revolution in the early 20$^{\text{th}}$ century to take advantage of new ideas on using electricity to streamline production. By the second half of that century, city locations became a disadvantage as these innovations lost their edge and no longer outweighed the high land prices; thus, manufacturing moved out. In turn, business services moved into cities from 1980 onward, as the IT revolution created new opportunities for innovation fueled by knowledge spillovers.

Whether high human capital activities cause rural areas to become cities, or whether the causation runs the other way, is immaterial to the empirical outcome: cities host industries that are more human capital intensive. To demonstrate this, Table IX presents Probit model estimates for the probability that a region is a city. Both mean human capital and January temperature have a strong positive effect. In the Appendix, we provide a full list of cities and rural areas with their mean levels of human capital. Of the 30 regions with the highest H_r, only three are classified as rural areas (Connecticut, New Hampshire, and Massachusetts). Similarly, of the 30 regions with the lowest H_r, only three are classified as cities (Buffalo, Riverside, and Miami).

3.8 Counterfactuals

We use our model to perform several counterfactual calculations. Recall that the model specifies an exogenous outside utility $u^*(h)$ and an infinitely elastic labor supply. This framework could be extended by introducing a finitely elastic labor supply at the national level; however, for clarity, we retain the assumption of an infinitely elastic supply at both the regional and national levels. This implies that workers always receive their outside utility, with the number of workers remaining endogenous. Absentee landlords are the residual claimants in the model, keeping the supply of inhabitable land in each region constant. As a result, the net effect on welfare can be assessed based on aggregate land rent.

For each area, we calculate the mean level of human capital H_r from the data and obtain employment figures from the census. Given that regional labor supply is infinitely elastic while the supply of land is fixed, our counterfactuals hold land use constant. In theory, we should account for the endogeneity of each city's radius S_r. In practice, we treat S_r as fixed and therefore hold constant each city's land area, $2\pi S_r^2$. Since changes in cities' land use have only a second-order effect on aggregate GDP we can, under a linear approximation, disregard them. Using the linear approximation of the model (see equation (24)) and the structural parameters from Tables VII and VIII, we compute the fixed effects for log wages, log land expenditure, and log land use (w_r, v_r, and l_r, respectively) for both rural areas and cities. The formulas used for these calculations are provided in the Appendix.

We classify regions into the same four groups as in Table II, based on their mean level of human capital: the bottom three rural areas (Louisiana, Mississippi, Georgia), other rural areas, the top three cities (Boston, San Jose, San Francisco), and other cities.

First, we evaluate the effect of the change in the regional mean education level H_{rt} on GDP between 1979 and 2015. On average across regions, human capital increases by $\overline{\Delta H} = 0.275$, see Table X. Recall that our measure of human capital normalizes its wage return to unity. Using a private return to education of 10% per additional year, this therefore corresponds to an increase in the average education level of almost three years. For the public return, we should account for the public return relevant part of the increase in human capital, which is only $\overline{\mu}\overline{\Delta H} = 0.036$. Though our estimation results show that knowledge spillovers are for 87% determined by the value of H_r relative to its nation wide mean $\overline{H}_t$, the remaining 13% imply that the role of knowledge spillovers has increased between 1979 and 2015, equivalent to a $13\% \times 0.275 = 3.6\%$ increase in knowledge spillovers relevant human capital.

We calculate labor income for each region r (line 3) by multiplying three factors: (i) employment (line 1), (ii) wages, presuming there were no agglomeration benefits[19] $e^{\omega_0 + H_r^*}$ (line 2a),[20] and (iii) the agglomeration multiplier e^{w_r} (line 2b). Unsurprisingly, the agglomeration multiplier is higher in the top three cities than in the bottom three rural areas: 1.29 versus 0.97, respectively, for

[19] We approximate the log wage in the absence of agglomeration benefits by setting $w_r = 0$; no agglomeration corresponds to $M_r = 0$ (see footnote 3).

[20] ω_0 is an intercept that captures the nominal wage in 1979 for a worker with $h_i = 0$ employed in a rural region where $H_r = T_r = 0$.

1979; the values of these multiplier are higher for 2015 due to the increase in knowledge spillovers relevant human capital by 3.6%. Since $\eta < 1$, the land share of aggregate income is increasing in the price per unit of land and hence smaller in rural areas than in cities: 11.7% in the bottom three rural areas versus 16.8% in the top three cities in 1979. Again, the land share is higher in 2015 due to the increasing importance of knowledge spillovers.

Next, we calculate the return to human capital as a percentage of regional income. Let us first consider wage income. The private return per person is equal to the average change in H_r weighted by nominal income, denoted $\overline{\Delta H}$ (line 4a). The public return is $\omega_H \left(\overline{\Delta H} - \mu \Delta \overline{H}\right)$ for rural areas and $\Delta_H \omega_H \left(\overline{\Delta H} - \mu \Delta \overline{H}\right)$ for cities (line 4b). Finally, we account for the change in land use per person. For the rental income, lines 4a and 4b reflect the change in land expenditure per person. Line 4a shows the effect of a higher real income resulting from the private return to human capital, $\overline{\Delta H}$.[21] Line 4b shows the offsetting effect on land prices due to the public return to human capital embodied in wages, given by $\lambda_w \omega_H \left(\overline{\Delta H} - \mu \Delta \overline{H}\right)$ for rural areas and $\lambda_w \Delta_H \omega_H \left(\overline{\Delta H} - \mu \Delta \overline{H}\right)$ for cities. Similar to wage income, line 4c captures the change in land use per person.

Summing these three lines for both wage and rental income gives the total return to human capital reported in line 4. Summing the changes in wage and rental income yields the relative change in total income in line 5, which corresponds to GDP when excluding capital income other than land rents. At the national level, a 27% private return to human capital results in a 17% increase in GDP (measured in log points, as are all subsequent figures). The increase in GDP is less than the private return because the private return induces an increase in land use that has a negative effect on agglomeration externalities. The GDP gains are heavily concentrated in the top three cities, where GDP rises by 35% and land rents by 39%.

A more appropriate comparison would involve GDP per capita. This entails ignoring line 4c, which captures changes in population resulting from changes in H_r. This adjustment is made in line 6, where we calculate the return to human capital on GDP per capita per year of education, using a private return of 10%. In the cross section, the return to GDP per capita is two and a half times the private return in the nation as a whole and even three and a half times the private return in the top three cities. This multiple comes half way the fivefold private return observed in

[21] The numbers differ slightly from those for wages, as they are weighted by rental income rather than wage income.

simple OLS regressions like those in Barro and Lee (1996) cross-country data for GDP and years of education. In the time series, the effects are much smaller.

Line 7 shows the effect of the actual increase in human capital, $\overline{\Delta H}$, on total welfare, again expressed as a percentage of GDP. Since landlords are the residual claimants in our model, this effect corresponds to the increase in rental income. Unlike the effect on aggregate income shown in line 5, the rise in wage income does not contribute to welfare because it serves as a compensating differential for higher land rents and, in cities, for commuting cost. Moreover, since labor supply is perfectly elastic, the private return to human capital represents its opportunity cost and does not increase welfare. As a result, the effect on welfare is substantially smaller than the effect on aggregate income—2.4% compared to 24% of GDP in the cross section. Using GDP as a proxy for aggregate welfare therefore leads to significant double counting. Once again, the growth in total welfare is heavily concentrated in cities.

As a final exercise, we calculate the mean fixed effect in a log wage regression for both 1979 and 2015. From these estimates, we derive the annual real wage growth for a worker with constant human capital:

$$\begin{aligned} \omega_0^{1979,2015} &= R^{-1}\Sigma_r\left(\widehat{w}_r - w_r\right), \\ \text{annual real wage growth} &= T^{-1}\left(\omega_0^{2015} - \omega_0^{1979} - \log \mathrm{CPI}_{2015-1979}\right) \\ &= 0.00. \end{aligned}$$

The full increase in real wages between 1979 and 2015 for a worker with constant human capital is accounted for by the increase in knowledge spillover relevant human capital by 3.6%.

Table XI presents the results of two additional counterfactual scenarios: one in which there are no cities (i.e., all regions are rural areas), and another with no sorting (i.e., all regions have the same level of human capital). In both cases, GDP declines by approximately 10%. These results demonstrate that the spatial organization of a city and spatial sorting in terms of human capital make substantial contributions to GDP.

4 Discussion: welfare and political economy

In the presence of agglomeration externalities—or, more broadly, increasing returns to scale—simple cause-and-effect reasoning is of limited value. Does one worker's productivity drive the other's, or is it the reverse? A similar ambiguity arises with the agglomeration of consumption and production spillovers: which of these two forces initiates the process?

This paper sheds light on these issues by combining a standard instrumental variables approach using Bartik instruments with a structural model that seeks to explain how agglomeration externalities affect both wages and house prices. Our model allows for agglomeration on both the consumption and production sides, an exogenous amenity, and two distinct spatial forms (rural areas and cities). Our findings go halfway in explaining 50% public return to an additional year of education suggested by a simple cross-country regression of GDP per capita on average years of education, finding a return of 23% in the cross section.

Our analysis requires the decomposition of regional fixed effects from a simple OLS log-wage regression into unobserved human capital and regional spillover effects of human capital. This decomposition relies on the Proportionality Assumption: that regions draw proportionally from both observed and unobserved human capital. The credibility of this assumption rests on the degree to which the two forms of human capital are close substitutes in production, as the evidence we have presented suggests. That the substitution is not quite exact implies that the Proportionality Assumption overestimates the role of unobserved human capital while underestimating the importance of knowledge spillovers and the public return to human capital. If anything, our estimates likely understate the true magnitude of agglomeration externalities.

Our Bartik instruments are strong, but do not pass overidentification tests when separate instruments are used for each industry. While our instruments may capture the average effect of an industry shock on knowledge spillovers, these effects vary substantially across industries. As a result, merely stimulating employment in industries with high demand for high-human-capital workers will not necessarily promote knowledge spillovers. However, industry shocks that increase the regional employment of such workers do on average generate such spillovers. The empirical credibility of this claim rests on the joint predictions for the effects of knowledge spillovers along three dimensions: (i) the regional fixed effects in regressions on individual log wages and log land expenditure, (ii) the difference in these effects between rural areas and cities, and (iii) the differential impact of an exogenous amenity (January temperature) on wages and land expenditure. All of these predictions are supported by the data.

Our model generates a total return to GDP per capita that is two and a half times the private return. Using a private return of 10% per year of education, this yields a 25% total return to GDP per capita, which goes a long way toward explaining the 50% return observed in a simple OLS cross-country regression of GDP per capita on average years of education. There are however two caveats. First, we show that the return is much higher in the cross section than in the time series. As knowledge frontier moves outward, the human capital requirements for contributing to new knowledge spill overs follows suit; 87% of the public return is relative to the knowledge frontier. As it moves outward, a higher level of human capital is required to keep with the Joneses. Second, the effect of knowledge spillover is considerably smaller on welfare than GDP, since the agglomeration benefit in wages is a compensating differential for higher land prices and commuting costs in cities, and the private return to human capital is a compensating differential for the opportunity cost of acquiring human capital.

Our model shows that several major trends in the U.S. economy between 1979 and 2015 are closely linked to the rising human capital of the workforce: increased urbanization, overall employment growth, and (in particular) the surge in urban house prices. It shows that the full increase in real wages can be accounted for by both the private and public returns to human capital. The rise in the value of the housing stock in France and England since 1970 documented by Piketty (2014), along with the sharp increase in house prices in successful cities, may well be explained by the growing importance of knowledge spillovers in modern capitalist economies. Our model generates a real house-price increase in the average CMSA by a factor $e^{(1+.13\lambda_w \omega_H \Delta_H)\overline{\Delta H}} \cong 1.68$, much larger than in rural areas. A substantial share of the gains from the IT and biotech revolutions accrued to real-estate investors in San Francisco, San Jose, and Boston, while standards of living in rural areas remained largely unchanged. This growing divide between rural areas and cities may help to explain the populist political schism seen in many countries; consider Gethin et al. (2022) claim that education levels have increasingly influenced voting behavior. The downward trend in real wages for workers with constant human capital predicted by our model may have contributed to the recent rise in populist political support. Regional sorting of human capital has increased over the period 1979-2015, reinforcing the pre-existing difference between rural areas and cities. This might have strengthened the populist upheaval. At the same time, we show that regional sorting of human capital and the benefits of the spatial structure of cities for knowledge spillovers each contribute 10% to GDP. Undoing this sorting to stop the populist revolution is therefore a costly policy.

The presence of agglomeration externalities implies that decentralized markets fail to achieve an efficient allocation. In our model, where workers' payoffs are determined by an infinitely elastic labor supply curve, total land rents serve as a sufficient statistic for efficiency. Below, we consider five types of externalities.

First, the choice of a region's spatial organization as either a rural area or a city is not necessarily efficient. For a region to feasibly adopt the city form, its human capital index H_r must exceed the lower bound $H^{\min}$; see equation (21). Even if this condition is met, it may still fail to develop into a city if it lacks a natural focal point for its CBD. In such cases, the rural form prevaila because workers have no incentive to commute beyond their home location, and firms have no reason to favor one site over another. Overcoming this coordination problem such that a city can emerge requires a driving force—either some pre-existing natural advantage or a visionary, charismatic entrepreneur.

Second, agglomeration externalities imply that workers' land use is too high Lucas and Rossi-Hansberg (2002). In cities, land prices reflect the benefit of living close to the CBD, where workers earn higher wages. These high land prices reduce individual land consumption, allowing more workers to live near the CBD and access its wage premium. However, land prices do not account for the positive externality that each additional worker in the CBD generates on the wages of all other workers there. This inefficiency could be addressed through a balanced-budget Henry George tax on land used for a wage subsidy, see Rossi-Hansberg and Wright (2007b). Because the infinitely elastic labor supply curve ensures that workers receive their outside option regardless, the optimal tax system should maximize a region's total land rent. An optimally designed tax system would increase the region's population and net land rent revenue without reducing those in other regions.

Third, cities impose a negative externality on surrounding rural areas by drawing workers from the urban fringe to the CBD, thereby diminishing knowledge spillovers in the adjacent rural areas. Assuming that the city employs a Henry George tax system, the city boundary will extend too far from the CBD (i.e., S_r will be too high), imposing a negative externality on the surrounding land. A related dynamic may explain Atlanta's high growth rate: the relatively low human capital in rural Georgia keeps the cost of urban expansion low, as rural land commands low prices due to limited knowledge spillovers.

Fourth, a region's human capital index H_r is shaped by the location decisions of entrepreneurs. Landowners benefit when firms with high H_r settle in their region, as these firms generate strong knowledge spillovers that drive up land prices. As a result, regions have an incentive to compete for high H_r firms, and should attract them by implementing industrial policies, financed through a Henry George land tax, that subsidize the location of such firms. Since the city form is more conducive to generating knowledge spillovers, cities should allocate more resources to industrial policy than rural areas.

The final type of externality stretches the model slightly by treating a region's transport infrastructure as an endogenous outcome. Public transport—particularly rail—involves a trade-off between flexibility and economies of scale. While train lines and stations are inflexible and entail high fixed costs, the returns to scale make rail transport systems efficient for cities. This is especially true in monocentric cities, where the CBD serves as a common destination for commuters and offers a natural home for the central station. The funding of rail infrastructure exhibits the characteristics of a public good, since incumbent users—those living next to suburban stations—should subsidize marginal customers who live further away from these stations and who therefore require some pre-transport to enjoy the benefits of the rail system. Otherwise, they will switch to private transport, undermining the economic viability of the network. Private rail companies cannot implement the necessary price discrimination to prevent this. Some cities, such as New York, Boston, and San Francisco, have successfully addressed this public action problem; others have not.

Los Angeles presents an interesting case: its underground rail system (the second to open in the U.S. after New York's) began operation in 1925 but was shut down in 1955.[22] The city became heavily reliant on car transport, which is land intensive and limits the population density

[22] https://en.wikipedia.org/wiki/Hollywood_Subway. Public transport rankings across cities: www.businessinsider.com/cities-with-best-public-transportation-systems-2014-1?international=true&r=US&IR=T https://smartasset.com/mortgage/best-cities-for-public-transportation

needed to fully realize agglomeration benefits (i.e., a low commuting cost κ). Consequently, Los Angeles has been less attractive to high-human-capital activities, as its structure is less conducive to knowledge spillovers. This is reflected in its relatively low human capital index, $H_r = -0.044$. The internal structure of cities may therefore vary depending on their political institutions.

Appendix

A Proof of Propositions

A.1 Proof of Proposition 1

Combination of equation (1), (2), (3) and (4) yields

$$w_r = \overline{\eta}^{-1} \ln \left(\overline{\lambda} + \lambda \mathrm{e}^{\overline{\eta} p_r} \right) - \alpha' a_r \cong \lambda p_r - \alpha' a_r,$$

where the second equality follows from a Taylor approximation of the term $\overline{\eta}^{-1} \ln \left(\overline{\lambda} + \lambda \mathrm{e}^{\overline{\eta} p_r} \right)$ around $p_r = 0$; the symbol $\cong$ refers to a Taylor approximation. By Shephard's Lemma, the partial derivative of the cost function with respect to the price of land is equal to the demand for land. Hence, the partial derivative of the log cost function with respect to the log price of land is equal to the land share in expenditure.[23] Combining these results yields:

$$\ln \frac{\partial \left[\overline{\eta}^{-1} \ln \left(\overline{\lambda} + \lambda \mathrm{e}^{\overline{\eta} p_r} \right) \right]}{\partial p_r} = \ln \lambda + \overline{\eta} p_r - \ln \left(\overline{\lambda} + \lambda \mathrm{e}^{\overline{\eta} p_r} \right) = l_r + p_r - w_r,$$

where the final expression on the right hand side is the land share in private expenditure (which is the same for all h due to homotheticity). A Taylor approximation of the term $\ln \left(\overline{\lambda} + \lambda \mathrm{e}^{\overline{\eta} p_r} \right)$ and rearranging terms yields:

$$l_r \cong \ln \lambda - \overline{\eta}_\lambda p_r + w_r,$$

$$\eta_\lambda \equiv \overline{\lambda} \overline{\eta} \Rightarrow 0 \leq \eta_\lambda < 1.$$

Combining both results yields:

$$l_r \cong \ln \lambda - \frac{\overline{\eta}_\lambda - \lambda}{\lambda} w_r - \frac{\overline{\eta}_\lambda}{\lambda} \alpha' a_r.$$

[23] Using upper cases for the exponents of the corresponding lower cases, let $E(P_r)$ be expenditure as a function of the price of land P_r for a fixed price of the composite commodity and utility level. Shephard's lemma implies

$$E'(P_r) = L_r,$$

$$e'(p_r) = \frac{E'(P_r)}{E(P_r)} \frac{\partial P_r}{\partial p_r} = \frac{L \cdot P_r}{E(P_r)}.$$

The final expression is the land share in total expenditure.

The density of income per unit of land, i.e. log density of purchasing power per unit of land, satisfies:

$$z_r = w_r - l_r.$$

Since both income and land use are proportional to human capital, H_r drops out. We obtain:

$$\begin{aligned} z_r &\cong \frac{-\lambda \ln \lambda + \overline{\eta}_\lambda w_r + \overline{\eta}_\lambda \alpha_T T_r}{\lambda - \overline{\eta}_\lambda \alpha_z}, \\ l_r &\cong \frac{\lambda \ln \lambda - \left(\overline{\eta}_\lambda + \overline{\eta}_\lambda \alpha_z - \lambda\right) w_r - \overline{\eta}_\lambda \alpha_T T_r}{\lambda - \overline{\eta}_\lambda \alpha_z}, \qquad (35) \\ p_r &\cong \frac{-\alpha_z \ln \lambda + w_r + \alpha_T T_r}{\lambda - \overline{\eta}_\lambda \alpha_z}, \qquad (36) \end{aligned}$$

where $\lambda - \overline{\eta}_\lambda \alpha_z > 0$ due to assumption (7). Using equation (35) and (36) and $v_r = l_r + p_r$, we obtain equation (8); we also know that $\lambda_z > 1$, because for $\alpha_z = 0$, $\lambda_z = \overline{\eta}_\lambda / \lambda > 1$ since $\overline{\eta}_\lambda = 1 - \eta_\lambda = \lambda + \eta (1 - \lambda) > \lambda$, and for $\alpha_z > 0$, λ_z is even larger.

A.2 Proof of Proposition 2

The proof for n_r^c, m_r^c and v_r^c follows from the transform of variable and the Taylor approximation in equation (19), substitution of equation (16) for κS_r and from $\overline{\lambda}_z + \lambda_z / \overline{\eta}_\lambda = \lambda_w$ in equation (17). w_r^c satisfies:

$$\begin{aligned} w_r^c &\cong \psi \left(1 + \chi H_r\right) \left(\chi_0 + \ln \pi + 2 \ln \Delta_r - \frac{2}{3} \lambda_z \Delta_r - 2 \ln \kappa - l_r\right) - \psi \qquad (37) \\ &= \psi \left(1 + \chi H_r\right) \left(1 + 2 \ln \left(\delta / \kappa\right) - \ln 2 + 2 \ln \Delta_r - \frac{2}{3} \lambda_z \Delta_r - \overline{\lambda}_z w_r^c + \lambda_z \alpha_T T_r\right) - \psi \\ &= \psi \left(1 + \chi H_r\right) \left[1 + 2\rho + 2 \ln \Delta_r - \frac{2}{3} \lambda_z \Delta_r - \overline{\lambda}_z \left(\frac{1}{2} w_r^r + \Delta_r\right) + \frac{\omega_T}{\psi_1} T_r\right] - \psi \Rightarrow \\ \Delta_r &= \psi \left(1 + \chi H_r\right) \left(1 + 2\rho + 2 \ln \Delta_r - \frac{2}{3} \lambda_z \Delta_r - \overline{\lambda}_z \Delta_r + \frac{\omega_T}{\psi_1} T_r\right) - \psi \\ &\quad - \frac{1}{2} \left[1 + \overline{\lambda}_z \psi \left(1 + \chi H_r\right)\right] w_r^r \\ &= \psi \left(1 + \chi H_r\right) \left[1 + 2\rho + 2 \ln \Delta_r - \left(1 - \frac{1}{3} \lambda_z\right) \Delta_r + \frac{1}{2} \frac{\omega_T}{\psi_1} T_r\right] - \frac{1}{2} \psi \chi H_r - \psi, \\ &= \psi \frac{2 \left(1 + \frac{\omega_H}{\psi_1} H_r\right) \left(\rho + \ln \Delta_r + \frac{\omega_T}{4\psi_1} T_r\right) + \frac{\omega_H}{2\psi_1} H_r}{1 + \psi \left(1 - \frac{1}{3} \lambda_z\right) \left(1 + \frac{\omega_H}{\psi_1} H_r\right)}. \end{aligned}$$

The first line substitutes equation (17) for m_r in equation (9). The second line substitutes equation (12) for χ_0 and (8) for l_r. The third line substitutes equation (16) for w_r^c and (18) for ρ and uses $\lambda_z \alpha_T = \omega_T/\psi_1$, see equation (13). The fourth line deducts $\frac{1}{2}w_r^r$ on both sides and redefines the left hand using equation (16). The fifth line substitutes equation (12) for $\left[1+\overline{\lambda}_z \psi\left(1+\chi H_r\right)\right] w_r^r$. The final line brings the term $\psi\left(1+\chi H_r\right)\left(1-\frac{1}{3}\lambda_z\right)\Delta_r$ to the left hand side and uses $\chi = \omega_H/\psi_1$, see equation (13)).

A Taylor approximation yields:

$$\frac{\psi\left(1+\frac{\omega_H}{\psi_1}H_r\right)}{1+\psi\left(1-\frac{1}{3}\lambda_z\right)\left(1+\frac{\omega_H}{\psi_1}H_r\right)} = \psi_2\left(1+\frac{\psi_2\omega_H}{\psi_1\psi}H_r\right)+\mathcal{O}\left(H_r^2\right),$$
$$\frac{\omega_H H_r + 2\psi_1}{1+\frac{\omega_H}{\psi_1}H_r} = 2\psi_1 - \omega_H H_r + \mathcal{O}\left(H_r^2\right).$$

Assumption (14) implies $1+\psi\left(1-\frac{1}{3}\lambda_z\right)>0$. Hence $\psi_2>0$. Substitution in equation (37) yields equation (18).

A solution for X to the equation $X-\ln X=\alpha$ exists iff $\alpha \geq 1$. Defining

$$X \equiv \frac{\Delta_r}{2\psi_2\left(1+\frac{\psi_2}{\psi_1\psi}\omega_H H_r\right)}$$

implies that equation (18) can be written as:

$$X-\ln X = \rho + \frac{1}{4\psi_1}\left(\omega_T T_r + \omega_H H_r\right) + \ln\psi_2 + \ln\left(1+\frac{\psi_2}{\psi_1\psi}\omega_H H_r\right) > 1-\ln 2, \tag{38}$$

proving equation (20). The relevant solution for X is larger than one. Hence $\Delta_r > 2\psi_2\left(1+\frac{\psi_2}{\psi_1\psi}\omega_H H_r\right)$, proving equation (17). Derivatives of the right hand side with respect to H_r and T_r are positive.

A.3 Proof of Proposition 3

Existence of Δ_0 ($\equiv \Delta_r$ for $H_r = T_r = 0$) requires condition (38) to hold for $H_r = T_r = 0$, which is guaranteed by equation (23). Since $X \equiv \frac{\Delta_0}{2\psi_2} \geq 1$, $\Delta_0 \geq 2\psi_2$. Equation (17) for Δ_r implies:

$$\rho + \ln\Delta_0 = \frac{\Delta_0}{2\psi_2}.$$

Differentiating equation (17) for Δ_r with respect to Δ_r, H_r, and T_r in the point $H_r = T_r = 0$ reads:

$$\frac{\Delta_0 - 2\psi_2}{\Delta_0}\mathrm{d}\Delta_r = \frac{\psi_2}{2\psi_1}\left[(\rho + \ln\Delta_0)\,\frac{4\psi_2}{\psi}\omega_H \mathrm{d}H_r + \omega_T \mathrm{d}T_r + \omega_H \mathrm{d}H_r\right] \text{ for } H_r = T_r = 0 \Rightarrow$$
$$\frac{\mathrm{d}\Delta_r}{\mathrm{d}T_r}_{|H_r=T_r=0} = \frac{1}{2}\frac{\psi_2}{\psi_1}\frac{\Delta_0}{\Delta_0 - 2\psi_2}\omega_T > 0,$$
$$\frac{\mathrm{d}\Delta_r}{\mathrm{d}H_r}_{|H_r=T_r=0} = \frac{1}{2}\frac{\psi_2}{\psi_1}\frac{\Delta_0}{\Delta_0 - 2\psi_2}\frac{2\Delta_0 + \psi}{\psi}\omega_H > \frac{\mathrm{d}\Delta_r}{\mathrm{d}T_r} > 0.$$

The inequalities follow from $\Delta_0 - 2\psi_2 \geq 0$ and $4\psi_2 > \psi$ (see the final line of the proof of Proposition 2). If assumption (23) applies, $\mathrm{d}\Delta_r/\mathrm{d}H_r > \mathrm{d}\Delta_r/\mathrm{d}H_r > \frac{1}{2}$.

A first-order Taylor approximation of Δ_r reads:

$$\Delta_r \cong \Delta_0 + \Delta_H^o \omega_H H_r + \Delta_T^o \omega_T T_r, \tag{39}$$
$$\Delta_0 \equiv 2\psi_2\,(\rho + \ln\Delta_0) > 2\psi_2,$$
$$\Delta_H^o \equiv \Delta_T^o \frac{2\Delta_0 + \psi}{\psi} > \Delta_T^o, \qquad \Delta_T^o \equiv \frac{1}{2}\frac{\psi_2}{\psi_1}\frac{\Delta_0}{\Delta_0 - 2\psi_2} > 0.$$

where $\Delta_{H,T} \equiv \Delta_{H,T}^o + \frac{1}{2} > 1$.

The relations for w_r^r and v_r^r in equation (24) replicate equation (13). The relations form w_r^c and $v_r^c\,(0)$ follow from equation (16) and substitution of the Taylor approximation for Δ_r.

A.4 Supplementary material: population of cities

Substitution of equation (8) for l_r and equation (24) for w_r^c in equation (17) yields:

$$\begin{aligned} n_r^c &= \ln\pi - 2\ln\kappa + 2\ln\Delta_r + \frac{2}{3}\overline{\lambda}_z\Delta_r - \lambda_l - \overline{\lambda}_z w_r^c + \lambda_z\alpha_T T_r - H_r \\ &= \ln\pi - 2\ln\kappa - \lambda_l + 2\ln\Delta_r + \frac{2}{3}\overline{\lambda}_z\Delta_r - \overline{\lambda}_z\left(\frac{1}{2}w_r^r + \Delta_r\right) + \frac{\omega_T}{\psi_1}T_r - H_r \\ &= \ln\pi - 2\ln\kappa - \lambda_l + 2\ln\Delta_r - \frac{1}{3}\overline{\lambda}_z\Delta_r - \left(1 + \frac{1}{2}\overline{\lambda}_z\omega_H\right)H_r + \left(\psi_1^{-1} - \frac{1}{2}\overline{\lambda}_z\right)\omega_T T_r, \end{aligned}$$

A Taylor approximation with respect to Δ_r in the point $H_r = T_r = 0$ yields:

$$\begin{aligned}
n_r^c &= \varepsilon_0 + \left(\frac{2}{\Delta_0} - \frac{1}{3}\overline{\lambda}_z\right)\left(\Delta_H^o \omega_H H_r + \Delta_T^o \omega_T T_r\right) - \left(\frac{\Delta_H^o \omega_H H_r + \Delta_T^o \omega_T T_r}{\Delta_0}\right)^2 \\
&\quad - \left(1 + \frac{1}{2}\overline{\lambda}_z \omega_H\right) H_r + \left(\psi_1^{-1} - \frac{1}{2}\overline{\lambda}_z\right)\omega_T T_r + \mathcal{O}\left(H_r^3\right) + \mathcal{O}\left(T_r^2\right) + \mathcal{O}\left(H_r T_r\right) \\
&= \varepsilon_0 + \varepsilon_H H_r + \varepsilon_T T_r + \varepsilon_{HH} H_r^2 + \mathcal{O}\left(H_r^3\right) + \mathcal{O}\left(T_r^2\right) + \mathcal{O}\left(H_r T_r\right) \\
\varepsilon_0 &\equiv \ln \pi - 2 \ln \kappa - \lambda_l + 2 \ln \Delta_0 - \frac{1}{3}\overline{\lambda}_z \Delta_0, \\
\varepsilon_H &= \left(\frac{2}{\Delta_0} - \frac{1}{6}\overline{\lambda}_z\right)\Delta_H - \frac{1}{2}\overline{\lambda}_z > 0, \\
\varepsilon_T &= \varepsilon_H + \psi_1^{-1} > \varepsilon_H, \\
\varepsilon_{HH} &= -\left(\frac{\Delta_H}{\Delta_0}\right)^2 < 0.
\end{aligned}$$

The factor $\frac{2}{\Delta_0} - \frac{1}{6}\overline{\lambda}_z$ is the derivative of the Taylor approximation of the log Gamma function $\ln\left(\int_0^{S_r} s e^{\overline{\lambda}_z \kappa s} \mathrm{d}s\right)$, see equation (19). The derivative of this function is always positive. Since this approximation is highly precise for the relevant value of $\Delta_r = \kappa S_r$, this factor is always positive for the relevant values of Δ_0 and $\overline{\lambda}_z$. Would we have use the exact function, the derivative would have been positive anyway.

B Distribution of the reduced-form parameters

Define $\widehat{\Delta}_1$:

$$\widehat{\Delta}_1 \equiv \ln \widehat{\beta}_{wrH} + \ln \widehat{\beta}_{vcH} - \ln \widehat{\beta}_{wcH} - \ln \widehat{\beta}_{vrH}.$$

Equation (32) implies $\mathrm{p}\lim \widehat{\Delta}_1 = 0$. The four estimators $\widehat{\beta}_{xzH}$ are uncorrelated since they are estimated on separate sets of observations.[24] Using the indepence of $\widehat{\beta}_{xzH}$ and $\mathrm{Var}\left(\ln \widehat{\beta}_{xzH}\right) \cong \mathrm{Var}\left(\widehat{\beta}_{xzH}\right) \widehat{\beta}_{xzH}^{-2} = \mathrm{t}_{xzH}^{-2}$, the variance V_1 of $\widehat{\Delta}_1$ reads

$$\mathrm{V}_1 = \mathrm{t}_{vcH}^{-2} + \mathrm{t}_{wcH}^{-2} + \mathrm{t}_{vrH}^{-2} + \mathrm{t}_{wrH}^{-2} \Rightarrow \mathrm{t}_{\Delta_1} = \widehat{\Delta}_1 / \sqrt{\mathrm{V}_1}.$$

[24]Strictly speaking, this statement is not fully correct for two reasons. First, to the extent that ε_{ws} and ε_{vs} do not reflect measurement error in the data on w_s and v_s, both error terms might be correlated. Second, the instrument $\widetilde{H}_s$ is perturbed by measurement error which does not affect the consistencey of $\widehat{\beta}_{xzH}$, but feeds into the error term ε_{xs}. We ignore this correlation.

The estimators $\widehat{\omega}_H$, $\widehat{\omega}_1$ and $\widehat{\lambda}_w$ satisfy

$$
\begin{aligned}
\ln \widehat{\omega}_H &= \mathrm{V}_1^{-1}\left[\mathrm{t}_{wrH}^{-2}\left(\ln\widehat{\beta}_{wcH}+\ln\widehat{\beta}_{vrH}-\ln\widehat{\beta}_{vcH}\right)+\left(\mathrm{t}_{wcH}^{-2}+\mathrm{t}_{vrH}^{-2}+\mathrm{t}_{vcH}^{-2}\right)\ln\widehat{\beta}_{wrH}\right],\\
\mathrm{t}_{\omega_H}^{-2} &= \mathrm{V}_1^{-1}\mathrm{t}_{wrH}^{-2}\left(\mathrm{t}_{wcH}^{-2}+\mathrm{t}_{vrH}^{-2}+\mathrm{t}_{vcH}^{-2}\right),\\
\ln \widehat{\Delta}_H^{+} &= \mathrm{V}_1^{-1}\left[\left(\mathrm{t}_{vcH}^{-2}+\mathrm{t}_{vrH}^{-2}\right)\left(\ln\widehat{\beta}_{wcH}-\ln\widehat{\beta}_{wrH}\right)+\left(\mathrm{t}_{wrH}^{-2}+\mathrm{t}_{wcH}^{-2}\right)\left(\ln\widehat{\beta}_{vcH}-\ln\widehat{\beta}_{vrH}\right)\right],\\
\mathrm{t}_{\Delta_H}^{-2} &= \mathrm{V}_1^{-1}\left(\mathrm{t}_{wcH}^{-2}+\mathrm{t}_{wrH}^{-2}\right)\left(\mathrm{t}_{vrH}^{-2}+\mathrm{t}_{vcH}^{-2}\right),\\
\ln \widehat{\lambda}_w &= \mathrm{V}_1^{-1}\left[\left(\mathrm{t}_{wcH}^{-2}+\mathrm{t}_{vcH}^{-2}\right)\left(\ln\widehat{\beta}_{vrH}-\ln\widehat{\beta}_{wrH}\right)+\left(\mathrm{t}_{wrH}^{-2}+\mathrm{t}_{vrH}^{-2}\right)\left(\ln\widehat{\beta}_{vcH}-\ln\widehat{\beta}_{wcH}\right)\right],\\
\mathrm{t}_{\lambda_w}^{-2} &= \mathrm{V}_1^{-1}\left(\mathrm{t}_{wcH}^{-2}+\mathrm{t}_{vcH}^{-2}\right)\left(\mathrm{t}_{wrH}^{-2}+\mathrm{t}_{vrH}^{-2}\right).
\end{aligned}
$$

We observe that $\widehat{\beta}_{xcQ}$ and $\widehat{\beta}_{xrQ}$ for $Q\in\{T,0\}$ are uncorrelated since they are estimated on separate sets of observations. Hence, $\mathrm{t}_{x\Delta Q}^{-2}$ satisfies

$$
\mathrm{t}_{x\Delta Q}^{-2}=\widehat{\beta}_{x\Delta Q}^{-2}\left(\mathrm{t}_{xrQ}^{-2}\widehat{\beta}_{xcQ}^{2}+\mathrm{t}_{xcQ}^{-2}\widehat{\beta}_{xrQ}^{2}\right).
$$

Define $\widehat{\Delta}_2$:

$$
\widehat{\Delta}_2=\ln\widehat{\beta}_{v\Delta T}-\ln\widehat{\lambda}_w-\ln\widehat{\beta}_{w\Delta T},\qquad \mathrm{p}\lim\widehat{\Delta}_2=0,
$$

Assuming these estimators to be uncorrelated to $\widehat{\lambda}_w$,[25] the variance V_2 of $\widehat{\Delta}_2$ reads

$$
\mathrm{V}_2=\mathrm{t}_{\lambda_w}^{-2}+\mathrm{t}_{w\Delta T}^{-2}+\mathrm{t}_{v\Delta T}^{-2}\Rightarrow \mathrm{t}_{\Delta_2}=\widehat{\Delta}_2/\sqrt{\mathrm{V}_2}.
$$

Similarly:

$$
\begin{aligned}
&\widehat{\Delta}_3=\ln\widehat{\Delta}_T^{+}+\ln\widehat{\beta}_{wrT}-\ln\widehat{\beta}_{wcT},\qquad \mathrm{p}\lim\widehat{\Delta}_3=0,\\
&\mathrm{V}_3=\mathrm{t}_{\Delta_T}^{-2}+\mathrm{t}_{wrT}^{-2}+\mathrm{t}_{vcT}^{-2}\Rightarrow \mathrm{t}_{\Delta_3}=\widehat{\Delta}_3/\sqrt{\mathrm{V}_3},\\
&\widehat{\Delta}_4=\ln\widehat{\lambda}_w+\ln\widehat{\beta}_{w\Delta 0}-\ln\widehat{\beta}_{v\Delta 0},\qquad \mathrm{p}\lim\widehat{\Delta}_4=0,\\
&\mathrm{V}_4=\mathrm{t}_{\lambda_w}^{-2}+\mathrm{t}_{w\Delta 0}^{-2}+\mathrm{t}_{v\Delta 0}^{-2}\Rightarrow \mathrm{t}_{\Delta_4}=\widehat{\Delta}_4/\sqrt{\mathrm{V}_4}.
\end{aligned}
$$

[25]This is the case of H_s and T_r are uncorrelated.

The estimators $\ln \widehat{\omega}_T$ and $\ln \widehat{\omega}_0$ and their t-values read:

$$\ln \widehat{\omega}_T = \frac{\left(\mathrm{t}_{wcT}^{-2} + \mathrm{t}_{\Delta_T}^{-2}\right) \ln \widehat{\beta}_{wrT} + \mathrm{t}_{wrT}^{-2} \left(\ln \widehat{\beta}_{wcT} - \ln \widehat{\Delta}_T^{+}\right)}{\mathrm{t}_{wcT}^{-2} + \mathrm{t}_{\Delta_T}^{-2} + \mathrm{t}_{wrT}^{-2}},$$

$$\mathrm{Var}\left(\ln \widehat{\omega}_T\right) = \frac{\mathrm{t}_{wcT}^{-2}(\mathrm{t}_{\Delta_T}^{-2} + \mathrm{t}_{wrT}^{-2})}{\mathrm{t}_{wcT}^{-2} + \mathrm{t}_{\Delta_T}^{-2} + \mathrm{t}_{wrT}^{-2}},$$

$$\ln \widehat{\Delta}_0 = \frac{(\mathrm{t}_{v\Delta 0}^{-2} + t_{\lambda_\omega}^{-2}) \ln \widehat{\beta}_{w\Delta 0} + \mathrm{t}_{w\Delta 0}^{-2} \left(\ln \widehat{\beta}_{v\Delta 0} - \ln \widehat{\lambda}_w\right)}{\mathrm{t}_{v\Delta 0}^{-2} + t_{\lambda_\omega}^{-2} + \mathrm{t}_{w\Delta 0}^{-2}},$$

$$\mathrm{Var}\left(\ln \widehat{\Delta}_0\right) = \frac{(\mathrm{t}_{v\Delta 0}^{-2} + t_{\lambda_\omega}^{-2})\mathrm{t}_{w\Delta 0}^{-2}}{\mathrm{t}_{v\Delta 0}^{-2} + t_{\lambda_\omega}^{-2} + \mathrm{t}_{w\Delta 0}^{-2}}.$$

C Counterfactuals

The annual log wage for a worker with $h_i = 0$ in 1000\$ in a region with $H_r = T_r = 0$, accounting for 2000 hours worked reads

$$\omega_0 = N^{-1}\Sigma_r\left(\widehat{w}_r - w_r\right) + \ln 2000 - \ln 1000 = N^{-1}\Sigma_r\left(\widehat{w}_r - w_r\right) + \ln 2.$$

Let:

n_r: data on log employment derived from the Census for 1979 (for states with CMSAs deducted)

$\mathcal{R}$: a set of regions

$N_{\mathcal{R}}$: number of regions in $\mathcal{R}$

D_{wH}: 1 for rural regions, Δ_H for city regions

D_{vH}: 1 for rural regions, $\frac{1}{3}\left(\Delta_H + 1\right)$ for city regions

Define:

$$H_r^* = H_r - \mu\overline{H}_t, \quad \Delta H_r^* = \Delta H_r - \mu\Delta\overline{H}_t$$

$$W_r \equiv e^{\omega_0 + H_r^* + w_r + n_r}, \quad V_r \equiv e^{\omega_0 + H_r^* + v_r + n_r},$$

$$W_{\mathcal{R}} \equiv \Sigma_{r\in\mathcal{R}} W_r, \quad V_{\mathcal{R}} \equiv \Sigma_{r\in\mathcal{R}} V_r,$$

$$\overline{\Delta H}_{w\mathcal{R}} \equiv W_{\mathcal{R}}^{-1}\Sigma_{r\in\mathcal{R}} W_r \Delta H_r, \quad \overline{\Delta H}_{v\mathcal{R}} \equiv V_{\mathcal{R}}^{-1}\Sigma_{r\in\mathcal{R}} V_r \Delta H_r$$

$$\overline{\Delta H}^*_{w\mathcal{R}} \equiv W_{\mathcal{R}}^{-1}\Sigma_{r\in\mathcal{R}} W_r \Delta H_r^*, \quad \overline{\Delta H}^*_{v\mathcal{R}} \equiv V_{\mathcal{R}}^{-1}\Sigma_{r\in\mathcal{R}} V_r \Delta H_r^*.$$

where w_r is either w_r^r or w_r^c and the same for v_r.

$W_{\mathcal{R}}$ and $V_{\mathcal{R}}$ use the actual w_r and v_r

$W_{\mathcal{R}}^r$ uses the actual w_r^r also for city regions

$W_{\mathcal{R}}^H$ and $V_{\mathcal{R}}^H$ substitute H_r for $\overline{H}$ in the expressions for w_r and v_r.

Using those notations, the formula for the two counterfactual tables are in Table A.9 and Table A.10

References

Acemoglu, D. and J. Angrist (2000). How large are human-capital externalities? evidence from compulsory schooling laws. *NBER Macroeconomics Annual 15*, 9–59.

Acemoglu, D., F. A. Gallego, and J. A. Robinson (2014). Institutions, human capital, and development. *Annual Review of Economics 6*(1), 875–912.

Ahlfeldt, G. M., S. J. Redding, D. M. Sturm, and N. Wolf (2015). The economics of density: Evidence from the berlin wall. *Econometrica 83*(6), 2127–2189.

Albouy, D., G. Ehrlich, and L. Yingyi (2016). Housing demand, cost-of-living inequality, and the affordability crisis. Working Paper No. w22816, NBER.

Almagro, M. and T. Domínguez-Iino (2024). Location sorting and endogenous amenities: Evidence from amsterdam. Technical Report w32304, National Bureau of Economic Research.

Amior, M. and A. Manning (2018). The Persistence of Local Joblessness. *American Economic Review 108*(7), 1942–70.

Angrist, J. D. and A. B. Krueger (1999). Empirical strategies in labor economics. In O. Ashenfelter and D. Card (Eds.), *Handbook of Labor Economics*, Volume 3 of *Handbook of Labor Economics*, Chapter 23, pp. 1277–1366. Elsevier.

Autor, D. H. and D. Dorn (2013). The growth of low-skill service jobs and the polarization of the us labor market. *American Economic Review 103*(5), 1553–97.

Autor, D. H., F. Levy, and R. J. Murnane (2003). The skill content of recent technological change: An empirical exploration. *The Quarterly Journal of Economics 118*(4), 1279–1333.

Bairoch, P. (1988). *Cities and economic development: From the dawn of history to the present*. University of Chicago Press.

Barro, R. J. and J.-W. Lee (1996). International measures of schooling years and schooling quality. *The American Economic Review 86*(2), 218–223.

Bernanke, B. S. and K. Rogoff (2001). *NBER Macroeconomics Annual 2000, Volume 15*. MIT Press.

Bils, M. and P. J. Klenow (2000). Does schooling cause growth? *American Economic Review 90*(5), 1160–1183.

Borusyak, K., P. Hull, and X. Jaravel (2022). Quasi-experimental shift-share research designs. *The Review of Economic Studies 89*(1), 181–213.

Borusyak, K., P. Hull, and X. Jaravel (2024). A practical guide to shift-share instruments. Technical Report w33236, National Bureau of Economic Research.

Bound, J. and A. B. Krueger (1991). The extent of measurement error in longitudinal earnings data: Do two wrongs make a right? *Journal of Labor Economics 9*(1), 1–24.

Caliendo, L., F. Parro, E. Rossi-Hansberg, and P.-D. Sarte (2017). The Impact of Regional and Sectoral Productivity Changes on the U.S. Economy. *The Review of Economic Studies 85*(4), 2042–2096.

Card, D. and T. Lemieux (1996). Wage dispersion, returns to skill, and black-white wage differentials. *Journal of Econometrics 74*(2), 319–361.

Ciccone, A. and R. E. Hall (1996). Productivity and the density of economic activity. *The American Economic Review 86*(1), 54–70.

Comin, D. A., D. Lashkari, and M. Mestieri (2015). Structural change with long-run income and price effects. Working Paper 21595, National Bureau of Economic Research.

Davis, M. A. and F. Ortalo-Magne (2011). Household expenditures, wages, rents. *Review of Economic Dynamics 14*(2), 248 – 261.

de Groot, H. L., G. Marlet, C. Teulings, and W. Vermeulen (2015). *Cities and the urban land premium*. Edward Elgar Publishing.

Desmet, K. and J. Rappaport (2017). The settlement of the united states, 1800–2000: The long transition towards gibrat's law. *Journal of Urban Economics 98*, 50–68.

Desmet, K. and E. Rossi-Hansberg (2014). Spatial development. *American Economic Review 104*(4), 1211–43.

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

Duranton, G. and D. Puga (2023). Urban growth and its aggregate implications. *Econometrica 91*(6), 2219–2259.

Emanuel, N., E. Harrington, and A. Pallais (2023). The power of proximity to coworkers: Training for tomorrow or productivity today? Technical Report w31880, National Bureau of Economic Research.

Frey, W. H. (2012). Population growth in metro america since 1980. Technical report, The Brookings Institution.

Gabaix, X. and A. Landier (2008). Why has ceo pay increased so much? *The Quarterly Journal of Economics 123*(1), 49–100.

Gagliardi, L., E. Moretti, and M. Serafinelli (2023). The world's rust belts: The heterogeneous effects of deindustrialization on 1,993 cities in six countries. Working Paper w31948, National Bureau of Economic Research.

Gautier, P. A., M. Svarer, and C. N. Teulings (2010). Marriage and the city: Search frictions and sorting of singles. *Journal of Urban Economics 67*(2), 206–218.

Gautier, P. A. and C. N. Teulings (2006). How large are search frictions? *Journal of the European Economic Association 4*(6), 1193–1225.

Gautier, P. A. and C. N. Teulings (2015). Sorting And The Output Loss Due To Search Frictions. *Journal of the European Economic Association 13*(6), 1136–1166.

Gennaioli, N., R. La Porta, F. Lopez-de Silanes, and A. Shleifer (2013). Human capital and regional development. *The Quarterly Journal of Economics 128*(1), 105–164.

Gethin, A., C. Martínez-Toledano, and T. Piketty (2022). Brahmin left versus merchant right: Changing political cleavages in 21 western democracies, 1948–2020. *The Quarterly Journal of Economics 137*(1), 1–48.

Glaeser, E. L. (2005). Reinventing boston: 1630-2003. *Journal of Economic Geography 5*(2), 119–153.

Glaeser, E. L. and J. D. Gottlieb (2009). The wealth of cities: Agglomeration economies and spatial equilibrium in the united states. *Journal of Economic Literature 47*(4), 983–1028.

Glaeser, E. L. and M. E. Kahn (2001). Decentralized employment and the transformation of the american city. *Brookings-Wharton Papers on Urban Affairs 2001*, 1–63.

Glaeser, E. L., M. E. Kahn, and J. Rappaport (2008). Why do the poor live in cities? the role of public transportation. *Journal of Urban Economics 63*(1), 1 – 24.

Glaeser, E. L. and D. C. Maré (2001). Cities and skills. *Journal of Labor Economics 19*(2), 316–342.

Glaeser, E. L., A. Saiz, G. Burtless, and W. C. Strange (2004). The rise of the skilled city [with comments]. *Brookings-Wharton Papers on Urban Affairs*, 47–105.

Goldsmith-Pinkham, P., I. Sorkin, and H. Swift (2020). Bartik instruments: What, when, why, and how. *American Economic Review 110*(8), 2586–2624.

Gottfries, A. and C. Teulings (2023). Returns to on-the-job search and wage dispersion. *Labour Economics 80*, 102292.

Heblich, S., S. J. Redding, and D. M. Sturm (2020). The making of the modern metropolis: Evidence from london. *The Quarterly Journal of Economics 135*(4), 2059–2133.

Herkenhoff, K. F., L. E. Ohanian, and E. C. Prescott (2018). Tarnishing the golden and empire states: Land-use restrictions and the us economic slowdown. *Journal of Monetary Economics 93*, 89–109.

Hsieh, C.-T. and E. Moretti (2019). Housing constraints and spatial misallocation. *American Economic Journal: Macroeconomics 11*(2), 1–39.

Iacoviello, M. (2011). Housing wealth and consumption. International Finance Discussion Papers 1027, Board of Governors of the Federal Reserve System (U.S.).

Jacobs, J. (2016). *The economy of cities*. Vintage.

Lucas, R. (1988). On the mechanics of economic development. *Journal of Monetary Economics 22*(1), 3–42.

Lucas, R. E. (2009). Ideas and growth. *Economica 76*(301), 1–19.

Lucas, R. E. and E. Rossi-Hansberg (2002). On the Internal Structure of Cities. *Econometrica 70*(4), 1445–1476.

Manduca, R. (2021). The spatial structure of us metropolitan employment: New insights from administrative data. *Environment and Planning B: Urban Analytics and City Science 48*(5), 1357–1372.

Moretti, E. (2003). Human capital externalities in cities. Working Paper 9641, National Bureau of Economic Research.

Moretti, E. (2004). Estimating the social return to higher education: Evidence from longitudinal and repeated cross-sectional data. *Journal of Econometrics 121*(1–2), 175–212.

Moretti, E. (2011). Local labor markets. In *Handbook of Labor Economics*, Volume 4, pp. 1237–1313. Elsevier.

Piketty, T. (2011). On the long-run evolution of inheritance: France 1820–2050 *. *The Quarterly Journal of Economics 126*(3), 1071–1131.

Piketty, T. (2014). *Capital in the twenty-first century*. Belknap Press.

Robert E. Lucas, J. (2001). Externalities and Cities. *Review of Economic Dynamics 4*(2), 245–274.

Romer, P. M. (1990). Endogenous technological change. *Journal of Political Economy 98*(5, Part 2), S71–S102.

Rosen, S. (1974). Hedonic prices and implicit markets: Product differentiation in pure competition. *Journal of Political Economy 82*(1), 34–55.

Rossi-Hansberg, E. (2004). Optimal urban land use and zoning. *Review of Economic Dynamics 7*(1), 69–106.

Rossi-Hansberg, E. and M. L. J. Wright (2007a). Establishment size dynamics in the aggregate economy. *American Economic Review 97*(5), 1639–1666.

Rossi-Hansberg, E. and M. L. J. Wright (2007b). Urban structure and growth. *Review of Economic Studies 74*, 597–624.

Saiz, A. (2010). The geographic determinants of housing supply. *The Quarterly Journal of Economics 125*(3), 1253–1296.

Sattinger, M. (1975). Comparative advantage and the distributions of earnings and abilities. *Econometrica 43*(3), 455–68.

Teulings, C. and T. van Rens (2008). Education, growth, and income inequality. *The Review of Economics and Statistics 90*(1), 89–104.

Teulings, C. N. (1995). The Wage Distribution in a Model of the Assignment of Skills to Jobs. *Journal of Political Economy 103*(2), 280–315.

Teulings, C. N. (2005). Comparative Advantage, Relative Wages, and the Accumulation of Human Capital. *Journal of Political Economy 113*(2), 425–461.

Teulings, C. N., I. V. Ossokina, and H. L. de Groot (2018). Land use, worker heterogeneity and welfare benefits of public goods. *Journal of Urban Economics 103*, 67 – 82.

Table I: Test of the single index model

Variable	Predicted	Actual
R^2 in a regression of h_i^o on o_i	0.60	0.36
R^2 in a regression of $w_i - h_i^o$ on o_i	0.25	0.12
$\text{Cor}\left(\underline{\beta}_o' d_i, \widehat{\beta}_o{}' d_i\right)$	1	0.70

Table II: Mean values of the human capital index

	H_s			
Region / Year	average	1979	2015	2015 − 1979
Boston, San Jose, San Francisco	0.080	−0.116	0.187	0.303
Mean over all other cities	−0.008	−0.184	0.097	0.281
Mean over all other rural areas	−0.040	−0.214	0.055	0.269
Louisiana, Mississippi, Georgia	−0.120	−0.309	−0.038	0.271
Mean over all areas	−0.026	−0.202	0.073	0.275

Table III: First- and second-stage results for rural areas (non-c) and cities

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
VARIABLES	H_s	w_s	w_s	v_s	v_s	H_s	w_s	w_s	v_s	v_s
sample	full	non-c	city	non-c	city	full	non-c	city	non-c	city
Observations	2,997	1,739	1,258	1,739	1,258	2,997	1,739	1,258	1,739	1,258
	A. Both BIV					**D. Both BIV no Agr**				
Bartik IVW	1.043					1.005				
	(9.18)					(10.01)				
	[7.46]					[8.01]				
Bartik IVB	0.777					0.769				
	(8.33)					(7.94)				
	[6.13]					[4.99]				
H_s^{inst}		0.567	0.669	1.655	5.678		0.976	0.710	1.796	5.876
		(4.10)	(5.96)	(3.29)	(11.51)		(7.17)	(6.31)	(3.59)	(11.92)
		[0.68]	[1.77]	[0.68]	[3.80]		[1.12]	[1.92]	[0.62]	[3.93]
R-squared	0.938	0.989	0.990	0.928	0.940	0.938	0.989	0.990	0.928	0.940
rmse	0.0271	0.0329	0.0314	0.120	0.138	0.0271	0.0329	0.0314	0.120	0.138
OLS Sargan		25.029	0.312	6.765	3.803		0.084	0.333	20.536	6.377
DK Sargan		5.143	0.251	1.189	2.591		0.015	0.299	2.263	4.310
	B. within BIV					**E. within BIV no Agr**				
H_s^{inst}	1.215	0.0186	0.628	3.031	7.594	1.150	1.083	0.829	3.836	8.410
	(10.75)	(0.11)	(4.14)	(4.74)	(11.42)	(11.53)	(6.56)	(5.43)	(6.39)	(12.66)
	[8.03]	[0.02]	[1.71]	[0.87]	[3.88]	[8.86]	[0.93]	[2.20]	[0.99]	[4.29]
R-squared	0.936	0.989	0.989	0.928	0.939	0.937	0.989	0.989	0.929	0.940
rmse	0.0275	0.0331	0.0317	0.119	0.139	0.0274	0.0327	0.0315	0.119	0.137
	C. between BIV					**F. between BIV no Agr**				
H_s^{inst}	0.932	1.117	0.849	0.0557	4.907	0.945	0.758	0.730	-1.202	4.527
	(10.02)	(6.29)	(5.44)	(0.08)	(6.87)	(9.77)	(4.03)	(4.67)	(-1.74)	(6.33)
	[6.86]	[2.51]	[1.82]	[0.03]	[3.26]	[5.78]	[1.74]	[1.64]	[-0.75]	[3.12]
R-squared	0.936	0.989	0.989	0.926	0.934	0.936	0.989	0.989	0.926	0.934
rmse	0.0275	0.0327	0.0315	0.121	0.144	0.0275	0.0329	0.0316	0.121	0.145

Note: t-statistics are in parentheses, Driscoll-Kraay t-values in square brackets; all regressions include time and region fixed effects.

Table IV: First- and second-stage results for rural areas (non-c) and cities without fixed effects

	(1)	(2)	(3)	(4)	(5)
VARIABLES	H_s	w_s	w_s	v_s	v_s
sample	full	non-c	city	non-c	city
Observations	2,997	1,739	1,258	1,739	1,258
E. within BIV no Agr					
H_s^{inst}	1.710	0.256	1.355	0.282	6.943
	(37.87)	(4.44)	(27.26)	(1.09)	(21.94)
	[29.85]	[6.31]	[23.53]	[0.96]	[9.64]
Avg temp	-0.0296	0.0379	0.0822	0.0347	0.752
	(-14.08)	(8.66)	(16.89)	(1.76)	(24.26)
	[-10.09]	[12.61]	[11.59]	[3.24]	[11.57]
Constant	0.0620	0.0555	0.353	11.07	12.50
	(7.22)	(3.64)	(26.54)	(161.10)	(147.68)
	[7.17]	[6.67]	[30.74]	[185.97]	[87.09]
R-squared	0.834	0.960	0.971	0.604	0.651
rmse	0.0438	0.0623	0.0517	0.281	0.329

Note: t-statistics are in parentheses, Driscoll-Kraay t-values in square brackets; all regressions include time fixed effects.

Table V: Reduced-form structural parameters and specification tests

model	test	t-value	parameter	value	t-value
equation (32)	$\frac{\widehat{\beta}_{wcH}\widehat{\beta}_{vrH}}{\widehat{\beta}_{wrH}\widehat{\beta}_{vcH}} - 1$	1.45	$\widehat{\omega}_H$	0.247	6.38
			$\widehat{\Delta}_H$	5.49	6.17
			$\widehat{\lambda}_w$	5.04	8.97
equation (33)	$\widehat{\lambda}_w \frac{\widehat{\beta}_{w\Delta T}}{\widehat{\beta}_{v\Delta T}} - 1$	3.07	$\widehat{\omega}_T$	0.038	12.98
			$\widehat{\Delta}_T$	2.17	6.61
equation (34)	$\widehat{\lambda}_w \frac{\widehat{\beta}_{w\Delta 0}}{\widehat{\beta}_{v\Delta 0}} - 1$	6.59	$\widehat{\Delta}_0$	0.101	61.27

Table VI: Number of coefficients, reduced-form parameters, and structural parameters for three versions of the model

Model	Coefficients	Reduced-form parameters	Structural parameters
only rural, no Jan temp	2	2 (ω_H, λ_w)	5 $(\lambda, \eta, \psi, \chi, \alpha_z)$
adding cities	4	3 $(\Delta_0, \Delta_H, \Delta_T)$	1 (ρ)
adding Jan temp	4	2 (ω_T, λ_T)	1 (α_T)

Table VII: Structural parameter results rural areas

ψ	λ	η	α_z	α_T	χ	ρ	
0.03	0.10	0.60	0.017	0.10	6.7	3.915	

para-meter	expression		value	
	general	$\eta = 1$	pred.	est.
η_λ	$\overline{\lambda}\overline{\eta}$	0	0.36	n.a.
λ_z	$\frac{\overline{\eta}_\lambda}{\lambda - \overline{\eta}_\lambda \alpha_z}$	$\frac{1}{\lambda - \alpha_z}$	7.18	n.a.
λ_w	$1 + \frac{\eta_\lambda}{\overline{\eta}_\lambda}\lambda_z$	1	5.04	5.04
λ_l	$\frac{\lambda}{\overline{\eta}_\lambda}\lambda_z \ln \lambda$	$\frac{\lambda \ln \lambda}{\lambda - \alpha_z}$	-2.58	n.a.
ψ_1	$\frac{\psi}{1+\overline{\lambda}_z \psi}$	$\frac{(\lambda-\alpha_z)\psi}{\lambda-\alpha_z+\psi(\lambda-\alpha_z-1)}$	0.037	n.a.
ω_H	$\psi_1 \chi$	$\frac{(\lambda-\alpha_z)\psi}{\lambda-\alpha_z+\psi(\lambda-\alpha_z-1)}\chi$	0.247	0.247
λ_T	$\frac{\eta_\lambda}{\overline{\eta}_\lambda}\lambda_z \alpha_T$	0	0.404	n.a.

Table VIII: Structural parameter results - cities

para-meter	expression	value	
		pred.	est.
ψ_2	$\frac{\psi}{1+\psi\left(1-\frac{1}{3}\lambda_z\right)}$	0.031	n.a.
Δ_0	$2\psi_2\left(\rho + \ln \Delta_0\right)$	0.100	0.101
Δ_T	$\frac{1}{2}\frac{\psi_2}{\psi_1}\frac{\Delta_0}{\Delta_0 - 2\psi_2} + \frac{1}{2}$	1.64	2.17
Δ_H	$\frac{1}{2}\frac{\psi_2}{\psi_1}\frac{\Delta_0}{\Delta_0 - 2\psi_2}\frac{2\Delta_0+\psi}{\psi} + \frac{1}{2}$	9.22	5.49
$H^{\min}$	$\frac{4\psi_1}{\omega_H}\left(\overline{\rho} - \ln 2\psi_2 - \ln\left(1 + \frac{\psi_2}{\psi_1\psi}\omega_H H^{\min}\right)\right)$	-0.015	n.a.
$\Delta^{\min}$	$2\psi_2\left(1 + \frac{\psi_2}{\psi_1\psi}\omega_H H^{\min}\right)$	0.056	n.a.
β_{wrT}	$\omega_T \equiv \psi_1 \lambda_z \alpha_T$	0.026	0.038
β_{wcT}	$\Delta_T \omega_T$	0.043	0.082
β_{vrT}	$\lambda_w \omega_T + \lambda_T$	0.537	0.035
β_{vcT}	$\lambda_w \Delta_T \omega_T + \lambda_T$	0.622	0.752

Table IX: Probit model for the probability of a region being a city

VARIABLES	(1) City dummy
Avg human capital H_r	30.40
	(4.60)
Avg Jan temp T_r	0.00872
	(4.06)
Constant	-3.372
	(-4.33)
Observations	81

Note: z-statistics are in parentheses.

Table X: Counterfactuals

Region(s)		Bottom 3		Other rural		Other cities		Top 3		Nation	
Levels 1979											
1.	employment ($\times 10^{-6}$)	5.41		79.03		54.95		4.28		143.67	
	wage/rent p. person ($\times (10)^{-3}$)	wages	rents	wages	rents	wages	rents	wages	rents	wages	rents
2a.	without agglomeration	8.83	1.03	9.72	1.14	10.02	1.17	10.72	1.26	9.84	1.15
2b.	agglomeration multiplier	0.97	0.97	0.98	0.85	1.09	1.16	1.29	1.75	1.04	1.01
3.	aggregate income ($\times (10)^{-9}$)	46.59	5.45	751.17	79.18	614.61	77.64	62.48	10.48	1474.85	172.75
3a.	rent-wage share (%)	11.71		10.54		12.63		16.77		11.71	
Levels 2015											
1.	employment ($\times 10^{-6}$)	7.17		113.34		81.65		5.93		208.09	
	wage/rent p. person ($\times (10)^{-3}$)	wages	rents	wages	rents	wages	rents	wages	rents	wages	rents
2a.	without agglomeration	30.50	3.57	33.50	3.93	34.94	4.09	38.18	4.47	34.11	4.00
2b.	agglomeration multiplier	0.98	1.01	0.99	0.88	1.20	1.38	1.48	2.18	1.09	1.12
3.	aggregate income ($\times (10)^{-9}$)	214.75	25.88	3748.98	419.45	3402.77	469.37	335.40	59.74	7701.89	974.44
3a.	rent-wage share (%)	12.05		11.19		13.79		17.81		12.65	
Changes 1979-2019 (%)											
	Return to human capital										
4a.	private return	27.10	27.09	27.20	27.12	27.72	27.64	28.59	28.10	27.47	27.41
4b.	public return	0.78	3.92	0.80	3.96	8.61	15.70	10.59	17.67	4.56	10.22
4c.	change in employment density	-22.27	-22.28	-22.23	-22.27	-8.06	-8.39	-4.40	-6.44	-15.39	-14.88
4.	total return (4a.+4b.+4c.)	5.61	8.72	5.78	8.82	28.27	34.96	34.78	39.33	16.64	22.76
5.	total aggregate income	5.95		6.08		29.05		35.45		17.31	
6.	return per cap per y.o.s										
6a.	in the time series	10.46		10.45		13.24		13.31		11.78	
6b.	in the cross section	13.56		13.45		35.08		35.67		23.79	
7.	total welfare										
7a.	in the time series	1.06		0.96		3.93		5.03		2.43	
7b.	in the cross section	8.19		7.40		30.46		38.98		18.80	

Table XI: Counterfactual no cities

Nation 2015	Aggregate income ($\times (10)^{-9}$ \$)				Change (%)		
	Actual		Counterfactual				
	wages	rents	wages	rents	wages	rent	total
All regions rural, no cities	7701.89	974.44	7063.84	732.02	-8.28	-24.88	-10.15
All regions same H_r	7701.89	974.44	7047.21	838.62	-8.50	-13.94	-9.11

Table A.1: CBSA Observations Distribution Among States

CBSA	State I	State II	State III	State IV	Pct SI	Pct SII	Pct SIII	Pct SIV	NAME
31100	CA				100.00%				Los Angeles-Long Beach-Anaheim, CA
40140	CA				100.00%				Riverside-San Bernardino-Ontario, CA
41740	CA				100.00%				San Diego-Carlsbad, CA
41860	CA				100.00%				San Francisco-Oakland-Hayward, CA
41940	CA				100.00%				San Jose-Sunnyvale-Santa Clara, CA
19740	CO				100.00%				Denver-Aurora-Lakewood, CO
47900	DC	VA	MD		45.91%	25.90%	28.19%		Washington-Arlington-Alexandria, DC-VA-MD-WV
33100	FL				100.00%				Miami-Fort Lauderdale-West Palm Beach, FL
45300	FL				100.00%				Tampa-St. Petersburg-Clearwater, FL
12060	GA				100.00%				Atlanta-Sandy Springs-Roswell, GA
16980	IL	IN	WI		98.23%	1.77%	0.00%		Chicago-Naperville-Elgin, IL-IN-WI
26900	IN				100.00%				Indianapolis-Carmel-Anderson, IN
35380	LA				100.00%				New Orleans-Metairie, LA
14460	MA	NH			86.75%	13.25%			Boston-Cambridge-Newton, MA-NH
12580	MD				100.00%				Baltimore-Columbia-Towson, MD
19820	MI				100.00%				Detroit-Warren-Dearborn, MI
33460	MN	WI			99.99%	0.01%			Minneapolis-St. Paul-Bloomington, MN-WI
28140	MO	KS			45.36%	54.64%			Kansas City, MO-KS
41180	MO	IL			80.98%	19.02%			St. Louis, MO-IL
24660	NC				100.00%				Greensboro-High Point, NC
15380	NY				100.00%				Buffalo-Cheektowaga-Niagara Falls, NY
35620	NY	NJ			69.24%	30.76%			New York-Newark-Jersey City, NY-NJ
40380	NY				100.00%				Rochester, NY
17140	OH	KY			77.70%				Cincinnati, OH-KY-IN
17460	OH				100.00%				Cleveland-Elyria, OH
18140	OH				100.00%				Columbus, OH
38900	OR	WA			91.57%	8.43%			Portland-Vancouver-Hillsboro, OR-WA
37980	PA	NJ	DE	MD	62.06%	23.32%	14.62%	0.00%	Philadelphia-Camden-Wilmington, PA-NJ-DE-MD
38300	PA				100.00%				Pittsburgh, PA
19100	TX				100.00%				Dallas-Fort Worth-Arlington, TX
26420	TX				100.00%				Houston-The Woodlands-Sugar Land, TX
47260	VA				100.00%				Virginia Beach-Norfolk-Newport News, VA-NC
42660	WA				100.00%				Seattle-Tacoma-Bellevue, WA
33340	WI				100.00%				Milwaukee-Waukesha-West Allis, WI

Note: Information for 34 city areas: CBSA code in 2013, city belong to which state(s) and the percentage of sample observations in the CPS 1979-2015, name of cities. *Data sources*: the Current Population Survey MORG and the US Census Bureau.

Table A.2: Individual Mincerian Wage Regression

Variables	Coefficient	t-stat	Variables	Coefficient	t-stat
Male	0.419	(448.59)	Edu = 0	-0.652	(-98.12)
Male $\times$ Time Trend	-0.00565	(-118.75)	Edu = 1	-0.532	(-36.40)
Single	0.0516	(40.86)	Edu = 2	-0.540	(-76.20)
Male $\times$ Single	-0.283	(-172.09)	Edu = 3	-0.532	(-86.27)
Single $\times$ Time Trend	-0.00287	(-48.72)	Edu = 4	-0.452	(-73.50)
Male $\times$ Single $\times$ Time Trend	0.00449	(55.31)	Edu = 5	-0.475	(-113.51)
Divorced	0.0213	(14.06)	Edu = 6	-0.426	(-129.18)
Male $\times$ Divorced	-0.107	(-41.03)	Edu = 7	-0.357	(-115.49)
Divorced $\times$ Time Trend	-0.00175	(-0.196)	Edu = 8	-0.264	(-128.79)
Male $\times$ Divorced $\times$ Time Trend	0.00185	(15.19)	Edu = 9	-0.260	(-180.71)
South	0.00858	(2.58)	Edu = 10	-0.194	(-188.11)
Black	-0.0975	(-96.7)	Edu = 11	-0.157	(-167.93)
Black $\times$ South	-0.0382	(-28.34)	Edu = 13	0.0611	(67.86)
Other Race	-0.0834	(-69.98)	Edu = 14	0.168	(193.93)
Other $\times$ South	-0.00692	(-2.75)	Edu = 15	0.212	(136.26)
Year of Experience	0.0299	(51.92)	Edu = 16	0.420	(373.93)
Exp $\times$ Edu	0.00151	(33.78)	Edu = 17	0.405	(171.13)
Exp^2 / 100	-0.0445	(-16.72)	Edu = 18	0.576	(326.16)
Exp^2 / 100 $\times$ Edu	-0.00842	(-40.15)	Constant	1.109	(267.11)
Exp^3 / 100000	0.183	(5.25)			
Exp^3 / 100000 $\times$ Edu	0.102	(36.01)	Observations	5,426,947	
Edu in y9297	0.00463	(21.76)	R-squared	0.575	
			R-MSE	0.444	

Note: Table presents the estimated β using OLS regression. Dependent variable is the log hourly wage. Mincer wage regression includes individual characteristics x, gender, year of education, year of experience, race, marital status, and the interaction of these factors. All the regressions include time x region dummies. Robust t-statistics in parentheses.

Table A.3: Ranking of Regions in observed mean Human Capital Level

Region	HC Index	Type	Region	HC Index	Type
Boston, MA	0.139	City	North Dakota	0.005	Non-city
San Jose, CA	0.128	City	Kansas	0.003	Non-city
San Francisco, CA	0.128	City	Nebraska	-0.002	Non-city
Seattle, WA	0.122	City	Oklahoma	-0.002	Non-city
Connecticut	0.104	Non-city	Illinois	-0.004	Non-city
Portland, OR	0.103	City	Iowa	-0.005	Non-city
Denver, CO	0.101	City	Pennsylvania	-0.008	Non-city
Washington, DC	0.093	City	Arizona	-0.008	Non-city
Pittsburgh, PA	0.088	City	Utah	-0.009	Non-city
Minneapolis, MN	0.084	City	Idaho	-0.010	Non-city
Rochester, NY	0.073	City	Wisconsin	-0.011	Non-city
New Hampshire	0.064	Non-city	Dallas, TX	-0.012	City
Colorado	0.063	Non-city	Ohio	-0.013	Non-city
New York, NY	0.062	City	Kentucky	-0.013	Non-city
Kansas City, MO	0.061	City	Florida	-0.013	Non-city
Vermont	0.057	Non-city	Delaware	-0.015	Non-city
Philadelphia, PA	0.054	City	Miami, FL	-0.019	City
New York	0.047	Non-city	Houston, TX	-0.025	City
Milwaukee, WI	0.046	City	Greensboro, NC	-0.026	City
Chicago, IL	0.045	City	Minnesota	-0.026	Non-city
Indianapolis, IN	0.044	City	New Orleans, LA	-0.027	City
San Diego, CA	0.042	City	California	-0.030	Non-city
Baltimore, MD	0.037	City	South Dakota	-0.031	Non-city
Massachusetts	0.036	Non-city	Indiana	-0.031	Non-city
Detroit, MI	0.034	City	Virginia Beach, VA	-0.031	City
Cleveland, OH	0.033	City	Virginia	-0.036	Non-city
Atlanta, GA	0.033	City	Missouri	-0.037	Non-city
Montana	0.031	Non-city	North Carolina	-0.038	Non-city
St Louis, MO	0.027	City	Nevada	-0.040	Non-city
Maine	0.027	Non-city	Tennessee	-0.040	Non-city
Buffalo, NY	0.026	City	Alabama	-0.044	Non-city
West Virginia	0.025	Non-city	Los Angeles, CA	-0.044	City
Washington	0.022	Non-city	Riverside, CA	-0.047	City
Columbus, OH	0.021	City	Maryland	-0.053	Non-city
Wyoming	0.020	Non-city	South Carolina	-0.060	Non-city
Cincinnati, OH	0.016	City	Texas	-0.067	Non-city
New Mexico	0.015	Non-city	Arkansas	-0.076	Non-city
Rhode Island	0.011	Non-city	Louisiana	-0.080	Non-city
Michigan	0.008	Non-city	Mississippi	-0.081	Non-city
Oregon	0.008	Non-city	Georgia	-0.122	Non-city
Tampa, FL	0.007	City			

Note: Average local human capital index and region type. 34 cities are denoted by the name of largest city with the abbreviation of the state. 47 non-city areas are denoted by the name of the states. Detailed definitions of occupation index in section 2. *Data sources*: Current Population Survey MORG and author's own calculations.

Table A.4: Second stage results for w_r with selected 2-digit industry

VARIABLES	(1) w_r	(2) w_r	(3) w_r	(4) w_r	(5) w_r	(6) w_r	(7) w_r	(8) w_r	(9) w_r
H_s^{inst}	0.275	2.712	-1.011	0.900	-10.71	-1.510	-19.30	0.188	-0.163
	(0.48)	(0.68)	(-2.41)	(0.92)	(-0.31)	(-1.93)	(-0.26)	(0.76)	(-0.23)
2-digit Ind	17	44	70	71	72	73	74	80	81
Observations	1,739	1,739	1,739	1,739	1,739	1,739	1,739	1,739	1,739
R-squared	0.988	0.933	0.984	0.982	0.213	0.976	-1.561	0.988	0.989
rmse	0.0337	0.0795	0.0392	0.0415	0.272	0.0476	0.491	0.0331	0.0321
H_s^{inst}	-0.763	0.211	1.507	-4.393	11.40	1.528	-0.617	-0.198	0.218
	(-5.24)	(1.02)	(0.33)	(-0.58)	(0.37)	(0.97)	(-0.54)	(-0.26)	(0.71)
2-digit Ind	82	83	84	85	86	87	88	89	90
Observations	1,739	1,739	1,739	1,739	1,739	1,739	1,739	1,739	1,739
R-squared	0.986	0.988	0.970	0.863	0.065	0.970	0.987	0.989	0.988
rmse	0.0368	0.0341	0.0541	0.116	0.304	0.0545	0.0353	0.0329	0.0341
Region Dummy	Y	Y	Y	Y	Y	Y	Y	Y	Y
Time Dummy	Y	Y	Y	Y	Y	Y	Y	Y	Y

DK t-statistics in parentheses

Table A.5: Second stage results for w_c with selected 2-digit industry

VARIABLES	(1) w_c	(2) w_c	(3) w_c	(4) w_c	(5) w_c	(6) w_c	(7) w_c	(8) w_c	(9) w_c
H_s^{inst}	0.552	4.142	1.232	0.458	0.713	1.877	3.851	-0.399	0.0842
	(1.11)	(0.33)	(1.91)	(2.39)	(0.46)	(1.17)	(1.52)	(-1.43)	(0.49)
2-digit Ind	17	44	70	71	72	73	74	80	81
Observations	1,258	1,258	1,258	1,258	1,258	1,258	1,258	1,258	1,258
R-squared	0.984	0.804	0.969	0.985	0.981	0.947	0.828	0.989	0.989
rmse	0.0388	0.136	0.0536	0.0371	0.0418	0.0705	0.127	0.0319	0.0325
H_s^{inst}	-3.054	-0.303	2.911	4.013	2.024	-2.619	1.476	0.122	4.020
	(-1.60)	(-0.89)	(1.69)	(1.80)	(1.25)	(-2.48)	(1.40)	(0.59)	(0.93)
2-digit Ind	82	83	84	85	86	87	88	89	90
Observations	1,258	1,258	1,258	1,258	1,258	1,258	1,258	1,258	1,258
R-squared	0.909	0.989	0.894	0.815	0.941	0.932	0.962	0.989	0.814
rmse	0.0924	0.0315	0.0997	0.132	0.0745	0.0800	0.0598	0.0328	0.132
Region Dummy	Y	Y	Y	Y	Y	Y	Y	Y	Y
Time Dummy	Y	Y	Y	Y	Y	Y	Y	Y	Y

DK t-statistics in parentheses

Table A.6: Second stage results for v_r with selected 2-digit industry

VARIABLES	(1) v_r	(2) v_r	(3) v_r	(4) v_r	(5) v_r	(6) v_r	(7) v_r	(8) v_r	(9) v_r
H_s^{inst}	3.890	-27.27	-3.602	2.255	2.444	-4.432	-62.51	1.578	1.893
	(1.33)	(-1.11)	(-5.29)	(0.68)	(0.08)	(-1.75)	(-0.25)	(1.30)	(0.68)
2-digit Ind	17	44	70	71	72	73	74	80	81
Observations	1,739	1,739	1,739	1,739	1,739	1,739	1,739	1,739	1,739
R-squared	0.831	-1.376	0.911	0.881	0.876	0.896	-11.926	0.897	0.890
rmse	0.184	0.688	0.134	0.154	0.157	0.144	1.605	0.144	0.148
H_s^{inst}	-11.08	-1.389	3.611	-29.65	4.195	26.03	-11.48	-3.600	-2.739
	(-9.20)	(-1.55)	(0.29)	(-0.68)	(0.14)	(1.82)	(-1.85)	(-2.24)	(-2.48)
2-digit Ind	82	83	84	85	86	87	88	89	90
Observations	1,739	1,739	1,739	1,739	1,739	1,739	1,739	1,739	1,739
R-squared	0.605	0.927	0.841	-1.820	0.819	-1.663	0.578	0.911	0.921
rmse	0.281	0.120	0.178	0.750	0.190	0.729	0.290	0.134	0.125
Region Dummy	Y	Y	Y	Y	Y	Y	Y	Y	Y
Time Dummy	Y	Y	Y	Y	Y	Y	Y	Y	Y

DK t-statistics in parentheses

Table A.7: Second stage results for v_c with selected 2-digit industry

VARIABLES	(1) v_c	(2) v_c	(3) v_c	(4) v_c	(5) v_c	(6) v_c	(7) v_c	(8) v_c	(9) v_c
H_s^{inst}	-1.292	-19.00	3.983	-0.705	14.34	13.16	34.82	-2.134	-3.565
	(-0.54)	(-0.58)	(1.46)	(-0.92)	(1.10)	(1.57)	(1.92)	(-0.97)	(-4.70)
2-digit Ind	17	44	70	71	72	73	74	80	81
Observations	1,258	1,258	1,258	1,258	1,258	1,258	1,258	1,258	1,258
R-squared	0.931	-0.016	0.859	0.931	0.249	0.349	-2.801	0.927	0.911
rmse	0.149	0.568	0.211	0.148	0.489	0.455	1.099	0.152	0.168
H_s^{inst}	-37.79	-5.249	12.30	20.85	15.32	-17.41	5.937	4.262	9.464
	(-1.73)	(-2.93)	(1.65)	(1.64)	(1.50)	(-2.46)	(1.55)	(3.33)	(0.89)
2-digit Ind	82	83	84	85	86	87	88	89	90
Observations	1,258	1,258	1,258	1,258	1,258	1,258	1,258	1,258	1,258
R-squared	-3.021	0.877	0.418	-0.455	0.158	0.143	0.792	0.851	0.614
rmse	1.131	0.197	0.430	0.680	0.517	0.522	0.257	0.218	0.350
Region Dummy	Y	Y	Y	Y	Y	Y	Y	Y	Y
Time Dummy	Y	Y	Y	Y	Y	Y	Y	Y	Y

DK t-statistics in parentheses

Table A.8: OLS regression on knowledge spillover over time and cross sections

VARIABLES	(1) y_r^x-log(cpi)
$T_{w_r^r}$ [Reference]	0.00247
	(1.11)
$T_{w_r^c}$	-0.000117
	(-65.62)
$T_{v_r^r}$	-2.16e-05
	(-1.15)
$T_{v_r^c}$	-0.000691
	(-34.59)
$\omega_{yx} \times \bar{H}_t$	-0.871
	(-17.01)
Constant	-9.430
	(-2.12)
Observations	148
R-squared	0.976
rmse	0.0973

D.K s.e. t-statistics in parentheses

Table A.9: Counterfactuals - Formula

Region(s)		Region set	
Levels 1979&2015			
1.	employment ($\times 10^{-6}$)	$\Sigma_{r\in\mathcal{R}} e^{n_r}$	
	wage/rent p. person ($\times 10^{-3}$ \$)	wages	rents
2a.	without agglomeration	$N_{\mathcal{R}}^{-1}\Sigma_{r\in\mathcal{R}} e^{\omega_0+H_r^*}$	$N_{\mathcal{R}}^{-1}\Sigma_{r\in\mathcal{R}} e^{\omega_0+\lambda_0+H_r^*}$
2b.	agglomeration multiplier	$N_{\mathcal{R}}^{-1}\Sigma_{r\in\mathcal{R}} e^{w_r}$	$N_{\mathcal{R}}^{-1}\Sigma_{r\in\mathcal{R}} e^{v_r-\lambda_0}$
3.	aggregate income ($\times 10^{-9}$ \$)	$W_{\mathcal{R}}$	$V_{\mathcal{R}}$
3a.	rent-wage share (%)	$V_{\mathcal{R}}/W_{\mathcal{R}}$	
Changes 1979-2015 (%)			
	Return to human capital		
4a.	private return	$\overline{\Delta H}_{w\mathcal{R}}$	$\overline{\Delta H}_{v\mathcal{R}}$
4b.	public return	$\omega_H D_{wH}\overline{\Delta H}^*_{w\mathcal{R}}$	$\lambda_w\omega_H D_{vH}\overline{\Delta H}^*_{v\mathcal{R}}$
4c.	change in employment density	$-\overline{\Delta H}_{w\mathcal{R}} - \overline{\lambda}_z\omega_H D_{vH}\overline{\Delta H}^*_{w\mathcal{R}}$	$-\overline{\Delta H}_{v\mathcal{R}} - \overline{\lambda}_z\omega_H D_{vH}\overline{\Delta H}^*_{v\mathcal{R}}$
4.	total return (4a.+4b.+4c.)	$\omega_H\left(D_{wH}-\overline{\lambda}_z D_{vH}\right)\overline{\Delta H}^*_{w\mathcal{R}}$	$\overline{\eta}_\lambda^{-1}\lambda_z\omega_H D_{vH}\overline{\Delta H}^*_{v\mathcal{R}}$
5.	total aggregate income	$\omega_H\frac{\Sigma_{r\in\mathcal{R}}\left[W_r\left(D_{wH}-\overline{\lambda}_z D_{vH}\right)+V_r\overline{\eta}_\lambda^{-1}\lambda_z D_{vH}\right]\Delta H_r^*}{W_{\mathcal{R}}+V_{\mathcal{R}}}$	
6.	return per cap per y.o.s.		
6a.	in the time series	$\frac{\Sigma_{r\in\mathcal{R}}[W_r(1+\overline{\mu}\omega_H D_{wH})+V_r(1+\overline{\mu}\lambda_w\omega_H D_{vH})]}{10(W_{\mathcal{R}}+V_{\mathcal{R}})}$	
6b.	in the cross section	$\frac{\Sigma_{r\in\mathcal{R}}[W_r(1+\omega_H D_{wH})+V_r(1+\lambda_w\omega_H D_{vH})]}{10(W_{\mathcal{R}}+V_{\mathcal{R}})}$	
7.	total welfare		
7a.	in the time series	$\overline{\mu}\frac{\Sigma_{r\in\mathcal{R}}V_r\overline{\eta}_\lambda^{-1}\lambda_z\omega_H D_{vH}\Delta H_r}{W_{\mathcal{R}}+V_{\mathcal{R}}}$	
7b.	in the cross section	$\frac{\Sigma_{r\in\mathcal{R}}V_r\overline{\eta}_\lambda^{-1}\lambda_z\omega_H D_{vH}\Delta H_r}{W_{\mathcal{R}}+V_{\mathcal{R}}}$	

Table A.10: Counterfactual no cities - Formular

Nation 2015	Aggregate income ($\times (10)^{-9}$ \$)				Change (%)		
	Actual		Counterfactual				
	wages	rents	wages	rents	wages	rent	total
All regions rural, no cities	$W_{\mathcal{R}}$	$V_{\mathcal{R}}$	$W_{\mathcal{R}}^{r}$	$V_{\mathcal{R}}^{r}$	$W_{\mathcal{R}}^{r}/W_{\mathcal{R}}-1$	$V_{\mathcal{R}}^{r}/V_{\mathcal{R}}-1$	$(W_{\mathcal{R}}^{r}+V_{\mathcal{R}}^{r})\,/\,(W_{\mathcal{R}}+V_{\mathcal{R}})-1$
All regions same H_r	$W_{\mathcal{R}}$	$V_{\mathcal{R}}$	$W_{\mathcal{R}}^{H}$	$V_{\mathcal{R}}^{H}$	$W_{\mathcal{R}}^{H}/W_{\mathcal{R}}-1$	$V_{\mathcal{R}}^{H}/V_{\mathcal{R}}-1$	$\left(W_{\mathcal{R}}^{H}+V_{\mathcal{R}}^{H}\right)/\,(W_{\mathcal{R}}+V_{\mathcal{R}})-1$

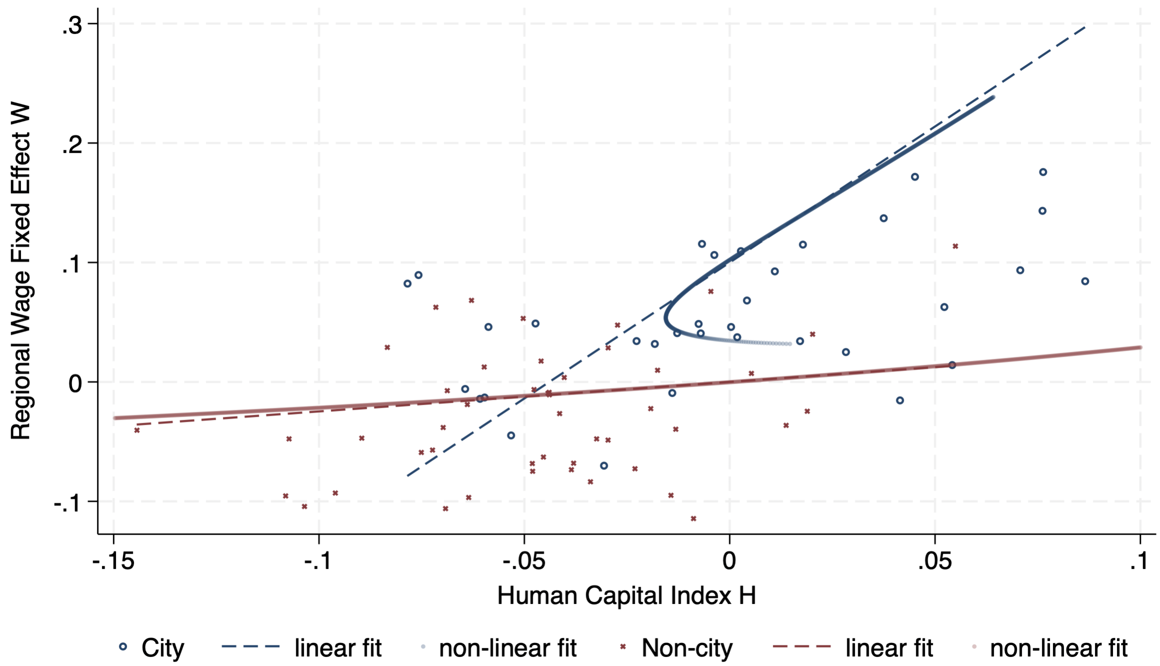

Figure I: w_r^c and w_r^r against H_r for $T_r = 0$