Should I Stay or Should I Grow?

How Cities Affect Learning, Inequality and Aggregate Productivity

Hugo Lhuillier*

Princeton University

November 3, 2023

Updated regularly, please click here for the latest version.

Abstract

The spatial concentration of talent is a robust pattern of modern economies. While the sorting of individuals into cities begets regional disparities, it may benefit aggregate productivity by fostering human capital accumulation. To study this equity-efficiency tradeoff, I propose a theory wherein individuals learn from their neighbors. Cities affect the stock of human capital by determining the frequency at which heterogeneous workers meet. Learning complementarities determine the existence of a tradeoff between productivity and spatial inequality. I estimate the model on French administrative data. I recover learning complementarities from a local projection of future wages on present wages and the wages of nearby individuals. I find that workers employed in relatively skill-dense cities experience faster wage growth, in particular if they are skilled. I address endogeneity concerns by using skill-density variations within firms across neighborhoods driven by past productivity shocks. The model explains two-thirds of the between-city wage growth variance and gives rise to a steep tradeoff between aggregate human capital and spatial inequality. I assess the implications of this tradeoff for the general equilibrium effects of moving vouchers. I find that large vouchers are effective at reducing spatial disparities in learning opportunities at the cost of decreased aggregate efficiency.

*Contact: hugo.lhuillier@princeton.edu. I thank my advisors Gene Grossman, Richard Rogerson, Esteban Rossi-Hansberg and Gianluca Violante for their support and guidance in this project. I am indebted to Adrien Bilal, John Grigsby and Steve Redding for many insightful discussions. I am also grateful to Thomas le Barbanchon, Juanma Castro-Vincenzi, Levi Crews, Théo Durandard, Mayara Felix, Cécile Gaubert, Gregor Jarosch, Benny Kleinman, Moritz Lenel, Adrien Matray, Eduardo Morales, Dávid Nagy, Ezra Oberfield, Karthik Sastry, Andreas Schaab, and Derek Wenning for valuable comments. This research benefited from financial support from the International Economics Section and the Simpson Center for Macroeconomics at Princeton University. Their support is gratefully acknowledged. This work is supported by a public grant overseen by the French National Research Agency (ANR) as part of the "Investissements d'Avenir" program (reference: ANR-10-EQPX-17—Centre d'accès sécurisé aux données—CASD). Many thanks to Marine Doux and Médianes for hosting the physical data in France.

1 Introduction

Spatial inequalities are pervasive across the globe. Cities such as New York, Paris, and Tokyo have become epicenters of talent, leaving remote regions deprived of such individuals. As workers exchange ideas with, collaborate with, and potentially learn from their neighbors, the spatial concentration of skills may have enduring effects on human capital accumulation while at the same time segregating learning opportunities in a few cities. Concerned by such regional disparities, national governments spend a sizable share of their budget every year to rebalance income across space. Yet, little is known about the aggregate effects of these policies on human capital. How does the sorting of workers into cities shape learning? To what extent are learning opportunities spatially concentrated, and how much do their concentration amplify regional disparities? Can policies spur higher productivity in conjunction with lower geographical inequalities?

This paper offers answers to these questions with four contributions. First, I propose a theory that sheds light on the importance of learning complementarities for the tradeoff between spatial inequality and human capital accumulation. Second, I estimate learning complementarities using French administrative data. Third, I use the estimated model to quantify the tradeoff between human capital accumulation and spatial inequality. Fourth, I evaluate the implications of this tradeoff for a large-scale moving voucher policy.

In the first part of the paper, I propose a theory wherein the city where you work affects your accumulation of human capital. The structure is an overlapping generation model with two key features. First, young workers are endowed with heterogeneous skills, cities differ in their productivity, and skill and productivity are complements in production. Second, young workers learn by randomly meeting the other workers in their city. The learning depends on which workers are engaged in the interaction. I do not restrict learning complementarities. Skilled individuals may disproportionately learn from other skilled workers. In this case, within-skill learning complementarities dominate between-skill complementarities. On the contrary, between-skill complementarities prevail if the marginal gains from meeting a high-skill are decreasing in workers' own skill.

The sorting of workers in cities shapes the spatial distribution of learning opportunities, which in turn may amplify geographic disparities. Workers are free to locate anywhere, and they choose a city that maximizes their lifetime utility. Holding constant future income, high-skill sort into productive cities as these locations offer relatively higher earnings inherited from the production complementarities. Yet, workers are forward-looking, and their location decisions also reflect the local learning opportunities that shape future earnings. When skills are concentrated in productive cities, these locations turn into attractive learning centers to any individual who learns relatively more from high-skill. In particular, the stronger the within-skill learning complementarities, the more skilled workers disproportionately gain over time from working in skill-dense environments, and the more they concentrate in a few cities to learn from each other.

Spatial inequality feeds into the aggregate stock of human capital. When high-skill agglomerates next to each other, it increases the frequency at which they meet while reducing the opportunity for low-skill to learn from them. Whether a greater segregation of interactions enhances aggregate productivity relies on which workers learn the most from the high-skill. The stock of human capital expands with greater spatial inequality if skilled workers learn disproportionately from interacting with their peers. In contrast, the spatial concentration of learning opportunities reduces aggregate productivity if between-skill complementarities prevail.

Spatial policies have the potential to accelerate aggregate human capital accumulation. The competitive allocation is inefficient as workers do not internalize their impacts on others' learning. The marginal social value of interactions hinges on learning complementarities. When high skills learn relatively more from each other, low skills crowd out the learning opportunities of productive cities, and the competitive equilibrium features too little spatial skill concentration. The reverse occurs when between-skill complementarities are relatively stronger. Skilled workers do not consider how their location choice helps low-skills learn, and the decentralized equilibrium features too much spatial inequality. Regardless of the complementarities, the optimal allocation spurs human capital accumulation and can be decentralized by place-based policies that vary by skill.

The second part of the paper develops and structurally estimates a quantitative version of the framework with three main additions. First, workers consume local housing in addition to the freely traded good already present in the theory. Second, workers face migration costs. The migration costs imply that workers' birthplace shapes their lifetime opportunities. Third, I impose a flexible functional form on the learning technology according to which three elasticities govern skill growth: with respect to workers' own skill, the skill of the worker they interact with, and a cross-term that captures learning complementarities.

I estimate the quantitative framework using French administrative data. The estimation is split into two steps. In the first step, I use the structure of the model to derive estimating equations that identify half of the parameters. In the second step, I calibrate the remaining parameters to match salient features of the wage distribution in the aggregate, across cities, and across the lifecycle. The estimated framework replicates several non-targeted moments such as how workers' willingness to sort varies across the lifecycle and how workers' birthplace determines their lifetime income.

The learning complementarities are transparently estimated. The learning technology is recovered from a model-consistent local projection of future wages on workers' current wages and the average wage of the individuals working in the same city. I find the three elasticities governing the learning technology to be positive. First, workers with higher skills experience faster skill growth regardless of with whom they interact. This own-skill effect is quantitatively small. Second, every worker learns more from skilled individuals. Third, the returns to meeting a high-skill are relatively larger for skilled workers. Altogether, the learning technology displays stronger within- than between-skill complementarities.

The estimated returns to local interactions are significant. The model explains 64% of the between-city wage growth variance. The predicted wage growths are unevenly distributed across space. The average worker who migrates from Troyes (bottom 10% of the growth distribution) to Paris (top 10%) experiences a wage growth 4.5 log points higher over five years – 40% of the average 5-year wage growth in France. The growth gains are heterogeneous across workers due to the estimated complementarities. A worker in the bottom 10% of the wage distribution sees its wage growth increase by 2.8 log points when moving from Troyes to Paris, whereas the dynamic gains are 7.3 log points for individuals in the top 10%.

The equation estimating the learning technology allows me to derive clear-cut conditions that guarantee unbiased estimates. Specifically, the drivers of wage growth other than local interactions must be identically and independently distributed across cities and individuals conditional on workers' wages. I use the granularity of the French matched employer-employee data to relax this identification assumption in three ways. First, I control for worker-level characteristics, such as age and tenure, known to shape wage growth. Second, I saturate the local projection with fixed effects to control for unobserved spatial heterogeneity in wage dynamics flexibly. In the most restrictive specification, the learning elasticities are estimated off variations in average wages across neighborhoods of employment for individuals who live in the same neighborhood and work in the same firm within the same city. Third, to control for neighborhood-level confounding shocks, I instrument present neighborhood wages with variations in the local share of white-collar twenty years ago. Armed with these controls and instrument, I find that neighborhoods with a greater skill density boost the wage growth of the local workers – especially if those are skilled. The implied learning elasticities cannot be statistically differentiated from the baseline estimates.

In the third part of the paper, I use the estimated model to quantify how the segregation of learning opportunities across cities shapes human capital accumulation and spatial disparities. To do so, I compare the baseline equilibrium in which cities segment interactions with a counterfactual economy in which individuals learn from every worker in the economy.

The spatial concentration of skilled workers has a substantial effect on human capital accumulation. The aggregate stock of human capital is 1.2% higher when learning takes place within cities. Learning complementarities explain the majority of this productivity gain: the aggregate average skill would only be 0.5% higher were all workers to learn equally from skilled workers. High-skill workers entirely drive the aggregate effect of cities on human capital. Skilled workers indeed agglomerate in productive locations, which increases the frequency at which they learn from each other. As a result, individuals in the top 10% of the skill distribution experience a lifetime skill growth 26% higher when they only interact with their neighbors. In contrast, workers at the bottom of the skill distribution neither gain nor lose from the spatial concentration of learning opportunities as they do not sort and learn relatively less from skilled workers.

The confinement of learning opportunities to productive cities also begets spatial disparities. First, it engenders learning inequality. The segmentation of interactions, rather than the sorting of fast-learning workers, explains the vast majority of the learning gaps between cities. As a result, workers born in remote cities experience slower human capital accumulation: workers born in the bottom 25% of the city-growth distribution have a lifetime skill growth 10% lower than the average French worker. Second, the presence of city-specific learning amplifies spatial wage inequality. Productive cities get bigger as workers need to live there to accumulate human capital faster. In addition, the spatial concentration of high-skill rises as they benefit more over time from working in skill-dense locations. Combined, spatial wage inequality increases: the between-city wage variance grows from 0.007 to 0.024 when learning occurs within cities.

Combined, I conclude the existence of a steep tradeoff between spatial inequality and human capital accumulation. In the fourth part of the paper, I assess the consequences of this equity-efficiency tradeoff for spatial policies. I evaluate a policy that provides vouchers to every worker born in the bottom 25% of the city-growth distribution to move to the three largest French cities.

The moving voucher policy is spatially redistributive. On the one hand, the policy improves the opportunities of treated workers by helping them reallocate to skill-dense locations. When the subsidy covers 80% of the average migration cost, which represents 1.5% of GDP, the probability that treated workers migrate to the three largest French cities rises from 7% to 45%. Treated individuals have an easier access to learning centers, and their lifetime skill growth coincides with that of the average worker. As a result, their welfare increases by 2% on average. On the other hand, the moving vouchers depress the learning of the non-treated. The policy attenuates the spatial concentration of skills by reallocating marginally less skilled workers to productive cities. The local learning opportunities worsen, and workers born there, as well as the other non-treated workers who were migrating to these cities to learn, experience slower human capital accumulation. Combined with the financial cost of the policy, non-treated workers suffer average welfare losses of 2.3%.

While the policy reduces spatial disparities, it acts negatively on aggregate efficiency. The stock of human capital is 0.3% lower when the policy reaches 1.5% of GDP. The general equilibrium effect of the policy more than offset its partial equilibrium gains. Holding the quality of local interactions constant, human capital would rise by 0.8% as the policy expands the pool of workers with access to fast-learning opportunities. Absent learning complementarities, the general equilibrium feedback would be a third smaller, and aggregate human capital would neither rise nor fall from the policy. I conclude that learning complementarities are crucial for the tradeoff between spatial inequality and human capital accumulation.

Related literature This paper primarily adds to two strands of the literature. The first strand studies how social interactions shape productivity growth. Started by the seminal work of Jovanovic and Rob (1989), and later developed by Lucas (2009a), Lucas and Moll (2014), Perla and Tonetti (2014), and Akcigit et al. (2018), amongst many others, this literature initially abstracted from space. The idea that cities facilitate knowledge diffusion is however not novel and can be traced back to Marshall (1890) and Jacobs (1969).[1] In particular, if interactions spur learning and mainly occur between individuals who work next to each other, then the spatial distribution of talent must matter for human capital accumulation. Two papers have recently quantified this insight.[2] Martellini (2022) finds that local peer effects explain the majority of the faster lifecycle wage growth of large cities in the United States. Crews (2023) builds a dynamic model in which the vibrancy of cities, defined by their size and skill composition, shapes how workers learn. He concludes that

[1]Since then, a long list of papers have documented how proximity shapes knowledge diffusion (e.g. Jaffe et al., 1993; Audretsch and Feldman, 2004; Moretti, 2021; Atkin et al., 2022; Baum-Snow et al., 2023).

[2]Several other recent papers have studied the impact of space on aggregate growth through trade and technology diffusion, e.g. Lucas (2009b), Alvarez et al. (2013), Sampson (2016), Desmet et al. (2018), Buera and Oberfield (2020), Perla et al. (2021), and Cai et al. (2022).

relaxing land-use regulations in New York and San Francisco spur aggregate growth.[3]

I contribute to this literature by characterizing theoretically when do cities spur faster human capital accumulation. In particular, by not imposing restrictions on learning complementarities, I show that the spatial segmentation of learning opportunities may be detrimental to learning, an insight already noted by Benabou (1993).

I also contribute to the literature that estimates the dynamic returns to interactions between neighbors and coworkers. Across space, Baum-Snow and Pavan (2012), de la Roca and Puga (2017), and Eckert et al. (2022), among others, find the wage age-profile of large cities to be relatively steeper. Within the firm, Nix (2020) and Jarosch et al. (2021) document that workers employed in high-wage firms experience faster wage growth. I add to this literature in three ways. First and foremost, I estimate how the returns to local interactions vary across workers. The granularity of the French matched employer-employee allows me in particular to estimate these learning complementarities while addressing the endogeneity concerns of the peer effect literature (Angrist, 2014). Second, I find the skill composition of cities to be one order of magnitude more important than city size in shaping workers' wage growth. Third, I show that the neighborhood where individuals work affects their wage growth even after controlling at which establishment and firm they are employed.

Finally, this paper connects with several papers in the spatial literature. First, it adds to the papers that investigate the sources of spatial agglomeration (Duranton and Puga, 2004; Behrens and Robert-Nicoud, 2015). Two papers in particular have studied theoretically how learning within cities triggers spatial agglomeration. Glaeser (1999) shows that urbanization rises when the ability to learn by imitation increases. More recently, Davis and Dingel (2019) finds that local interactions spur between-city differences in inequality even when cities are *ex-ante* similar. I contribute to this literature by quantifying the strength of this agglomeration force. I find that local learning opportunities amplify the sorting of workers across space, an empirical pattern documented by Combes et al. (2008), Eeckhout et al. (2014), and Diamond (2016), among others. Second, this paper adds to the literature that studies optimal spatial policies.[4] Most related are the papers by Rossi-Hansberg et al. (2019) and Fajgelbaum and Gaubert (2020) who study optimal policy in the presence of worker sorting and static production spillovers.[5]. Related to these papers, I characterize the optimal spatial policy when spillovers are dynamic.

The rest of this paper is organized as follows. Section 2 presents the theory and characterizes the tradeoff between spatial inequality and human capital accumulation. Section 3 turns to the French matched employer-employee and provides evidence on the role of space in shaping learning. Section 4 estimates the model, and Section 5 quantifies the impact of local interactions on human capital accumulation and agglomeration. Section 6 then proceeds to study the consequences of spatial policies on spatial inequality, productivity, and welfare. Section 7 concludes.

[3]Another strand of recent papers focuses on the joint evolution of city sizes and aggregate growth (Rossi-Hansberg and Wright, 2007; Herkenhoff et al., 2018; Hsieh and Moretti, 2019; Duranton and Puga, 2023).

[4]Glaeser and Gottlieb (2008) and Kline and Moretti (2014) summarize the debate on the consequences of place-based policies on productivity.

[5]Gaubert (2018) and Bilal (2023) also derives the optimal spatial policy in the presence of firm sorting.

2 A Theory of Learning across Space

In this section, I shed light on the tradeoff between spatial inequality and human capital accumulation by offering a model of learning across space. The model puts learning complementarities front and center. The remaining ingredients are standard in the quantitative spatial literature (Redding and Rossi-Hansberg, 2017).

2.1 Setup

I model workers that sort into cities. Every period, a unit mass of young workers enter the labor market. Workers live for two periods and discount the future at rate β. Denote their first life period as young (y) and their second period as old (o). Old workers express no altruism towards future generations. Workers choose in which city to work at any point in time.[6] Cities are indexed by ℓ, and the total number of cities L is fixed. I study the steady state of this economy and omit time subscripts.

Human capital Young workers are endowed with an initial skill s^y. This skill is drawn from an aggregate skill distribution N^y. The young skill distribution is a primitive of the model. I suppose that the support of N^y is connected, positive, and bounded, $[\underline{s}^y, \bar{s}^y]$, with $\bar{s}^y < \infty$.[7] I also assume that skills are uni-dimensional.[8]

Workers learn in their youth. This learning is shaped by the interactions they encounter. Learning interactions are governed by three assumptions.

First, they are spatially segmented: interactions take place between workers who work in the same location. Second, interactions happen randomly within cities. Hence, while workers cannot target specific interactions, they can choose to locate in a particular city to increase their chance to meet a particular skill-type. In this section, I assume for the sake of tractability that interactions happen only between young workers. The likelihood of an interaction with a skill s_p in city ℓ is then given by the local density of this skill, $\pi^y_\ell(s_p)$. I relax this assumption in the quantitative model. Third and most importantly, the return to an interaction depends on which workers are engaged in it. If a worker with skill s interacts with a partner with skill s_p, they obtain new skill $s^o = e\gamma(s, s_p)$.[9] The learning technology γ encapsulates how local interactions shape human capital accumulation. In addition, interactions often have an idiosyncratic component, which is captured here by the learning shocks e. These shocks also reflect the sources of human capital accumulation other than interactions. They are drawn from the city-independent distribution F with support $\mathbb{R}_+$.[10]

[6] I define a city as a commuting zone in the data and thereby abstract from commuting decisions.

[7] A bounded support simplifies the proofs, but all that is require is the existence of the first moment, $\int s dN^y(s) < \infty$.

[8] For papers with multi-dimensional skill and sorting across firms or occupations, see Lindenlaub (2017), Guvenen et al. (2020) and Lise and Postel-Vinay (2020).

[9] I assume for simplicity that workers interact with a single partner in their youth.

[10] The idiosyncratic shocks are not required for the theory but smooth the aggregate old skill distribution. They help quantitatively to match the wage distribution of old workers.

Differently from the previous papers that have studied the impact of cities on learning (Martellini, 2022; Crews, 2023; Duranton and Puga, 2023), I impose minimal restrictions on the learning technology.

Assumption 1 (Learning technology)**.**
The learning technology γ is positive, bounded, and twice differentiable in (s, s_p).

Assumption 1 allows my framework to nest different skill learning complementarities present in the literature. First and foremost, it encompasses the case in which interactions are irrelevant for learning, $\gamma(s, s_p) = \gamma(s)$. In this case, workers' capacity to learn may still depend on their own skill, e.g. if productive workers are better at learning-by-doing (Huggett et al., 2011). Second, interactions may affect the learning of all workers equally, $\gamma(s, s_p) = \gamma(s_p)$. In this case, interactions do shape learning but the learning technology does not feature any skill complementarities. Third, the learning complementarities may be stronger within-skill than across-skill. In that case, the learning technology is supermodular:

$$\frac{\partial^2 \gamma(s, s_p)}{\partial s \partial s_p} > 0, \quad \forall (s, s_p).$$

Intuitively, supermodularity requires that the marginal gains from a skilled interaction are greater for skilled individuals. This occurs for instance when workers are better able to learn from each other when they share similar skills (Jovanovic, 2014). Alternatively, the learning technology may display complementarities that are stronger across than within-skill. Then, the learning technology is submodular,

$$\frac{\partial^2 \gamma(s, s_p)}{\partial s \partial s_p} < 0, \quad \forall (s, s_p),$$

and the marginal gains from a skilled interactions are greater for low-skilled individuals. This is for instance the case if knowledge diffuses from skilled workers to relatively less skilled individuals (e.g. Lucas and Moll, 2014; Perla and Tonetti, 2014).

The simple learning technology

$$\gamma(s, s_p) = g_0 + g_1 \times s + g_2 \times s_p + g_{12} \times s \times s_p \tag{1}$$

illustrates well the relative strength of within- and across-skill learning complementarities. The parameter g_0 governs the average skill of old workers, whereas $g_1 > 0$ implies that skilled workers experience on average faster learning. The return to interactions are dictated by g_2 and g_{12}. When $g_2 > 0$, interactions with skilled partners yield higher future skills. The learning complementarities are fully determined by g_{12}. If $g_{12} > 0$, the learning technology is supermodular, and skilled workers learn relatively more from skilled partners. On the contrary, the learning technology is submodular when $g_{12} < 0$.

Learning takes time, and the skill obtained from an interaction can only be used in the next, old period. The aggregate distribution of old skill is denoted by N^o. To the contrary of the young skill distribution, this is a general equilibrium object that reflects the learning of every young workers.

Preferences Individuals consume a single, freely-traded good taken to be the numéraire. I abstract from local consumption for the sake of tractability in this section. I add housing in the quantitative version of the model. Individuals hold linear utility over their consumption, which they finance using their current income. They supply inelastically one unit of labor per period, so that wages and incomes are interchangeable. They do not have access to saving devices.

In addition, individuals have idiosyncratic preferences for the city they work in, denoted by $\boldsymbol{\varepsilon} = \{\varepsilon_\ell\}_{\ell=1}^L$. These preferences are drawn from independent Gumbell distributions with city- and age-specific location parameters B_ℓ^a, $a \in \{y, o\}$, and scale parameter ϑ^{-1}. The location parameter B_ℓ^a reflects the amenities offered by city ℓ to workers with age a. Young workers may prefer cities with a vibrant nightlife whereas old workers may value more the quality of the opera, and I impose no relationship between B_ℓ^y and B_ℓ^o. Local amenities are taken as given. Finally, I assume that location preferences are independent across the lifecycle.[11]

Consumption and location preferences are perfect substitutes. Hence, the per-period utility of workers with income y and preferences $\boldsymbol{\varepsilon}$ when working in city ℓ is $y + \varepsilon_\ell$.

Production Production takes place within cities. In each city, there is a single representative competitive firm with productivity T_ℓ. The local TFP T_ℓ reflects the industries and firms present in the city and is taken as given. The representative firm produces the good using labor as sole input. If the firm hires a bundle of skill $\boldsymbol{n}_\ell = \{n_\ell(s)\}_s$, it produces output

$$Y_\ell(\boldsymbol{n}_\ell) = T_\ell \int s n_\ell(s) ds.$$

This production function has two notable properties. First, cities' TFP and workers' skills are complements: high-skill workers produce relatively more in productive cities than their low skill counterparts.[12] Second, skills are perfect substitutes. I assume away skill complementarities in production, which has been the focus of the spatial literature thus far (e.g. Diamond, 2016; Giannone, 2022), to put the emphasis on skill complementarities in the learning process.[13] Quantifying jointly the spatial consequences of static and dynamic skill complementarities is a promising avenue of research for future work.

Labor markets are segmented by city and by skill. The zero profit condition faced by the representative firm pins down the local wage schedule, $w_\ell(s) = sT_\ell$. Wages therefore inherit the complementarity between TFP and skill.

[11] I later add migration costs in the model. Migration costs and inter-temporal persistence of location preferences are isomorphic except for welfare accounting.

[12] TFP and skills are complements in that a marginal increase in the city's TFP increases the marginal productivity of a skill; that is, the production technology is supermodular in (s, T_ℓ). This definition of complementarity is standard in the monotone comparative statics literature (e.g. Topkis, 1998).

[13] The estimation of the learning technology in Section 3 can however accommodate complementarities in production.

Feasibility The spatial allocation of workers must be feasible. Specifically, the aggregate demand for workers with skill s must be less or equal to the aggregate supply of that skill:

$$\sum_{\ell} n_{\ell}^{a}(s) \leq n^{a}(s), \quad \forall\, s,\ \forall a \in \{y, o\},$$

where $n_{\ell}^{a}(s)$ is the measure of workers with age a and skill s employed in city ℓ, and $n^{a}(s)$ is the aggregate density of skill s for workers with age a. The relative presence of skill s in city ℓ amongst workers with age a can accordingly be defined as

$$\pi_{\ell}^{a}(s) = \frac{n_{\ell}^{a}(s)}{\int n_{\ell}^{a}(\sigma) d\sigma}, \quad \forall\, s,\ \forall a \in \{y, o\}. \tag{2}$$

2.2 The Steady State Equilibrium

The steady state equilibrium is represented by three equations governing the spatial allocation of young and old workers as well as the aggregate distribution of skill for old workers.[14]

Old workers Old workers solve a static location choice problem. They choose a city ℓ that maximizes their present utility conditional on their skill s and their idiosyncratic location preferences $\boldsymbol{\varepsilon}$,

$$V^{o}(s, \boldsymbol{\varepsilon}) = \max_{\ell} \left\{ w_{\ell}(s) + \varepsilon_{\ell} \right\}.$$

Workers' idiosyncratic preferences ensure that every skill is present in every location. The measure of old workers with skill s employed in city ℓ is given by

$$n_{\ell}^{o}(s) = n^{o}(s) \left(\frac{\mathrm{e}^{\vartheta \left(w_{\ell}(s) + B_{\ell}^{o} \right)}}{\sum_{\ell'} \mathrm{e}^{\vartheta \left(w_{\ell'}(s) + B_{\ell'}^{o} \right)}} \right). \tag{3}$$

The first term captures the aggregate supply of skill s amongst the old. The second term is the probability that skill s chooses to work in city ℓ, which follows from the Gumbell-distributed location preferences.

Young workers Young workers solve a dynamic location choice problem,

$$V^{y}(s, \boldsymbol{\varepsilon}) = \max_{\ell} \left\{ w_{\ell}(s) + \varepsilon_{\ell} + \beta \int \int \mathcal{V}^{o}[e\gamma(s, s_p)] dF(e) \pi_{\ell}^{y}(s_p) ds_p \right\},$$

where $\mathcal{V}^{o}(s) \equiv \mathbb{E}[V^{o}(s, \boldsymbol{\varepsilon}) \mid s]$ is the expected utility of an old worker with skill s prior to observing their idiosyncratic preferences.[15] Young workers' location decision reflect three considerations. The

[14] The young spatial allocation is decoupled from that of the old. Hence, the young allocation is independent from the initial conditions of the economy, and the economy converges to the steady state in one period.

[15] The expected utility takes the usual form in discrete choice models with Gumbel preferences, $\mathcal{V}^{o}(s) = c + \vartheta^{-1} \log \left(\sum_{\ell} \mathrm{e}^{\vartheta w_{\ell}(s)} \right)$ for c a constant.

first two are static and are common to old workers. The third consideration is dynamic and reflects the spatial learning opportunities they face. The expected utility of working in city ℓ is indeed the weighted average of the expected gains from particular interactions, $\mathcal{V}^o[e\gamma(s,s_p)]$, with weights given by the local likelihood of such interactions, $\pi_\ell^y(s_p)$. A location thus constitutes an attractive learning center if it provides a large density of individuals from whom workers learn relatively more. When workers learn equally from every interactions, $\gamma(s,s_p)=\gamma(s)$, this dynamic consideration disappears.

As for the old, the number of young workers with skill s employed in city ℓ follows from the Gumbell idiosyncratic location preferences,

$$n_\ell^y(s) = n^y(s)\left(\frac{\mathrm{e}^{\vartheta\left(w_\ell(s)+B_\ell^y+\beta\mathbb{E}[\mathcal{V}^o(E\gamma(s,S_\ell^y))]\right)}}{\sum_{\ell'}\mathrm{e}^{\vartheta\left(w_{\ell'}(s)+B_{\ell'}^y+\beta\mathbb{E}[\mathcal{V}^o(E\gamma(s,S_{\ell'}^y))]\right)}}\right). \tag{4}$$

where $\mathbb{E}[\mathcal{V}^o(E\gamma(s,S_\ell^y))] = \int\int \mathcal{V}^o[e\gamma(s,s_p)]dF(e)\pi_\ell^y(s_p)ds_p$ is the expected utility of young workers with skill s employed in city ℓ. This expectation is a general equilibrium object that depends on the location choices of the other young workers. When deciding where to work, young workers take those as given.

Skill distribution The aggregate skill distribution of old workers is a general equilibrium object that reflects the learning experienced by young individuals. Specifically, the fraction of old workers with skill less than s_o is given by

$$N^o(s_o) = \sum_\ell \int\int\int \mathbb{1}\left\{e\gamma(s_y,s_p)\leq s_o\right\}dF(e)n_\ell^y(s_y)\pi_\ell^y(s_p)ds_p ds_y. \tag{5}$$

The spatial allocation of young workers determine the aggregate skill distribution of old workers by shaping the aggregate probability of two skills interacting with each other.

Equilibrium Equations (3) to (5) describe the steady-state spatial equilibrium. Young workers choose a city where to work taking as given the spatial distribution of learning opportunities. In equilibrium, their expectations must be consistent with the other workers' location choices. The spatial allocation of young workers determine the aggregate skill distribution of old workers. Finally, the aggregate demand for old workers must be feasible.[16]

Definition 1 (Steady state equilibrium).
A steady state equilibrium is a young spatial allocation, $\{n_\ell^y\}_{\ell=1}^L$, an old spatial allocation, $\{n_\ell^o\}_{\ell=1}^L$, city-specific skill densities amongst young workers, $\{\pi_\ell^y\}_{\ell=1}^L$, and an aggregate old skill distribution, n^o, such that:

1. *Taking the behavior of other young workers as given, the young spatial allocation satisfies* (4)*;*
2. *Given the old skill distribution, the old spatial allocation satisfies* (3)*;*

[16]The equilibrium is block-recursive: the young allocation can be solved independently from the old's. This also implies that the economy converges to the steady state equilibrium in one period.

3. The local skill densities are consistent with the location decisions of young workers (2)*;*

4. The distribution of old skill satisfies (5)*.*

The steady state equilibrium is the outcome of an infinitely-dimensional non-linear fixed point in the local skill densities. Existence of an equilibrium is guaranteed by the idiosyncratic location preferences which, together with the cities' TFP, coordinate expectations.

Proposition 1 (Existence)**.**
A steady state equilibrium exists.[17]

Characterizing the steady state equilibrium amounts to solving for the spatial allocation of young workers. I highlight two special types of allocation that may prevail in equilibrium. In the first, the local skill densities are the same in every city, and therefore so are the learning opportunities. I refer to this allocation as symmetric.

Definition 2 (Symmetric allocation)**.**
An allocation is symmetric if the local skill densities are identical, $\pi^a_\ell(s) = n^a(s)$ *for all* s *and all* ℓ*.*

Alternatively, productive locations may feature a higher density of skilled workers if workers sort. In this case, I say that the allocation satisfies stochastic positive assortative matching (SPAM).[18]

Definition 3 (Stochastic positive assortative matching)**.**
An allocation satisfies stochastic positive assortative matching if the local skill densities can be ordered in the first-order stochastic dominance sense by cities' productivity: $\pi^a_\ell \succ_{FOSD} \pi^a_{\ell'} \iff T_\ell > T_{\ell'}$*.*

To gain analytical tractability without imposing structure on the learning technology, I solve for the steady state equilibrium when city TFP are not too dispersed, $T_\ell \approx \bar{T}$ for all ℓ. When cities share a common TFP, the production complementarity between workers' skill and city TFP vanishes. Workers no longer have an exogenous motive to sort across cities, and an equilibrium with a symmetric allocation exists (Proposition A.1). This equilibrium is unique when the dynamic gain of cities, summarized by $\vartheta\beta$, are not too large.[19] Approximating the equilibrium around the symmetric allocation thus simplifies how to keep track of the entire spatial distribution of workers. This approximation is used solely to characterize analytically the steady state; in particular, the estimation and counterfactuals are computed using the global solution.

Sections 2.3 and 2.4 characterize the steady state under this perturbation approach. Throughout, $\bar{x}$ refers to the value of x in the homogeneous TFP equilibrium, and I use capital letters to denote to random variables. For instance, $\mathbb{E}[S^y] = \int s n^y(s) ds$ is the average young skill in the economy, whereas $\mathbb{E}[S^y_\ell] = \int s \pi^y_\ell(s) ds$ denotes the average young skill in city ℓ.

[17] Appendix A presents the proofs of all propositions.

[18] This definition generalizes the notion of positive assortative matching in a context of many-to-one matching in which every skill is present in every city due to the presence of idiosyncratic location preferences.

[19] For a framework where local interaction alone generate asymmetric equilibria, see Davis and Dingel (2019).

2.3 Local Interactions as a Source of Agglomeration

I first show how local interactions shape the spatial distribution of economic activity. I start by characterizing the location decisions of old workers to then contrast it with the young allocation. When between-city differences in TFP are small, the number of old workers with skill s employed in city ℓ is to a first order given by[20]

$$n_\ell^o(s) \approx \underbrace{\frac{\bar{n}^o(s)}{L}}_{\text{Symmetric}} + \underbrace{\vartheta \left(\frac{\bar{n}^o(s)}{L}\right)\left(T_\ell - \bar{T}\right) s}_{\text{Sorting for working}},$$

where $\bar{n}^o(s)$ is the density of old workers with skill s when cities are homogeneous. When cities share a common TFP, old workers locate in cities that maximize their idiosyncratic preferences, and the resulting allocation is symmetric. When some cities are more productive than others, local wages are relative higher which attracts more workers. Furthermore, the between-city wage gaps are larger for relatively skilled workers due to the complementarity in the production function. The willingness to work in high TFP cities is therefore increasing in workers' skill, and SPAM follows for the old.[21] The dispersion in idiosyncratic location preferences modulates the amount of sorting.

The location decisions of young workers differ from the old in that they also care about local learning opportunities. Specifically, the measure of young workers with skill s in city ℓ is to a first order given by

$$n_\ell^y(s) \approx \underbrace{\frac{n^y(s)}{L}}_{\text{Symmetric}} + \underbrace{\vartheta \left(\frac{n^y(s)}{L}\right)\left(T_\ell - \bar{T}\right) s}_{\text{Sorting for working}} + \underbrace{\vartheta^2 \beta \left(\frac{n^y(s)}{L}\right)\left(T_\ell - \bar{T}\right) \Theta(s)}_{\text{Sorting for learning}}$$

Similarly to old workers, young workers prefer to work in relatively productive cities to earn higher wages everything else equal. The third, new term captures the learning considerations present only in the location decision of young workers. This sorting for learning motive would disappear were interactions not segmented by cities. Specifically, the function $\Theta(s)$ summarizes how much young workers with skill s value the learning opportunities of relatively productive cities. When $\Theta(s) > 0$, the learning opportunities present in relatively productive cities draw young workers with skill s into these locations. When $\Theta'(s) > 0$, the learning value of productive cities is increasing in workers' skills, and local interactions increase young workers' willingness to sort. This function is given by

$$\Theta(s) = \int \underbrace{\left(\gamma\left(s, s_p\right) - \mathbb{E}\left[\gamma\left(s, S^y\right)\right]\right)}_{\substack{\text{Relative learning from} \\ s_p \text{ interaction}}} \underbrace{\left(s_p + \vartheta \beta \Theta(s_p)\right)}_{\substack{\text{Willingness to} \\ \text{sort of } s_p}} dN^y(s_p). \tag{6}$$

The learning value of relatively productive cities relies on two objects. First, how much do workers

[20] For ease of exposition, I present in the main text the expression for the case of homogeneous local amenities. Appendix A presents the general expressions. Amenity differentials affect city size but not the local skill densities, and as a result, do not affect the qualitative predictions of the model.

[21] Proposition A.2 shows that SPAM always prevail for old workers for any spatial distribution of TFP.

with skill s learn from an interaction with a partner s_p relative to their average learning. This relative learning is exogenously determined by the learning technology. Second, how much more likely are interactions with s_p in relatively productive cities, which depends on this worker's willingness to sort. That is, the value that skill s places on the learning opportunities of relatively productive cities is an endogenous object that depends on the willingness to sort of all the other young workers.[22]

Were interactions not a source of learning, the willingness to sort of young workers would solely rely on their sorting for working motive. The spatial allocation of young workers would then display stochastic positive assortative matching. The sorting for learning motive could *a priori* offset this technological force if the value of skilled interactions was significantly larger for relatively low-skill workers. A higher density of low-skill workers in productive cities would however contradict the empirical evidence on sorting patterns found in this paper and others (e.g. Combes et al., 2008; Card et al., 2023, amongst others). I therefore focus on the part of the parameter space that generates a SPAM allocation for young workers.

Lemma 1 (Stochastic positive assortative matching)**.**
If $T_\ell \approx \bar{T}$ for all ℓ and $\vartheta\beta\mathbb{C}ov[\gamma_1(s, S^y), S^y] \geq -1$ for all s, the young allocation exhibits SPAM.

Lemma 1 bounds the relative strength of the between-skill learning complementarities. For instance, when the learning technology takes the form of (1), the condition simplifies to $\vartheta\beta g_{12}\mathbb{V}\text{ar}[S^y] \geq -1$, which imposes a bound on how submodular the learning technology can be. I assume that the condition of Lemma 1 holds for the remainder of Section 2.

In equilibrium, local interactions shape the size of cities. Under a SPAM allocation, productive cities offer relatively skill-intensive interactions. Productive cities thus constitute attractive learning centers if the learning derived from an interaction is increasing in the partner's skills. Local interactions then act as a source of agglomeration which increases the size of productive cities and decreases that of low TFP locations compared to a counterfactual economy in which interactions are not spatially segmented.

Local interactions also mold spatial wage inequality. To a second order, the between-city standard deviation of young workers' average wage is[23]

$$\mathbb{S}\text{d}[\mathbb{E}(W_\ell^y)] \approx \underbrace{\mathbb{E}[S^y]\mathbb{S}\text{d}[T_\ell]}_{\text{TFP differentials}} + \underbrace{\vartheta\mathbb{V}\text{ar}[S^y]\mathbb{S}\text{d}[T_\ell]}_{\text{Sorting for working}} + \underbrace{\vartheta^2\beta\mathbb{C}\text{ov}[S^y, \Theta(S^y)]\mathbb{S}\text{d}[T_\ell]}_{\text{Sorting for learning}}, \tag{7}$$

where $\mathbb{E}[W_\ell^y]$ is the average wage of young workers in city ℓ, and $\mathbb{S}\text{d}[T_\ell]$ is the standard deviation of cities' TFP.[24] An increase in the spatial dispersion of TFP directly contributes to greater between-city wage inequality amongst the young, which is captured by the first term. Worker sorting complements the spatial TFP differentials. In particular, local interactions affect spatial wage inequality by

[22]Technically, $\Theta(s)$ is the solution to a Fredholm integral equation of the second kind. If a solution to this equation exists, then it is unique. For $\vartheta\beta$ not too large, a solution exists (see Section A.6). Similar integral equations can be found in Allen and Arkolakis (2014).

[23]The first order terms naturally disappear from the approximation as the variance is a second order moment.

[24]Empirically, the between-city variance of wages is naturally computed in logs rather than in levels. While this matters quantitatively, it does not affect the qualitative prediction of the model.

shaping young workers' willingness to sort. When the within-skill complementarities dominate the between-skill complementarities, young skilled workers have relatively greater incentives to interact with their own peers. They therefore place a higher value on the learning opportunities present in productive cities, $\Theta' > 0$. In that case, local interactions amplify the between-city wage gaps amongst young workers compared to a world in which interactions were not spatially segmented, i.e. $\mathbb{C}\mathrm{ov}[S^y, \Theta(S^y)] > 0$. On the contrary, if the between-skill complementarities prevail, relatively low skill workers have greater incentives to work in productive cities to learn from skilled partners. Local interactions then dampen spatial wage inequality.

I summarize the impacts of local interactions on spatial inequality in Proposition 2. I use x^{C} to refer to the value of x in a counterfactual economy in which interactions are not spatially segmented.

Proposition 2 (Spatial inequality)**.**
When $T_\ell \approx \bar{T}$ for all ℓ:

1. *If $\gamma(s, \cdot)$ is increasing for all s, productive (unproductive) cities are larger (smaller) when interactions are spatially segmented: $N_\ell > N_\ell^C \Leftrightarrow T_\ell > \bar{T}$; conversely, $\gamma(s, \cdot)$ decreasing implies $N_\ell < N_\ell^C \Leftrightarrow T_\ell > \bar{T}$;*
2. *If γ is supermodular, the between-city variance of young wages is larger when interactions are spatially segmented; conversely, it is smaller when γ is submodular.*

Local interactions thus shape spatial inequality. I now characterize their impacts on aggregate productivity.

2.4 The Human Capital Consequences of Local Interactions

Between-city differences in skill growth arise endogenously from the sorting of young workers. In particular, if workers learn more from relatively skilled interactions, young workers accumulate human capital faster in productive cities and slower in low TFP locations.

Do these spatial differences in learning matter for aggregate productivity? In the aggregate, the average skill of old workers is given by

$$\mathbb{E}[S^o] = \mathbb{E}[E] \sum_\ell \int \int \gamma(s_y, s_p) n_\ell^y(s_y) \pi_\ell^y(s_p) ds_p ds_y, \tag{8}$$

where $\mathbb{E}[E] = \int e dF(e)$ is the average learning shock. This equilibrium stock of human capital can be contrasted with the average skill of old workers that prevails under the symmetric allocation:

$$\mathbb{E}[\bar{S}^o] = \mathbb{E}[E] \int \int \gamma(s_y, s_p) n^y(s_y) n^y(s_p) ds_p ds_y. \tag{9}$$

Equation (9) also coincides with the average skill of old workers when interactions are not spatially segmented since the spatial allocation of workers is then irrelevant for human capital accumulation. Comparing (8) and (9), cities affect the aggregate stock of human capital if they enable particular workers to meet specific partners more frequently. If interactions benefit every worker equally, i.e. $\gamma(s, s_p) = \gamma(s_p)$, (8) and (9) coincides and local interactions do not shape aggregate productivity. Absent learning complementarities, the faster human capital accumulation of individuals working in productive cities is indeed exactly offsetted by the slower learning of workers employed in the other locations.

To characterize when do local interactions boost aggregate productivity, I take a second order approximation of the average old skill around the symmetric equilibrium. Local interactions do not have first order effect on aggregate human capital as the spatial differences in learning cancel each other. Hence, to a second order, the average skill of old workers is

$$\mathbb{E}[S^o] \approx \mathbb{E}[\bar{S}^o] + \vartheta^2 \mathbb{V}\mathrm{ar}[T_\ell]\Omega. \tag{10}$$

The constant Ω fully encapsulates the impact of local interactions on aggregate productivity. The stock of human capital is larger when interactions are spatially segmented if $\Omega > 0$, and smaller otherwise. The constant Ω is given by

$$\Omega = \int\int \gamma(s, s_p)\left(\eta^y(s) - \mathbb{E}[\eta^y(S^y)]\right)\left(\eta^y(s_p) - \mathbb{E}[\eta^y(S^y)]\right) dN^y(s_p) dN^y(s),$$

where $\eta^y(s) \equiv s + \vartheta\beta\Theta(s)$ is the total willingness to sort of young workers with skill s. The integrand is comprised of three terms. The first term captures the learning that occurs when s and s_p interacts. The second and third represent the aggregate probability that these two skills interact.

In a SPAM equilibrium, cities increase the chance that workers with similar skills interact. Whether this is beneficial for aggregate productivity hinges on the learning complementarities. When within-skill complementarities are strong, the marginal gains from skilled interactions are increasing in workers' skills. In such case, the agglomeration of skilled individuals in productive locations boost aggregate human capital accumulation. On the contrary, if the between-skill complementarities dominate, low skill workers gain relatively more from meeting skilled partners. The spatial segmentation of learning interactions prevents frequent interactions between low- and high-skill workers, and thus dampens aggregate learning.[25]

Proposition 3 (Human capital accumulation)**.**
Suppose that $T_\ell \approx \bar{T}$ for all ℓ. If γ is supermodular, $\Omega > 0$ and the average skill of old workers is larger when interactions are spatially segmented. Conversely, γ submodular implies $\Omega < 0$ and the average skill of old workers is lower.

The existence of a tradeoff between spatial inequality and human capital accumulation follows naturally from Proposition 4. An increase in the spatial TFP dispersion leads to greater wage inequality across cities. It also leads to accentuated sorting, and therefore a greater spatial segmentation of learning interactions. Whether the greater concentration of skilled workers is

[25] Appendix A.9 presents the consequences of local interactions on human capital accumulation across the skill distribution. When every worker learns relatively more from skilled interactions, local interactions increase learning inequality across skills by boosting the average learning of skilled workers and dampening that of low-skill individuals.

beneficial for aggregate productivity relies on the learning complementarity (Proposition 4). When within-skill learning complementarities prevail, local interactions generate an equity-efficiency tradeoff: a larger spatial TFP dispersion improves productivity at the cost of greater spatial wage inequality. Such tradeoff does not exist when the between-skill learning complementarities are stronger. In such a case, a policy that reduces the between-city TFP gaps also contribute to increase the aggregate stock of human capital.

Corollary 4 (Equity-efficiency tradeoff)**.**
A greater spatial TFP dispersion increases between-city wage inequality amongst young workers. It also implies a higher (lower) average skill of old workers when γ is supermodular (submodular).

Learning complementarities therefore shape the tradeoff between spatial inequality and aggregate productivity. They also matter to design the optimal spatial policies in the presence of local learning spillovers.

2.5 Optimal Policies

Local interactions generate externalities as workers do not internalize that their location decision shapes others' interactions. The competitive equilibrium is generically inefficient, and spatial policies can improve aggregate human capital accumulation and welfare.

I solve for the allocation chosen by a utilitarian planner. I assume that the planner has homogeneous Pareto weights across skills and cities to focus on the learning externalities rather than on redistributive concerns. I further suppose that the planner cannot condition its allocation on workers' idiosyncratic location preferences, either because these preferences are unobserved or because it is not politically feasible to do so. This restriction is standard in models with idiosyncratic location preferences (e.g. Rossi-Hansberg et al., 2019; Fajgelbaum and Gaubert, 2020).

Given these restrictions, the planner chooses a sequence of spatial allocation of workers, $\{n^{y\star}_{t\ell}(s), n^{o\star}_{t\ell}(s)\}_{t,\ell}$, together with a sequence of consumption allocation across cities and skills, $\{c^{y\star}_{t\ell}(s), c^{o\star}_{t\ell}(s)\}_{t,\ell}$, to maximize the discounted sum of each cohort's expected lifetime utility,[26]

$$\sum_{\ell} \int \mathbb{E}[V^{o\star}_0(s,\boldsymbol{\varepsilon}) \mid s] n^{o\star}_{0\ell}(s) ds + \sum_{t>0} \beta^t \sum_{\ell} \int \mathbb{E}[V^{y\star}_t(s,\boldsymbol{\varepsilon}) \mid s] n^{y\star}_{t\ell}(s) ds. \tag{11}$$

The function $V^{o\star}_0(s,\boldsymbol{\varepsilon})$ is the utility of old workers with skill s and preferences $\boldsymbol{\varepsilon}$ in the initial period under the planner's allocation. Likewise, $V^{y\star}_t(s,\boldsymbol{\varepsilon})$ is the lifetime utility of young workers in cohort t. The planner cannot observe workers' idiosyncratic preferences, and thus maximizes workers' expected utility. The consumption allocation must not exceed the aggregate output produced in each period.[27] Meanwhile, the spatial allocation of workers must be consistent with the aggregate supply of each skill and workers' idiosyncratic location preferences. These preferences act as an incentive constraint: to increase the supply of workers in a particular city, the planner needs to

[26]By the property of the Gumbell distribution, the expected utility of workers who choose to work in city ℓ is identical to the unconditional expected utility, $\mathbb{E}[V^{o\star}_0(s,\boldsymbol{\varepsilon}) \mid s, \ell \succ \ell' \, \forall \ell' \neq \ell] = \mathbb{E}[V^{o\star}_0(s,\boldsymbol{\varepsilon}) \mid s]$.

[27]Same as for consumers, the planner cannot transfer resources across time.

compensate them for their foregone location preferences by offering them a higher consumption.[28] I solve for the planner's allocation without imposing restrictions on the spatial distribution of TFP. I focus here on the steady state allocation, which I refer to as the (constrained) optimal allocation.[29]

Old workers do not affect the learning of young workers, and therefore do not exert any externalities. As a result, the decentralized allocation of old workers coincides with that of the planner.[30]

Young workers do not internalize that their location choice shapes the learning of others. On the contrary, the planner takes into account these local learning spillovers. Specifically, the difference between the optimal and decentralized consumption of young workers with skill s employed in city ℓ is given by

$$c_\ell^{y\star}(s) - c_\ell^{y}(s) = \tau_\ell^\star(s) - \sum_{\ell'} \left(\frac{n_{\ell'}^{y\star}(s)}{n^y(s)}\right) \tau_{\ell'}^\star(s). \tag{12}$$

The right-hand side of (12) is the wedge between the social and private incentives of young workers with skill s to locate in city ℓ. When it is positive, the decentralized equilibrium features too few of these workers in location ℓ. This wedge is given by the net welfare gains (or losses) brought by a marginal increase in the number of skill s in city ℓ. The first term on the right-hand side of (12), $\tau_\ell^\star(s)$, captures the welfare gains of this reallocation for city ℓ alone. The second term is the employment-weighted average of the welfare gains for all the locations. The welfare gains of increasing the presence of skill s for workers in city ℓ is given by

$$\tau_\ell^\star(s) = \beta \int\int \left(\mathcal{V}^o[e\gamma(\sigma, s)] - \mathbb{E}[\mathcal{V}^o(e\gamma(\sigma, S_\ell^{y\star}))]\right) dF(e) \pi_\ell^{y\star}(\sigma) d\sigma, \tag{13}$$

where $\mathbb{E}[\mathcal{V}^o(e\gamma(\sigma, S_\ell^{y\star}))] = \int \mathcal{V}^o[e\gamma(\sigma, s_p)]\pi_\ell^{y\star}(s_p) ds_p$ is the expected future utility of working in city ℓ for skill s. The integrand in (13) is positive when skill σ learns more from s than from the average worker present in the city. The welfare value of increasing the presence of skill s in city ℓ is the weighted average of these skill-specific learning gains across the local skill distribution. Combining (12) and (13), the optimal spatial allocation features a greater share of skill s in city ℓ if interactions with that skill are particularly beneficial to the workers employed in that city.

The decentralized equilibrium is therefore inefficient to the extent that local interactions are a source of learning. When workers' learning depend solely on their own skill, $\gamma(s, s_p) = \gamma(s)$, the spatial distribution of young workers is irrelevant for human capital accumulation, and $c_\ell^{y\star}(s) = c_\ell^y(s)$. Otherwise, the optimal and decentralized spatial allocation of young workers differ from one another.

Characterizing the spatial misallocation in the decentralized equilibrium amounts to solving for

[28] Appendix B.1 describes the full set of constraints faced by the planner.

[29] As for the decentralized equilibrium, the planner's allocation converges to its steady state in one period.

[30] The efficiency of the old allocation highlights that, absent learning externalities, the planner has no incentives to redistribute resources across cities. Idiosyncratic location preferences alone thus do not create inefficiencies. While this may appear in contradiction with the conclusions of Fajgelbaum and Gaubert (2020), the two results are reconciled by noting the absence of local non-traded goods in the present framework. The marginal utilities of consumption are then trivially equalized across space, and the planner does not redistribute across space.

(13). The function $\tau_\ell^\star(s)$ appears on both side of the equation as it affects the spatial allocation of young workers $\pi_\ell^{y\star}$ through the consumption allocation (12). The optimal allocation is therefore the solution to a non-linear infinitely dimensional fixed point. As in Sections 2.3 and 2.4, I simplify this fixed point by studying the optimal allocation when the spatial TFP gaps are small, $T_\ell \approx \bar{T}$. Appendix B.3 derives the first-order approximation of the optimal allocation.

I find that local learning interactions generate spatial misallocation through two channels. First, productive cities may be too large in the decentralized equilibrium. When skilled interactions are relatively more productive, young workers agglomerate in productive cities (Proposition 2.1). In doing so, they forego their idiosyncratic location preferences for less productive locations. However, high-TFP cities do not offer *ex-ante* better learning opportunities, and productive interactions could *a priori* take place anywhere.[31] Holding learning complementarities constant, the planner thus desires to decentralize learning centers. When spatial TFP gaps are small, this is achieved by reallocating workers from high to low TFP locations independently of their skills.

Second, the decentralized equilibrium may feature too much or too little spatial skill segregation. When the within-skill learning complementarities dominate the between-skill complementarities, skilled workers benefit disproportionately from interacting with their own peers. The planner takes into account these learning spillovers, and increases the frequency of interactions between relatively skilled workers by augmenting the concentration of skilled workers in productive cities. On the contrary, relatively low-skill workers experience too little skilled interactions in the competitive equilibrium when the between-skill complementarities prevail. The planner corrects this misallocation by reallocating skilled workers towards low-TFP cities that are abundant in low skills. In both cases, the optimal allocation corrects the learning externalities and increases the aggregate stock of human capital.

Proposition 5 (Optimal allocation)**.**
The decentralized equilibrium is efficient if interactions do not shape learning, $\gamma(s, s_p) = \gamma(s)$. Otherwise, the decentralized equilibrium is inefficient. Furthermore, when $T_\ell \approx \bar{T}$ for all ℓ:

1. *If $\gamma(s, \cdot)$ is increasing for all s, productive cities are too large in the decentralized equilibrium: $N_\ell^{y\star} < N_\ell^y \iff T_\ell > \bar{T}$; conversely, productive cities are too small if $\gamma(s, \cdot)$ is decreasing for all s.*
2. *If γ is supermodular, there is too little young skill concentration in productive cities: $\mathbb{E}[S_\ell^{y\star}] > \mathbb{E}[S_\ell^y] \iff T_\ell > \bar{T}$; conversely, there is too much young skill concentration in productive cities when γ is submodular.*

The optimal allocation implies a larger stock of human capital, $\mathbb{E}[S^{o\star}] > \mathbb{E}[S^o]$.

Finally, which policy instruments can decentralize the optimal allocation? Equation (13) implies that efficiency is restored through place-by-skill transfers, $t_\ell^y(s)$, equal to the right-hand side of (13). These transfers complement workers' wage, and the income of young workers with skill s employed in city ℓ becomes $I_\ell^y(s) = w_\ell(s) + t_\ell^y(s)$.[32] When the spatial TFP gaps are small, the transfers can

[31] In a different setting, Waldfogel (2003) coined this externality the "preference externality".
[32] These transfers can alternatively be expressed as city-specific income tax schedules.

further be decomposed into two components (equation (48) in Appendix B.3). First, a city-specific tax which corrects the size distortion found in Proposition 5.1. Second, skill-by-city subsidies which generate the optimal amount of spatial skill segregation (Proposition 5.2).

Learning complementarities therefore shape both the tradeoff between spatial inequality and human capital accumulation and the optimal spatial policies. In Section 3, I use French administrative data to estimate these learning complementarities. Before doing so, I add several quantitative extensions to bring the framework to the data.

2.6 Quantitative model

I add two key elements to the model, housing and migration costs, that are frequently mentioned to explain the presence of spatial inequality and justify the need for local policies (e.g. Chetty et al., 2014). I also let young workers interact with old partners. Appendix D provides a formal description of the quantitative model.

Housing In addition to the freely traded good, households consume a local, non-tradable good sold at (rental) price p_ℓ. I refer to this good as housing. Households have Cobb-Douglas preferences over the two goods, with α representing the housing expenditure share.[33] In each city, local land-owners supply a stock of housing $H_\ell = \mathcal{H} p_\ell^\delta$, for $\mathcal{H}$ a constant and δ the housing supply elasticity.[34] I take these land owners to be absentee, and in particular exclude them from any welfare calculation.[35]

Migration costs While workers remain free to change location whenever they please, they now incur a migration cost upon moving. I let these migration costs vary across the lifecycle, e.g. if young workers face tighter borrowing constraint or if old workers have accumulated more social capital. I assume that workers are born in the location where their parents lived when young; that is, the probability to be born in city ℓ is given by N_ℓ^y.[36] A young worker born in city ℓ that moves to city ℓ' incurs a utility cost $\kappa_{\ell\ell'}^y$. Likewise, an individual who lived in ℓ in their youth and migrates to location ℓ' in their old period pays a cost $\kappa_{\ell\ell'}^o$. I assume that the migration costs are symmetric, $\kappa_{\ell\ell'}^a = \kappa_{\ell'\ell}^a$ for $a \in \{y, o\}$, and I normalize $\kappa_{\ell\ell}^a = 0$ for every ℓ and $a \in \{y, o\}$. The presence of migration costs implies that the expected lifetime utility of young workers now depends on their skill and their birthplace. In particular, expected utility is no longer equalized across space.

Learning I also relax the assumption that young workers interact only between themselves, and let young workers interact both with young and old workers. Furthermore, I assume that the

[33] For empirical evidence using U.S. data in support of the constant housing expenditure share implied by the Cobb–Douglas functional form, Davis and Ortalo-Magné (2011).

[34] This reduced-form supply function can be micro-founded by assuming that land-owners have to provide effort subject to convex cost to turn the land they own into rentable housing.

[35] The model would have the same positive predictions if I were to assume that land-owners use their revenues to consume the freely traded good. Abstracting from them in the welfare computation would however be inconsistent.

[36] I could alternatively assume that workers' birthplace is given by their parents' location when old. The probability to be born in ℓ would be N_ℓ^o. In the data, I define a young as a worker between 25 and 40 year old. I stick with the first formulation to stick closer to the empirical timing of birth. This will not matter quantitatively since $N_\ell^y \approx N_\ell^o$.

learning technology takes the form

$$\log \gamma(s, s_p) = g_0 + g_1 \log s + g_2 \log s_p + g_{12} \log s \log s_p. \tag{14}$$

The parameter g_1 dictates whether skilled workers are marginally better to learn on their own, whereas g_2 determines the marginal gains from a skilled interaction. Meanwhile, g_{12} determines how the marginal gains from a skilled interaction differ across workers.

The learning technology (14) differs from (1) as it is expressed in logs rather than in levels. While the latter functional form was useful to exemplify the between and within-skill learning complementarities, it is harder to bring to the data because it does not feature any curvature in learning.[37] On the contrary, (14) allows for a direct estimation of (g_1, g_2, g_{12}) from wage growth data. In addition, this functional form is more flexible as it entertains learning technologies that are neither super- nor submodular.

Distributions Finally, I suppose that the skill distribution of young workers is log-normal with mean μ_y and standard deviation σ_y, and I parametrize the idiosyncratic learning shocks to be log-normally distributed with zero mean and standard deviation σ_e.[38]

Relationship to theory To what extent the results derived in Sections 2.3 to 2.5 hold in the quantitative model? The introduction of migration costs and housing consumption alters the analytical tractability of the model. Rather, I rely on numerical methods to quantify the impact of local interactions on spatial inequality and human capital accumulation. In Section 5, I compare the estimated framework with a counterfactual economy in which interactions are not spatially segmented. Qualitatively, the numerical results align with the predictions contained in Proposition 2 and 4.

3 Estimating the Returns to Local Interactions

In this section, I present evidence on the role of cities in shaping workers' wage growth. I argue that this evidence can be used to estimate the learning technology. I present the data in Section 3.1 and provide some descriptive evidence about between-city differences in wage growth in Section 3.2. Section 3.3 derives the empirical design from the quantitative model, and Section 3.5 concludes with the estimation of the learning technology.

[37] Alternatively, one could include higher order terms in (1) to introduce curvature in the learning technology. This would however require to estimate many more parameters and open the door to over-fitting concerns.

[38] Normalizing the mean of the learning shocks distribution to zero is without loss of generality since it is not separately identified from g_0 in the learning technology.

3.1 Data

I provide evidence on spatial differences in wage growth using French matched employer-employee data (DADS).[39] This dataset comes in two formats. The first format is a 4% representative panel that tracks the entire history of individuals on the labor market (DADS panel).[40] The second format is a repeated yearly cross-section covering the universe of employed workers (DADS salariés). Both datasets provide information on workers' wages, the number of hours worked, the location where they work and live, along with demographic information. The data goes back to 1976 but frequent changes of variables make it difficult to use the observations collected prior to 1993.

This dataset has several advantages over other data sources used in the literature. First, its panel dimension allows me to measure individual wage growth at a long-term horizon. Second, its cross-sectional dimension is useful to measure precisely skill density at a granular geographical level. Third, it is one of the few matched employer-employee dataset that provides joint information on where individuals work (firm, establishment, and neighborhood) and where they live.

I apply the same sample restrictions on both the panel and the cross-sectional datasets. I focus my analysis on full-time employed workers between 25 and 55 year old. Workers employed in the public sector in France have their wage determined by their tenure, and as such, their wage may not fully reflect their productivity. I therefore keep in the sample workers employed in the private sector, and I exclude the education and health industries due to the large fraction of workers that are public servants in these sectors. I measure workers' wages by their gross yearly salaries divided by the number of hours worked, and I truncate the hourly wage distribution at the 5% and 99.9% percentile to minimize measurement errors in hours worked.[41] Appendix E.1 provides further details on the construction of the sample.

Over time, wages grow for two reasons. First, workers experience idiosyncratic wage growth across their lifecycle. Second, wages grow for all workers as the economy expand.[42] Aggregate growth is absent from the framework laid down in Section 2. To abstract from it empirically, I normalize the average log wage to zero every year. These aggregate trajectories may also differ across space, e.g. for cities impacted by the deindustrialization of the French economy. I therefore focus the study to the 2009-2019 time window. In that decade, aggregate growth was evenly spatially distributed in France (see Figure E.2), guaranteeing that I do not conflate faster lifecycle wage growth with faster aggregate growth.

I define a city as a commuting zone. A commuting zone is a statistical area defined by French statistical area (INSEE). It consists of a collection of contingent municipalities clustered together to

[39]DADS stands for "*Declarations Annuelles de Données Sociales*". The dataset is constructed from employer tax records which are compiled by the French statistical agency (INSEE).

[40]Individuals' employment history is recorded in the dataset if (a) they have at least one employment spell, and (b) they are born in the first fourth days of each quarter.

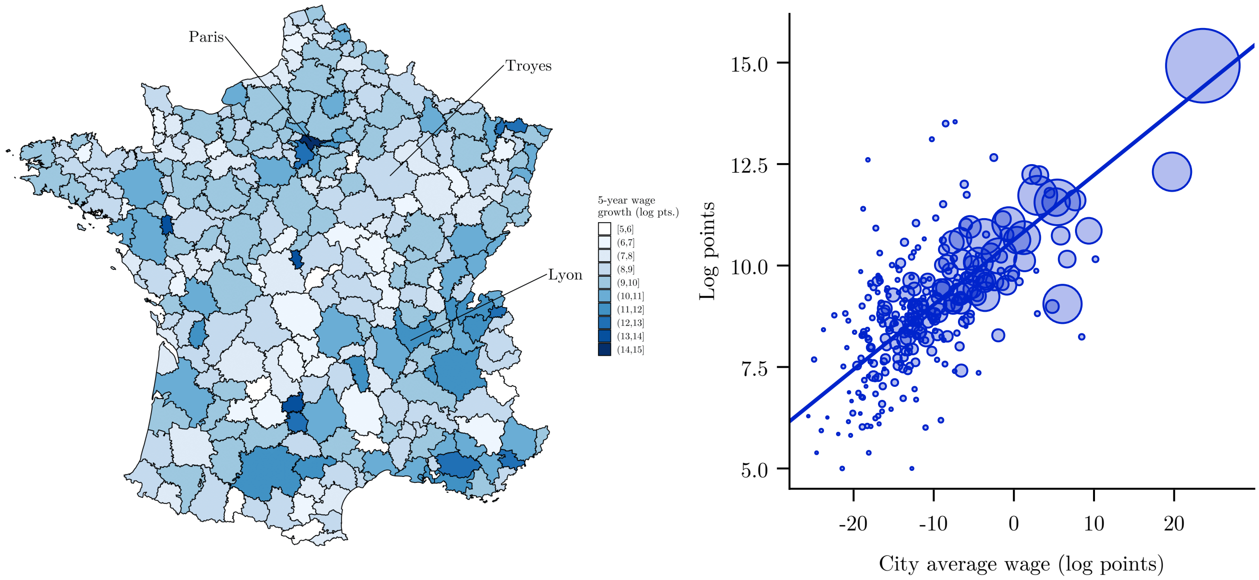
[41]In France, between 5% and 10% of the overall workforce is paid at the minimum wage (INSEE, 2021). For these workers, wages are not a valid proxy of their skills. I focus on workers paid above the minimum wage by truncating the left-tail of the wage distribution at the 5%.

[42]This relationship is mechanical. The log wage growth of worker i can always be decomposed $\omega_{it+1} - \omega_{it} = \tilde{\omega}_{it+1} - \tilde{\omega}_{it} + \mathbb{E}[\omega_{it+1}] - \mathbb{E}[\omega_{it}]$, for $\omega_{it} = \log w_{it}$ and $\tilde{\omega}_{it} \equiv \omega_{it} - \mathbb{E}[\omega_{it}]$. The first term is the worker-level wage growth and the second is aggregate wage growth.

Figure 1: The geography of wage growth

(a) Average wage growth by commuting zone

(b) Wage growth and local wages



Note: left-panel plots the average log wage growth at the five-year horizon by commuting zones for workers under 40 year old. Right-panel plots the average log wage growth at the five-year horizon against the average wage of the commuting zones. The size of the markers is proportional to the city size. In both panels, wage growth is measured at the worker level, i.e. $\log w_{it+1}/w_{it}$.

reduce the commuting flows across them. There are 297 such areas in metropolitan France, which span the whole territory.[43] Commuting zones are relatively large, with an average size of 40,000 employed workers. Commuting zones are the natural geographical unit when thinking of local labor markets. In addition, their relatively large size are well-suited to capture all the interactions that a worker may experience locally. I use the terms commuting zones and cities interchangeably for the rest of the paper.

I also use a second, more granular geographical unit for the empirical analysis. Specifically, I define a neighborhood as a municipality ("commune" in French). Municipalities are legal areas which also span the French metropolitan territory. They are however smaller than commuting zones: in 2019, France was split in 34,839 municipalities with an average size of 1,800 inhabitants and an average area of 15 square kilometers.[44] They compare to ZIP codes in the United States. Figure E.1 displays a map of the French commuting zones alongside the municipalities contained in Paris and Lyon to highlight how granular these spatial units are. I restrict the analysis to metropolitan municipalities with more than 50 employed workers. I am left with 12,320 municipalities, representing 99% of the French workforce and with an average size of 950 employed workers.

In the next section, I use the matched employer-employee dataset to provide descriptive evidence on the differences of wage growth across space. I then present how this evidence can be used to

[43]Metropolitan France excludes the French overseas regions (Guadeloupe, Réunion, etc.).

[44]Municipalities were introduced with the French revolution in 1789. Since the 20$^{\text{th}}$ century, the number of municipalities have been rather stable over time. Appendix E.2 provides more details on their history and characteristics.

estimate the learning technology.

3.2 The Geography of Wage Growth

Figure 1a plots the average five-year wage growth at the worker-level by commuting zones. Consistent with the model, I restrict the analysis to "young" workers under 40 years old. On average, young workers experience a wage growth of 10.6 log points over five years. This lifecycle wage growth is very unevenly distributed across space.[45] Workers in Paris see their wage increase by 15 log points. Meanwhile, the average wage growth in Lyon, the city in the 75$^{\text{th}}$ percentile of the growth distribution and the second biggest French city, is 12 log points, and the growth in Troyes, the city in the 10$^{\text{th}}$ percentile of the growth distribution, is 8 log points. In total, half of the cities (20% of the population) offer wage growth below 8.8 log points.

Figure 1b displays how the local wage growth covaries with the city average wage. On average, workers in cities with higher wages tend to experience faster wage growth across the lifecycle. In the cross-section, an increase of the city wage by one standard deviation is associated with an increase of wage growth by 2 log points at the five-year horizon. This relationship is not mechanically caused by faster aggregate growth at the city level (Figure E.2) or by housing prices dynamics (Figure E.3).

3.3 Identification

In the framework laid down in Section 2, local interactions engender skill growth. The returns to these interactions depend on the learning technology. Consequently, spatial variations in skill growth are informative about the learning technology.

Specifically, given the learning technology (14), the next period's skill of a worker i given their present skill, s_{it}, and the skill of their learning partner, $s_{p_{it}}$, is

$$\log s_{it+1} = g_0 + g_1 \log s_{it} + g_2 \log s_{p_{it}} + g_{12} \log s_{it} \log s_{p_{it}} + \log e_{it+1}, \tag{15}$$

where e_{it+1} is the idiosyncratic learning shock experienced by i. These shocks represent for instance enrollments in on-the-job training programs or strokes of genius. They are however identically and independently distributed across space, workers and partners. Through the lens of the model, the learning technology can therefore be estimated via a local projection of future skills on today's skills, both for i and their partner.

Estimating (15) poses however two empirical challenges: how to observe learning interactions, p_{it}, and how to measure skill, s_{it}. I use the structure of the model to solve these challenges.

First, very few datasets track workers' interactions, and even less the skills or wages of the workers engaged in them.[46]. However, when interactions are random within cities, the average skill of individuals working in the same location as i constitutes a valid proxy for their interactions.

[45]While the between-city difference in growth are large, they explain a small fraction (4%) of the total variance.

[46]Recent papers have made significant advances in how to measure face-to-face interactions (e.g. Atkin et al., 2022; Emanuel et al., 2023). Yet, they lack a measure of worker productivity that they can track over time.

Specifically, substituting $s_{p_{it}}$ in (15) for the average skill in location ℓ_{it}, $\mathbb{E}[\log s_{it}]$, returns

$$\log s_{it+1} = g_0 + g_1 \log s_{it} + g_2 \mathbb{E}_{\ell_{it}}[\log s_{it}] + g_{12} \log s_{it} \mathbb{E}_{\ell_{it}}[\log s_{it}] + \log e_{it+1} + \nu_{it}. \tag{16}$$

The term $\nu_{it} = (g_2 + g_{12} \log s_{it})(\log s_{p_{it}} - \mathbb{E}_{\ell_{it}}[\log s_{it}])$ captures the gap between the particular interaction experienced by worker i and the average interaction in city ℓ. Random interactions imply that ν_{it} is orthogonal to s_{it} and $\mathbb{E}_{\ell_{it}}[\log s]$ (see Appendix E.3).

Second, in the theory, skill and wages are related through $w_{it} = s_{it} T_{\ell_{it}}$, for $T_{\ell_{it}}$ the TFP of the city where i works.[47] While skills are rarely measured, wages are easily observed. Substituting wages for skills in (16), the theory predicts that future wages are related to today's wages through

$$\log w_{it+1} = g_0 + g_1 \log w_{it} + g_2 \mathbb{E}_{\ell_{it}}[\log w_{it}] + g_{12} \log w_{it} \mathbb{E}_{\ell_{it}}[\log w_{it}] + \log e_{it+1} + \nu_{it} + \zeta_{it}. \tag{17}$$

The added term, $\zeta_{it} = \log T_{\ell_{it+1}} - (g_1 + g_2) \log T_{\ell_{it}} - g_{12} \left(\log w_{it} + \mathbb{E}_{\ell_{it}}[\log s_{it}] - \log T_{\ell_{it}}\right) \log T_{\ell_{it}}$, captures the difference between workers' skills and wages. When the spatial dispersion in TFP is small, wages closely track skills and ζ_{it}.[48] Otherwise, not controlling for this gap could bias the learning technology estimates. I thus estimate (17) in two steps. First, I estimate the local projection without controlling for ζ_{it}. Given the estimates of the learning technology, I use the model to recover cities' TFP (Section 4.3). In the second step, I re-estimate (17) controlling for the terms in ζ_{it}.[49] In practice, I find that the spatial dispersion in TFP is small and explain less than 2% of the wage variance. The learning technology estimates are therefore very similar in both steps.

From equation (17), the learning technology can therefore be estimated from a local projection of future wages on current wages, the average wage of the location where workers are employed, and the control for cities' TFP. This local projection yields unbiased estimates of the learning technology to the extent that the error terms e_{it+1} and ν_{it} are uncorrelated with the regressors. The following model-consistent assumptions ensure unbiased estimates.

Assumption 2 (Learning process)**.**

1. *Interactions are random within cities,* $(w_{it}, w_{p_{it}}) \perp \mathbb{E}_{\ell_{it}}[\log w]$*;*
2. *Idiosyncratic learning shocks are i.i.d. across workers and cities,* $e_{it} \perp (w_{it}, \mathbb{E}_{\ell_{it}}[\log w_{it}])$*.*

Under Assumption 2, the local projection

$$\log w_{it+1} = \alpha_t + \beta \log w_{it} + \gamma \overline{\log w}_{\ell_{it}} + \delta \log w_{it} \overline{\log w}_{\ell_{it}} + \zeta_{it} + u_{it} \tag{18}$$

produces unbiased estimates of the learning technology, where α_t is a year fixed effect that controls for aggregate trends in wages and $\overline{\log w_\ell}$ is the average log wage in city ℓ. I measure future wages at the five-year horizon to smooth out transitory wage shocks while keeping sufficient statistical power.

[47] As a robustness, in Section 3.5 I control for work characteristics (e.g. industry or occupation) that may affect wages beyond workers' skills.

[48] Recall that mean log wages are normalized to zero. Hence, small TFP differentials across cities imply $\log T_\ell \approx 0$ and $\zeta_{it} \approx 0$.

[49] Controlling for the terms in ζ_{it} is of course identical to residualizing workers' wages by cities' TFP.

I compute city wages on the right-hand side of (18) on all workers between 25 and 55 year old, but I only include in the regression workers under 40.

Similar local projections have been used in the learning from coworkers literature. For instance, Nix (2020) and Jarosch et al. (2021) project respectively future wages on the average education within the firm and the average wage within the establishment where workers are employed. My empirical design differ from theirs in two ways. First, it studies the returns to local interactions within cities rather than within firms. Second, it includes an interaction term between workers' wage and the wage of the city where they work to estimate learning complementarities.

3.4 Endogeneity concerns

Assumption 2 may not hold empirically. Skilled workers may be able to target interactions with skilled partners (violation of Assumption 2.1). The degree of learning complementarities estimated by (18) would then be biased upward, $|\delta| > |g_{12}|$. To investigate how large is this bias, I estimate (18) separately for small and large cities as well as for more or less unequal locations. To the extent that skilled workers can better target skilled interactions in larger or more unequal cities, the estimated learning complementarity should be higher in those places. More generally, the difference in the estimated learning technology across city groups provide suggestive evidence on the validity of Assumption 2.1.

Meanwhile, the distribution of learning shocks may vary across workers and cities (violation of Assumption 2.2). For instance, young workers may learn faster for reasons other than local interactions. The assumption $e_{it} \perp w_{it}$ would then break to the extent that young workers also earn lower wages. Meanwhile, the faster wage growth of productive cities may be due to other form of agglomerations than local interactions, e.g. a steeper local job ladder, in which case $e_{it} \not\perp \mathbb{E}_{\ell_{it}}[\log w_{it}]$.

I alleviate this second set of endogeneity concerns in three ways. First, I control for worker-level determinants of wage growth (Adda and Dustmann, 2023). I include age fixed effects to control for the faster growth of younger workers. To control for long-term contracting and decreasing returns to experience, I add fixed effects for tenure at the job, at the occupation, and at the industry level. Finally, to account for the effect of job switching on wage dynamics, I include a dummy for employer switching.

Second, I control for unobserved spatial heterogeneity in learning shocks through a set of granular spatial fixed effects. Specifically, I estimate the modified local projection

$$\log w_{it+1} = \alpha^R_{tn^r_{it}} + \alpha^W_{tg_{it}} + \beta \log w_{it} + \gamma \overline{\log w}_{n^w_{it}} + \delta \log w_{it} \overline{\log w}_{n^w_{it}} + \boldsymbol{\Psi}' \boldsymbol{X}_{it} + u_{it}. \tag{19}$$

In the above, n^r_{it} and n^w_{it} are the neighborhoods where worker i resides and works. The parameters $\alpha^R_{tn^r}$ are therefore year by neighborhood of residence fixed effects, and $\overline{\log w}_{n^w}$ is the average wage of the workers employed in neighborhood n^w. Meanwhile, $\alpha^W_{tg_{it}}$ are fixed effects that control for the work environment of worker i. I adopt two different specifications for these fixed effects. First, I set g_{it} to be the city where i is employed. In this case, (19) effectively compares two workers who live in the same neighborhood but work in different neighborhoods in the same city. Second, I set g_{it} as the firm-city where i is employed. Then, (19) compares workers who live in the same neighborhood and work in the same firm but in establishments located in different neighborhoods.[50]

The two sets of fixed effects control flexibly for unobserved heterogeneity in wage dynamics. The residence fixed effects take into account that workers who live in wealthier neighborhoods may enjoy steeper wage profiles, e.g. due to a better access to public transportation. Meanwhile, productive firms may offer higher wages and more lucrative promotions to its workers, which is controlled for by the city-firm fixed effects. Finally, by comparing workers who live in the same neighborhood and work in the same firm, the fixed effects also reduce workers' unobserved heterogeneity.[51]

Third and last, I rely on an instrumental variable strategy to generate quasi-random variations in skill density across neighborhoods. Equations (18) and (19) are effectively peer effect models. These models are generally subject to a reflection problem (Manski, 1993; Angrist, 2014). Individuals working in the same location may indeed be exposed to long-lasting productivity shocks that generate a spurious correlation between their present and future wages. Assumption 2.1 rules out local correlated shocks, which allows (18) to produce unbiased estimate of the learning technology.[52] When Assumption 2.1 is relaxed, it is not possible to measure the quality of local interactions through average wages without running into the reflection problem.

I therefore identify spatial variations in skill density that are plausibly unrelated to local productivity shocks. To do so, I instrument the neighborhood's average wage by the past change in the white-collar employment share. I define white-collar workers through their occupations (e.g. engineers, managers). I consider the past change in their employment share between 1993 and 2000. By isolating the variation in local skill density coming from past changes in the occupation composition of neighborhoods, the instrument estimates the returns to local interactions net of any time-invariant agglomeration forces (Ahlfeldt et al., 2015).

The instrument is relevant to the extent that past changes in the occupational mix of neighborhoods have long-term consequences on their skill composition. Meanwhile, the exclusion restriction holds if the underlying shocks that triggered these changes are not correlated with the present productivity shocks.[53]

I provide indirect evidence in support of the relevance and exclusion restrictions in Table E.2. First, past changes in the white-collar employment share are strongly correlated with the current employment share (column 2). Neighborhoods that have seen their white-collar employment share increase by one percentage point (p.p.) between 1993 and 2000 have on average an employment share

[50]The city and city-firm fixed effects also allows to relax Assumption 2.2.

[51]An alternative is to include worker fixed effects in (19). In this case, present wages need to be instrumented à la Arellano and Bond (1991). However, this instrument requires $T \geq 3$. In my setting, by using five-year wage growth and restricting my sample to the 2009-2019 period, most of my individuals are observed for $T \leq 3$, making it difficult to obtain precise estimates of the learning technology.

[52]Table E.1 further confirms through Monte-Carlo simulation that the reflection problem is not present in (18). See Appendix E.3 for further discussion.

[53]Changes in the local skill composition may also affect neighborhood characteristics other than local interactions (e.g. the stock of physical capital), and indirectly through those wage growth. I cannot separate between the "direct" and the "indirect" channels. However, to the extent that both channels capture how the local skill composition shapes workers' human capital accumulation, I view as consistent with the model laid down in Section 2.

0.3 p.p. (0.05 standard deviation) higher in 2010. Meanwhile, there is no statistical relationship between the change in the white-collar employment share between 1993 and 2000 and the change between 2010 and 2019 (column 4).

Endowed with this instrument variable, I interpret γ in (19) as a reduced-form parameter that measures the average treatment effect of local skill composition on skill growth. Similarly, I interpret δ as a reduced-form parameter that captures the heterogeneity in treatment effects across the skill distribution. How to interpret structurally these estimates? Through the lens of Assumption 2, the controls and fixed effects present in (19) introduce structural mis-specification – the so-called "bad controls" (Angrist and Pischke, 2009). For instance, if workers' learning is shaped by the skill composition of the neighborhood *and* the city where they work, the city fixed effects in (19) will soak all the city-level variations and produce downward-biased estimates of g_2 and g_{12}. Interpreting structurally the gap between the estimates produced by (18) and (19) is therefore complicated. At the same time, if interactions with relatively skilled workers spur faster learning regardless of where these interactions take place, the city fixed effects will produce estimates of γ and δ whose sign is consistent with g_2 and g_{12}. I therefore test whether Assumption 2.2 introduces large biases by comparing the signs of the estimates produced by (18) and (19).

3.5 Local Wage Growth and Skill Density

Table 1 presents the results of the estimation of (18). Standard errors are clustered at the city level. The first column corresponds to the specification without the controls for city TFP. The second columns control for ζ_{it}. In both cases, the three point estimates of the learning technology are positive and significant at the 0.1%. In addition, the two sets of estimates are very similar and are not statistically different from each other at the 5%. I take column 2 as the baseline specification. Three predictions emerge out of it.

First, high-wage workers enjoy faster growth ($\hat{\beta} > 1$): an increase of wages by one standard deviation boosts wage growth by 0.4 log points over five years for workers in the average city.[54] Workers starting their career with lower wages therefore do not catch-up to high-wage individuals. The estimated wage process is therefore non-stationary.[55] This result rests however on the time horizon considered. In Figure E.4, I estimate the local projection for time horizons between one and five years. For shorter-term wage growth, I do estimate an auto-correlation parameter β close but under the unit root, in accordance with estimates from the literature (e.g. Guvenen, 2009). I take this as indirect evidence that, by considering longer-term variation, I estimate a learning technology that captures permanent variations in human capital rather than transitory, mean-reverting shocks. Quantitatively, this effect alone is small.

Second, individuals employed in high-wage cities experience faster growth ($\hat{\gamma} > 0$). An increase

[54]The average wage is normalized to zero, and therefore so is the wage of the average (employment-weighted) French city. Hence, $\beta - 1$ represents the marginal effect of increasing workers' wages on wage growth for workers in the average French city. Similarly, γ estimates the marginal growth effect of increasing the city's average for the average French worker.

[55]The non-stationarity of the wage process is not a problem in my setting due to the overlapping generation structure that guarantees the existence of an ergodic skill distribution.

Table 1: The returns to local interactions at the city level

Dep. variable: 5-year log wage	(1)	(2)	(3)	(4)
Log wage (g_1)	1.010 (0.003)	1.010 (0.003)	1.003 (0.002)	1.010 (0.003)
City wage (g_2)	0.122 (0.005)	0.107 (0.015)	0.119 (0.010)	0.095 (0.012)
Log wage × city wage (g_{12})	0.102 (0.008)	0.108 (0.032)	0.065 (0.019)	0.102 (0.007)
Log city size				0.002 (0.001)
Controls				
City TFP	·	✓	✓	✓
Obs.	1,052,176	1,052,176	1,052,176	1,052,176
Weights	Unit	Unit	1 / size	Unit

Standard errors clustered at the city level. All regressions include age, job tenure, occupation tenure, and industry tenure fixed effects, as well as a dummy if the worker changed employer between the five years. Sample includes all workers between 25 and 40 year old. Wage growth distribution truncated at the 5% within cities. Industry and occupation in column 3 defined at the one-digit level. Firm wage growth in column 4 computed as $\mathbb{E}[\log w_{it+1} \mid f]/\mathbb{E}[\log w_{it} \mid f]$ for firm f. Column 6 controls for ζ_{it} by using the TFP estimates of Section!4. Column 7 instruments for today's wage by the previous year wage. Each fixed effect is interacted with year fixed effects.

by one standard deviation of the city wage implies an additional growth of 1.6 log points for the average worker. Through the lens of the model, this implies that interactions with skilled workers are on average more productive.

Third and last, the growth of high-wage workers in high-wage city is larger than that of low-wage workers in the same location ($\hat{\delta} > 0$). An increase by one standard deviation of the city average wage implies an additional growth of 1 log points for workers at the 10th percentile of the wage distribution, compared to an increase of 2.6 log points for workers at the 90th percentile. From (18), this implies that the learning technology displays strong within-skill complementarities: skilled workers gain relatively more from interacting with skilled partners than their low skill counterpart.

Columns 3 and 4 investigate the role of large cities in shaping the estimated parameters.[56] A vast literature, started by de la Roca and Puga (2017), finds indeed that larger cities boost workers' wage growth. In column 3, I thus control for city size to net out the correlation between city size and local wages. Controlling for city size slightly reduces the point estimates of g_2, although the two estimates are not statistically different at the 5%. In accordance with the previous literature, I also find that workers in larger city experience faster growth. However, the effect of city size is one order of magnitude smaller than the effect of skill composition: a one standard deviation increase in size implies a 0.4 log points increase in growth for the average worker.[57] The difference in

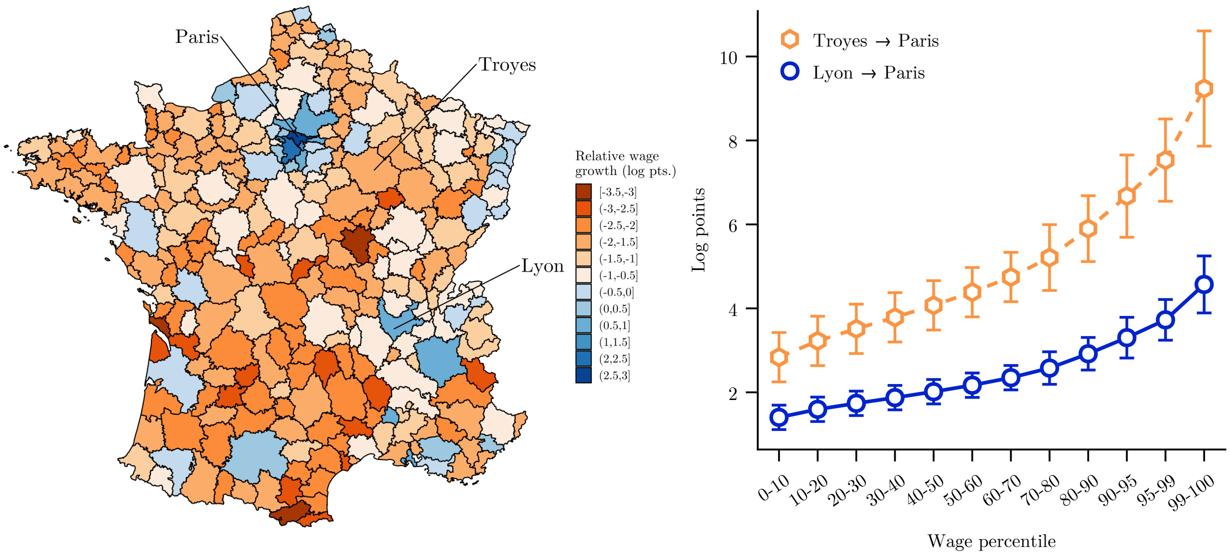
[56] In both cases, I do not control for cities' TFP. Controlling for cities' TFP is indeed computationally costly since it requires to solve the entire estimation routine presented in Section 4. Since the estimates of the first and second columns of Table 1 were not statistically different from each other, I abstract from this computation cost.

[57] Most papers in this literature, to the exception of Eckert et al. (2022), identify the effect of size on growth from spatial correlation between city size and wage growth. In fact, the empirical model of de la Roca and Puga (2017) coincides with (18) under the restriction that $\beta = \delta = 1$.

Figure 2: The estimated growth gaps across space and workers

(a) Relative wage growth for average worker

(b) Growth consequences of moving to Paris



Note: left-panel displays the relative average growth by commuting zone for the average French worker, $\hat{g}_2 \mathbb{E}_{\ell_{it}}[\log w_{it}]$. Right-panel displays the predicted growth difference between Paris and Troyes (orange hexagons) and Paris and Lyon (blue circles) across the wage distribution, $(\hat{g}_2 + \hat{g}_{12} \log w_{it})\,(\mathbb{E}_{\ell'}[\log w_{it}] - \mathbb{E}_{\ell}[\log w_{it}])$. Both set of statistics are computed using the estimates of column 1 in Table 1.

magnitude justifies why I abstract from city size in the quantitative model.[58] In column 4, I weight observations by the inverse of the city size to guarantee that the point estimates are not driven by a few large cities. Here as well, the point estimates are not statically different at the 5% from the baseline estimates. Finally, Figure E.4 displays the point estimates for time horizons between one and five year. The parameters dictating the returns to local interactions, γ and δ, increase with the horizon considered. I take these patterns as suggestive evidence that a longer exposure to local interactions yield higher future wages.[59]

What are the implication of these estimates for the between-city differences in wage growth? The implied spatial growth gaps explain 64% of the between-city variance in wage growth. Figure 2a displays the predicted relative wage growth by commuting zones for the average French worker.[60] The predicted distribution of wage growth remain very uneven across cities. Migrating from Troyes to Paris triggers an increase in wage growth by 4.5 log points per five year, while a move from Lyon to Paris implies wage growth gains of 2.2 log points.

The between-city difference in growth are very heterogeneous across workers due to the estimated complementarities. Figure 2b plots the difference in growth between Paris, Lyon and Troyes across the wage distribution. All workers, both low- and high-wage, experience faster growth when they are

[58] See Duranton and Puga (2023) for a spatial growth model without heterogeneous workers in which size shapes aggregate human capital accumulation.

[59] In (18) and Figure E.4, I do not condition on stayers – which would create a selection bias. However, the yearly migration rate in France is 10%, such that a longer time horizon is strongly correlated with longer exposure to the local, initial interactions.

[60] These estimates are computed as $\hat{\gamma}\mathbb{E}_{\ell}[\log w]$. They correspond to the change in wage growth from a migration for the average French worker. These estimates do not take into account the sorting of workers across cities.

Table 2: The returns to local interactions at the neighborhood level

Dep. variable: 5-year log wage	(1) OLS	(2) OLS	(3) OLS	(4) 2SLS	(5) 2SLS	(6) OLS
Log wage	1.010 (0.003)	1.000 (0.001)	0.935 (0.002)	0.998 (0.002)	0.935 (0.003)	0.975 (0.002)
Local wage	0.111 (0.005)	0.078 (0.008)	0.032 (0.008)	0.099 (0.022)	0.053 (0.021)	0.037 (0.008)
Log wage × local wage	0.104 (0.009)	0.062 (0.014)	0.071 (0.013)	0.078 (0.026)	0.077 (0.023)	
Establishment wage						0.072 (0.006)
Firm wage						0.011 (0.006)
Controls						
City TFP	✓	·	·	·	·	·
Worker-level	✓	✓	✓	✓	✓	✓
Fixed effects						
Residence	·	✓	✓	✓	✓	✓
City	·	✓	·	✓	·	✓
City × firm	·	·	✓	·	✓	·
Obs.	971,401	971,401	974,401	974,401	974,401	508,236
F-stat.	·	·	·	55,244	53,539	·

Standard errors clustered at the city level for column 1 and at the neighborhood of work for the remaining columns. All regressions include age, job tenure, occupation tenure, and industry tenure fixed effects, as well as a dummy if the worker changed employer between the five years. Column 5 and 6 instruments for the neighborhood wage by the change in the white-collar employment share between 1993 and 2000. Sample includes all workers between 25 and 40 year old. Wage growth distribution truncated at the 5% within cities. Industry and occupation in column 3 defined at the one-digit level. Each fixed effect is interacted with year fixed effects.

employed in Paris. High-wage workers benefit however more over time from working in high-wage locations. For instance, the wage growth gap between Paris and Troyes is 2.8 log points for workers in the bottom 10% of the wage distribution. It is more than three times larger for workers in the top 1%. The same pattern holds between Paris and Lyon, although the gaps are smaller as the two cities have a more similar skill composition.

Robustness Table 2 displays the point estimates obtained from (19), gradually adding the controls, the fixed effects and the instrument variable. The first columns introduces the worker-level controls: age, tenure and employer transition fixed effects. The point estimates are not statistically different from the baseline estimates (column 2 in Table 1). The same hold when adding occupation and industry fixed effects or when controlling for transitory wage shocks (see Appendix E.5). The standard drivers of wage growth thus do not bias the estimates of the learning technology.

The second column adds the neighborhood of residence and city fixed effects to control for unobserved spatial heterogeneity in wage growth. Both set of fixed effects are fully interacted with year fixed effects.[61] The parameters γ and δ are then estimated off average wage variations across neighborhoods of work, and standard errors are accordingly clustered at the neighborhood level. The point estimates are smaller than in the baseline specification. For two workers living in the same neighborhood and working in the same city, an increase by one standard deviation of the neighborhood wage implies an additional growth of 1.3 log points for the average worker, and up to 2.2 log points for a worker in the 10% of the wage distribution. However, the three coefficients remain positive: interactions with skilled partners remain more productive, particularly so for skilled individuals, even after controlling for unobserved wage growth heterogeneity across cities and neighborhoods. In addition, the learning complementarities as estimated by δ are not statistically different at the 5% from the baseline estimates.

Column 4 includes the year by city by firm fixed effects to control for heterogeneity in wage dynamics across firms. In addition, to neutralize establishment-specific productivity shocks, e.g. at the headquarter level, I also control for the growth in average wages at the establishment level.[62] The parameters that dictate the returns to local interactions, γ and δ, remain positive and statistically significant different at the 0.1%. That is, comparing two workers employed in the same firm, in the same city, and living in the same neighborhood, the worker employed in the relatively high-wage neighborhood experiences faster wage growth – in particular if these workers earn initially higher wages.

The fifth and sixth columns introduce the instrument variables to address the potential presence of correlated shocks at the neighborhood level. In both cases, the F-statistics are well above the conventional thresholds for weak instruments (Stock and Yogo, 2002; Andrews et al., 2019). I find that the OLS and 2SLS estimates are not statistically different from each other at the 5%. If anything, the 2SLS estimates are slightly larger than the OLS estimates. The downward bias of the OLS estimates could in part be caused by mean-reverting productivity shocks at the neighborhood level.

Finally, Figure E.6 and E.7 provides suggestive evidence that the random interaction assumption (Assumption 2.1) is reasonable. Specifically, Figure E.6 presents the estimates of the learning technology by group of city size, while Figure E.7 shows the estimates by tercile of within-city wage inequality. To the extent that interactions are more segmented in larger and more unequal cities, the estimated learning technology should display stronger skill complementarities in those places. In both cases, no systematic pattern emerges across the size or inequality distribution.

Table 2 therefore validates qualitatively the structural estimates of the learning technology found in Table 1. Furthermore, the estimated learning complementarities, δ, are stable across the several specifications and are statistically different from the baseline estimates at the 5% level. I now turn

[61] Table E.3 in Appendix E.5 further adds city by industry by occupation fixed effects. Beyond controlling for heterogeneity in wage growth across occupations and industries, these fixed effects also allow for skill complementarities, decreasing returns to scale, and productivity spillovers in the production function at the city level.

[62] Alternatively, Table E.3 in Appendix E.5 controls jointly for the average wage of the neighborhood, the establishment, and the firm where workers are employed. I find that individuals working in high-wage neighborhoods experience faster wage growth even after controlling for the average wage of their establishment. I take this as indirect evidence that space matters for human capital accumulation outside the boundary of the firm.

to estimating the remaining parameters of the quantitative model.

4 Model Estimation

The estimation of the quantitative model is split into three parts. In the first part, I define the geography and the time horizon of the model, and I externality calibrate one parameter. In the second part, I invert the model to estimate seven groups of parameters, including the learning technology given the estimates of Table 1. Finally, in the third part, I use indirect inference to calibrate the remaining seven groups of parameters. The estimated parameters are presented in Table F.2.

4.1 Definition and external calibration

Consistent with the evidence of Section 3, I define a location in the model as a commuting zone. Estimating city fundamentals for the 297 commuting zones is computationally complex since the model cannot be aggregated at the city level.

Rather, I create thirty types of cities and restrict all cities within the same type to be homogeneous. Importantly, there remain 297 cities in the model and learning interactions happen within cities and not city types. I define the first ten city types as the ten largest French cities. To construct the remaining twenty types, I divide France into four geographic districts (North-East, North-West, South-East and South-West), and I subdivide each district into five clusters using a k-mean algorithm. I include in the clustering algorithm variables that proxy for the geographic primitives of the model: the coordinates of the city center, its average wage, and its population. [63] Figure F.1 plots the outcome of the clustering algorithm, and Table F.1 lists the largest city in each city type along with its average size and wage.

The model features two lifecycle periods covering workers' entire working career. [64] In the young phase, individuals work and learn while they only work in the old period. In the data, the wage-age profile plateaus around 40 year old. Accordingly, I set a lifecycle period to fifteen years and map the young period to ages 25 to 40 and the old period to ages between 40 and 55.

Finally, I externality calibrate the discount factor as the model does not provide an explicit moment to estimate it. I take a yearly discount factor of 0.975 and set $\beta = 0.975^{15}$.

4.2 Model Inversion

In the second step of the estimation, I use the structure of the model to derive estimating equations that identify half of the parameters without the need for simulation. This procedure allows me to

[63]While this classification is not without loss of generality, it is likely to have little impact on the counterfactuals since cities within each city-type are rather homogeneous. Figure F.2 plots the within-cluster to total sum of square residuals as a function of the number of clusters within each district. With five types, the clustering algorithm captures 83% of the total variation.

[64]Adding more lifecycle periods would make the model's learning technology more comparable to the estimates reported in Table 1. However, it would also increase its computation complexity without much effects on the counterfactuals of interest.

recover the parameters dictating the housing costs, the migration costs, and the learning technology.

Housing From the Cobb-Douglas preferences, α is the housing expenditure share. I therefore read off α from the French national accounts and set $\alpha = 0.2$ (INSEE, 2020). Regarding the housing supply parameters, the housing market clearing conditions require

$$\log p_\ell = \left(\frac{1}{1+\delta}\right) \log\left(\frac{1-\alpha}{\mathcal{H}}\right) + \left(\frac{1}{1+\delta}\right) \log\left(N_\ell \mathbb{E}[W_\ell]\right) \tag{20}$$

to hold, where p_ℓ, N_ℓ and $\mathbb{E}[W_\ell]$ are the housing rental price, the employment share, and the average wage in city ℓ. I obtain p_ℓ from the "Carte des Loyers" (Rental Map), which provides estimates for the average monthly rent per meter square at the municipal level.[65] I compute p_ℓ as the population weighted average of the municipal rents. Meanwhile, N_ℓ and $\mathbb{E}[W_\ell]$ are computed from the cross-sectional matched employer-employee. To match the frequency of the rent data, I set wages at the monthly frequency. Given (p_ℓ, N_ℓ, w_ℓ) and α, δ and $\mathcal{H}$ can be recovered from (20) by OLS. Figure F.5c compares the rent prices in the data with the predicted rent prices in the model. While simple, the model captures 71% of the between-city variation in rents.

Migration costs The migration costs can be recovered non-parametrically from workers' migration decisions. Using the spatial allocation of workers (equations (49) and (50) in Appendix D) together with the assumption of symmetric migration costs, the costs are recovered from

$$\mathrm{e}^{-2\vartheta\kappa^a_{l\ell}} = \int \left(\frac{n^a_{l\ell}(s)}{n^a_{ll}(s)} \Big/ \frac{n^a_{\ell\ell}(s)}{n^a_{\ell l}(s)}\right) n^a(s) ds, \tag{21}$$

where $n^a_{l\ell}(s)$ is the number of workers with age a and skill s who moved from city l to ℓ. Measuring the left-hand side requires to observe workers' skill and migration decisions. I proxy workers' skill with their wage quartile.[66] Consistently with the model, I measure migration decisions differently for young and old workers. In the panel version of the matched employer-employee, I observe workers' birthplace. Hence, for young workers, I compute $n^y_{l\ell}(s)$ as the number of workers under 40 born in city l and currently living in city ℓ with skill s. For old workers, I compute $n^o_{l\ell}(s)$ from the long panel as the number of workers with age above 40 that lived in city l 15 years ago and are currently living in city ℓ with skill s.[67]

Figure F.3 plots the estimated migration costs by city of origin against the distance between any two cities. Empirically, 45% (50% in the model) of young workers live in cities different than their birthplace, whereas 20% (16% in the model) of old workers change locations between their

[65] These rents are residualized and correspond to the rent paid for a typical 40m^2 apartment in France.

[66] Proxying workers' skill with their wage quartile may lead to biased migration costs as wages are the product of workers' skill and city TFP. The biases are likely to be small in practice for two reasons. First, the estimated variance of log TFP explains less than 2% of the variance of log wages. Second, (21) estimates migration costs off double difference in migration flows, and $n_{l\ell}(s)/n_{\ell\ell}(s)$ partially nets out the biases that emerge from using wages as a proxy for skills.

[67] Equation (21) can be used to the extent that $n^a_{\ell\ell}(s) \neq 0$ for all ℓ. To guarantee that this is the case for all cities in my sample, I directly compute the flows $n^a_{l\ell}(s)$ at the city cluster level.

youth and their old age. Through the lens of the model, this is rationalized with higher migration costs for old workers. The average migration cost for young workers is 1,255€, which corresponds to 43% of the average youth income. For old workers, the average migration costs is 1,594€, or 44% of their old wage.[68] For mean of comparison, Kennan and Walker (2011) estimates that the migration cost for young workers represent 49% of their youth income in the United States.

The model predicts substantial heterogeneity in migration probabilities across young workers. On average, a increase in young skills by one standard deviation is associated with a migration probability 1.1 percentage points larger. However, these elasticities differ across space. To see this, I estimate

$$\mathrm{P}_\ell(s) = \alpha_\ell + \beta_\ell \log s + u_\ell(s), \tag{22}$$

where $\mathrm{P}_\ell(s)$ is the probability that a worker born in ℓ with skill s moves out of their birthplace, α_ℓ is a city fixed effect, and β_ℓ is the city-specific migration elasticity. Figure F.6a and Figure F.6b displays the migration elasticities for young and old workers. I find that young skilled workers born in productive cities are relatively more likely to stay in their birthplaces, whereas young skilled workers born in low-TFP locations are more likely to move out. The same pattern holds for old workers, but the migration elasticities are one order of magnitude smaller, suggesting that most of the observed sorting patterns are determined at the onset of workers' career.

Learning technology The learning technology parameters g_1, g_2 and g_{12} are estimated from the between-city specification (18). In the model, the young period lasts fifteen years. In the data, I use wage growth at the five year horizon to have enough statistical power. To get a learning technology with the appropriate time horizon, I therefore estimate (18) for three age bins: 25 to 30, 30 to 35 and 35 to 40 (see Figure E.5). I then set g_1, g_2 and g_{12} as the sum of their respective age-specific estimates.

4.3 Indirect Inference

The remaining seven groups of parameters are jointly estimated by minimizing the distance between empirical moments and their model counterparts. I provide first heuristic arguments that justify the identification strategy. I then show numerically how the chosen moments affect the calibration of the model.

City characteristics In addition to migration costs, cities differ in their productivity $\boldsymbol{T}$ and their age-specific amenities $(\boldsymbol{B}^y, \boldsymbol{B}^o)$. I set the productivity of each city to mach its average wage in the

[68] The estimated migration costs are in utils. The reported estimates are converted to monetary unit through $\kappa^a_{\ell\ell} P_\ell$. The reported average cost are computed as unweighted average, excluding $\kappa^a_{\ell\ell}$. As such, they do not represent the migration costs *actually* paid by workers.

data,

$$T_\ell = \frac{\mathbb{E}[W_\ell]}{\int s\pi_\ell(s)ds}, \tag{23}$$

where $\mathbb{E}[W_\ell]$ is the average wage of city ℓ computed from the matched employer-employee data, and $\int s\pi_\ell(s)ds$ is the average skill of city ℓ that I compute within the model.[69] Meanwhile, the age-specific city amenities are calibrated to match the age-specific city employment share,

$$\underbrace{N^a_\ell}_{\text{Emp. share}} = \underbrace{\mathrm{e}^{\vartheta\left(B^a_\ell - B^a_{\bar\ell}\right)}}_{\text{Relative amenity}} \underbrace{\int \mathrm{e}^{\vartheta\left(\frac{w_\ell(s)}{P_\ell} - \frac{w_{\bar\ell}(s)}{P_{\bar\ell}} + \beta\left[O^a_\ell(s) - O^a_{\bar\ell}(s)\right]\right)} \sum_l \mathrm{e}^{-\vartheta(\kappa^a_{l\ell} - \kappa^a_{l\bar\ell})} n^a_{l\bar\ell}(s)ds}_{\text{Relative attractiveness of } \ell \text{ net of amenity differentials}}, \tag{24}$$

for $\bar\ell$ some reference city. In (24), the left-hand side is the employment share of city ℓ, computed from the matched employer-employee data, while the integral on the right-hand side is computed within the model. Expression (24) thus identifies city amenities up to a constant, and I normalize $B^y_\ell = B^o_\ell = 0$ for Paris.

Figure F.4 displays the estimated city TFP and amenities. I estimate the TFP of the city in the top 10% of the unweighted productivity distribution (Paris) to be 21% larger than the TFP of the city in the bottom 10% of the TFP distribution (Blois). Overall, the variance of log TFP explains 1.7% of the variance in log wages, and 16% of the between-city component. Regarding amenities, I estimate that most cities (79% and 90% of cities for young and old workers respectively) offer worst amenities than Paris.

Skill distributions The young skill distribution is governed by its mean, μ_y, and its standard deviation, σ_y. Workers' average skill and aggregate TFP are not separately identified. I thus normalize the average log skill to zero and set μ_y to ensure that the normalization holds.[70] I then calibrate σ_y to match the aggregate variance of log wages for young workers.

The old skill distribution is an equilibrium object dictated by the young skill distribution and the learning interactions. I therefore set the remaining learning parameters to match the two first moments of the old skill distribution. First, g_0 dictates the overall level of learning across the lifecycle. I thus set g_0 to match the old to young wage ratio. Second, the variance in the learning shocks, σ_ν, shapes the dispersion in old skill. I therefore calibrate σ_ν to match the variance of log wages for old workers. I find that workers' initial skill and interactions together explain 30% of the total variance in skill growth, whereas the remaining 70% are due to the idiosyncratic learning shocks.[71]

[69] In equilibrium, the average skill depends on $\boldsymbol{T}$. Hence, (23) constitutes a fixed point over $\boldsymbol{T}$. Similarly, (24) constitutes a fixed point over $\boldsymbol{B}^y$ and $\boldsymbol{B}^o$. While I cannot prove that these fixed points have a unique solution, I numerically converge to the same vector of productivity and amenities when starting from different initial conditions.

[70] When estimating the learning technology in Section 3.5, I also ensure that this normalization holds by setting the average of the log wage distribution to zero.

[71] While idiosyncratic learning shocks explain the majority of the wage growth variance, workers' initial skill explain most of the variance of lifetime income in accordance with the findings of Huggett et al. (2011).

The joint calibration of σ_y, σ_ν and $\boldsymbol{T}$ implies that the model matches the aggregate, between-city and within-city variance of log wages.

Location preferences The last parameter to be estimated is the dispersion in idiosyncratic location preferences, ϑ. This parameter governs the amount of spatial sorting that prevails in equilibrium. When $\vartheta \approx 0$, the idiosyncratic location preferences are very dispersed, and the skill distributions are homogeneous across cities. As a consequence, there is no spatial dispersion in wage growth. On the contrary, as ϑ increases, skilled workers become concentrated in productive cities. Given the estimated the learning technology, the concentration of skilled workers in a few cities increases the average wage growth of productive places. I therefore set ϑ to match the between-city dispersion in average wage growth.[72]

Identification How tight is the calibration of the model? Figure F.7 reports how the predicted moments vary as a function of the parameters. Across the board, the predicted moments are sensitive to the parameters they are supposed to identify, with a 1% change in the parameter associated with a change in the targeted moment between 0.5% and 2%. Furthermore, most parameters affect only the moments they are supposed to match. Two notable exceptions are the variance of the young skill distribution and that of the learning shocks who also matter for the between-city variance of wage growth.[73] I provide two over-identification exercises to further validate the model's calibration.

4.4 Over-identification exercises: sorting in the data and in the model

Two key sets of parameters govern the impact of local interactions on human capital accumulation and agglomeration. The first set of parameters relate to the learning technology. Section 3.5 provided several robustness exercises on the estimation of g_1, g_2 and g_{12}. The second key parameter is the dispersion in the idiosyncratic preference shocks. In addition to governing the extent to which interactions are spatially differentiated, this parameter is crucial in determining how workers respond to spatial policies. I now provide two over identification exercises that validate the calibration of ϑ.

Between-city wage gaps In the theory, the average wage of city ℓ may be high because the city is relatively productive, or because the workers living there are particularly skilled. Specifically, the between-city wage gaps can be decomposed into a TFP gap and a skill gap,

$$\underbrace{\mathbb{E}[\log W_\ell] - \mathbb{E}[\log W]}_{\text{Between-city wage gap}} = \underbrace{\log T_\ell - \mathbb{E}[\log T]}_{\text{TFP gap}} + \underbrace{\mathbb{E}[\log S_\ell]}_{\text{Skill gap}}, \tag{25}$$

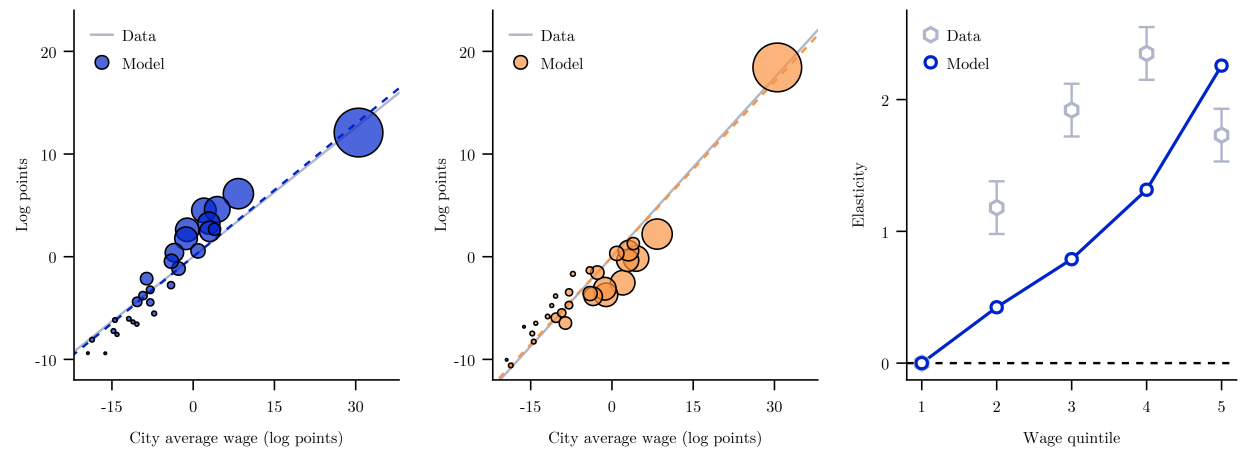
where $\mathbb{E}[\log S] = 0$ by normalization. Local TFPs are calibrated to match the local average wages. However, neither the TFP gaps nor the skill gaps are targeted in the estimation. In particular, for a

[72] The level of ϑ depends on the monthly wage frequency adopted, making it hard to compare its value with previous estimates from the literature.

[73] In addition, Figure F.8 reports the sensitivity measure of Andrews et al. (2017). In accordance with Figure F.7, the variance of the young skill distribution and that of the learning shocks are the most sensitive parameters to the targeted moments.

Figure 3: Worker sorting in the model and in the data.

(a) Spatial TFP gaps (b) Spatial skill gaps (c) Wage elasticity of city size



Note: panel (a) displays the estimated city TFP against the city average wage. The blue dashed line is computed as $\log T_\ell = \alpha + \beta \log w_\ell$. The grey solid line is computed as $\log \hat{T}_\ell = \alpha + \beta \log w_\ell$, where $\log \hat{T}_\ell$ are the location fixed effects estimated from a two-way fixed effect regression of log wages on a city and a skill fixed effect. Panel (b) displays the average skill in city ℓ, $\mathbb{E}[\log S_\ell]$, against the city average wage. The orange dashed line is computed as $\mathbb{E}[\log S_\ell] = \alpha + \beta \log w_\ell$. The grey solid line is computed as $\mathbb{E}[\hat{\beta}_s \mid \ell] = \alpha + \beta \log w_\ell$, where $\hat{\beta}_s$ are the skill fixed effects from the TWFE regression. The size of the markers in panels (a) and (b) is proportional to city size. Panel (c) plots the point estimates of $\{\gamma_q\}$ in (27) obtained from the model (blue circles) and from the data (grey hexagons).

given spatial distribution of average wages, greater sorting implies smaller TFP gaps. Comparing the spatial TFP and skill gaps in the data and in the model thus allows to quantify the extent to which the model captures sorting correctly, and with it, whether ϑ is correctly calibrated.

Computing the wage decomposition (25) empirically requires to directly estimate local TFPs. Through the lens of the model, the wage of worker i when employed in city ℓ_{it} is $\log w_{it} = \log T_{\ell_{it}} + \log s_{it} + u_{it}$, for u_{it} some measurement error. Hence, city TFP can be estimated as a city fixed effect upon observing workers' skill s_{it}.

I use workers' occupation at the four-digit level, interacted with their tenure at the occupation, to proxy for their skill. Proxying skills with occupations has the advantage that skills can evolve over time, much like in my model. In particular, worker fixed effects, as in Card et al. (2023), cannot be used to estimate the productivity of a location in the presence of spatial heterogeneity in learning. Measuring skill through worker fixed effect would indeed confound the TFP estimates with the local learning opportunities since those fixed effects are constant over time (see Appendix F.4).[74] The occupation-based skill definition is only used in this over-identification exercise.

Figure 3a and 3b plots the TFP and skill gaps in the model against the between-city wage gaps. Cities with a relatively high wage are more productive and attract a higher density of skilled workers. Combined, these two panels show that skilled workers locate disproportionately in productive cities. Quantitatively, the spatial TFP differentials explain 43% of the spatial wage gaps in the model (blue dotted line in Figure 3a), while the between-city skill gaps explain the remaining 57% of the

[74] Appendix F.4 provides two robustness exercises. First, I estimate the two-way fixed effect model on young workers only to minimize the amount of learning that takes place in cities. Second, I define a skill as a four-digit occupation interacted with the wage quintile in that occupation.

variation (orange dotted line in Figure 3b).[75] In the data, I find that the TFP and skill gaps explain respectively 42% and 58% of the spatial wage variation (grey lines in both panels). In addition to spatial differences in average wage, worker sorting also generates greater within-city wage inequality in larger, more productive places (Figure F.9), aligned with the findings of literature.[76]

Sorting by age In the model, workers' willingness to sort varies by skill and by age: young workers consider both the learning potential of cities and the local wage, whereas old workers only sort according to the earnings differential. When migration costs are small, the model predicts that the between-age difference in the number of workers in city ℓ with skill s is

$$\log\left(\frac{n_\ell^y(s)}{n_\ell^o(s)}\right) = \vartheta\left(B_\ell^y - B_\ell^o\right) - \log\left(\frac{\mathcal{N}^y(s)}{\mathcal{N}^o(s)}\right) + \vartheta\beta O_\ell(s), \tag{26}$$

where $\mathcal{N}^a(s) \equiv \sum_\ell \mathrm{e}^{\vartheta[w_\ell(s)+B_\ell^a+\beta O_\ell^a(s)]}$ is independent from ℓ. The last term captures the learning motive that is present only for young workers. The shape of the learning technology affects whether young skilled workers place a relatively higher learning value on productive cities than low skill workers. The dispersion in idiosyncratic preferences modulates the between-age difference in willingness to sort. The spatial variations in young to old ratio can therefore be used to validate jointly the estimation of the learning technology and the dispersion in idiosyncratic preferences.

Motivated by (26), I measure in the data the between-age difference in willingness to sort through the reduced-form specification

$$\log\left(\frac{N_{\ell q}^y}{N_{\ell q}^o}\right) = \alpha_\ell + \beta_q + \gamma_q \mathbb{E}[\log W_\ell] + e_{\ell q}, \tag{27}$$

where $N_{\ell q}^a$ is the number of workers living in city ℓ with age a in wage quintile q, and $e_{\ell q}$ is some residual. The parameters α_ℓ and β_q are a city and wage quintile fixed effects respectively. I proxy the quality of interactions in city ℓ by the city log wage, and let workers in different wage quintile value those interactions differently, as captured by γ_q. If $\gamma_q > 0$, young workers in quintile q are more sensitive to spatial variations in wages than old workers in the same quintile. If $\gamma_{q'} > \gamma_q$, this between-age difference in willingness to sort is greater for workers in quintile q' than q. The parameters $\{\gamma_q\}$ are identified up to a constant, and I normalize $\gamma_1 = 0$ to focus on the difference in between-age willingness to sort across skills. I estimate (27) in the data and in the model by OLS.

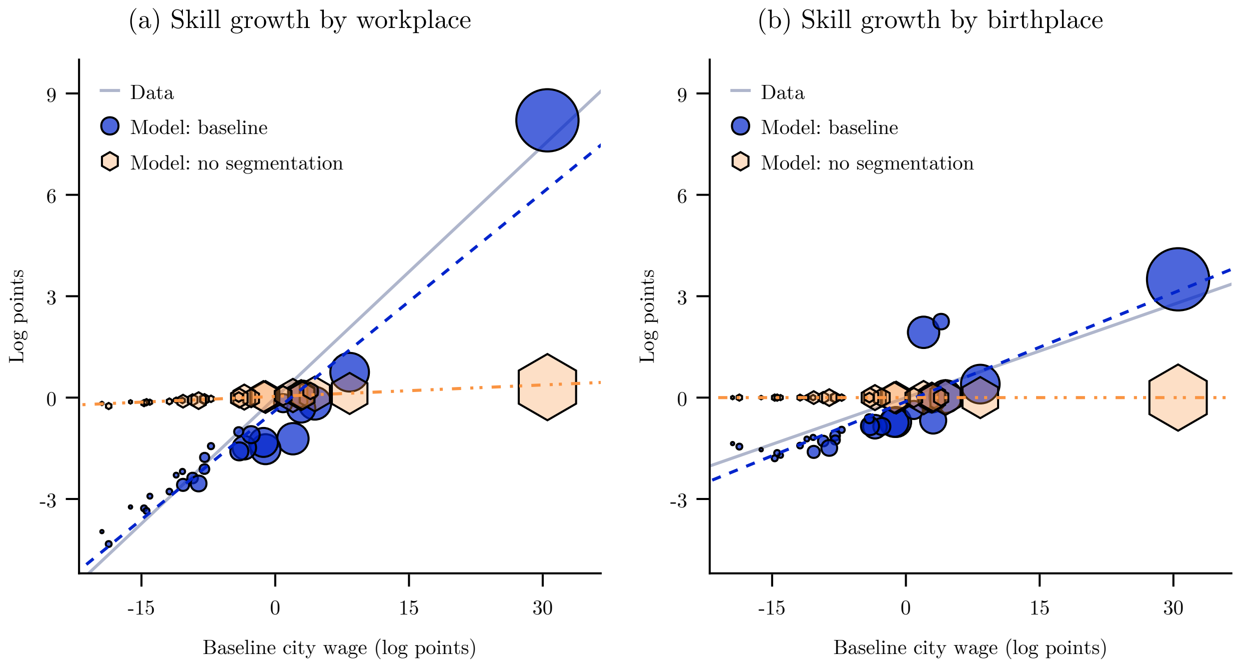
Figure 3c presents the point estimates. In the data, the extra incentives of young workers to sort to high-wage places is greater for skilled workers. The same pattern old in the model. Figure F.10 plots the model-implied spatial differences in option values for low, middle and high-skill. The future opportunities offered by productive cities are more attractive to skilled workers, which generates further sorting from them.[77]

[75] Table F.3 in Appendix F.4 provides the variance decomposition of between-city wage inequality.

[76] See for instance Baum-Snow and Pavan (2013), Eeckhout et al. (2014), Papageorgiou (2022) and Lhuillier (2022).

[77] When migration costs are positive, the option value of productive cities reflect both the learning opportunities and the future migration costs. Figure F.10 plots the spatial differences in option values when interactions are not

Figure 4: Between-city differences in skill growth in the model and in the data



Note: panel (a) displays the average skill growth by workplace in the equilibrium with segmented interactions (blue circles) and in the equilibrium without segmented interactions (orange circles) against the baseline city average wage. The blue and orange dashed lines are the (unweighted) fitted lines. The grey line is the empirical (unweighted) fitted line. Panel (b) displays the average skill growth by birthplace. The same nomenclature as in panel (a) applies. The size of the markers is proportional to city size.

The model therefore replicates the sorting patterns observed in the data. I now use the estimated model to quantify the impact of local interactions on agglomeration and human capital accumulation.

5 The Tradeoff Between Human Capital and Inequality

To what extent do local interactions shape learning and the spatial distribution of economic activity? And what is the implied tradeoff between human capital accumulation and spatial inequality? To answer these questions, I use the estimated model and compare the baseline equilibrium with a counterfactual economy in which interactions are not segmented by cities. In the latter case, interactions continue to be random, but the likelihood of interactions is given by the aggregate skill density. The aggregate skill density in the counterfactual economy may in turn differ from the baseline skill distribution to the extent that cities shape human capital accumulation. I compare the two economies in steady state.

Section 5.1 presents the impact of local interactions on human capital accumulation, and Section 5.2 studies how local interactions shape spatial agglomeration. Together, the two sections help understand the tradeoff between spatial inequality and aggregate productivity generated by local interactions.

segmented by cities (Section 5). In this case, skilled workers continue to place a higher option value on productive cities to front-load the future migration costs, but the between-city differences are four times smaller.

5.1 Local Interactions and Human Capital Accumulation

Figure 4 plots the spatial differences in skill growth that arise in the model. As in the data, individuals who work in high-wage cities experience faster human capital accumulation (Figure 4a). For instance, workers in Paris experience a skill growth 9 log points higher than the average French worker. Furthermore, fast-growth is very spatially concentrated, and Paris stands as an outlier in the city-growth distribution. The spatial differences in learning opportunities, rather than the sorting of fast learners to productive cities, generates these between-city differences in learning. When interactions are not segmented by cities, the growth differences are minimal.[78]

The spatial segmentation of learning interactions amplify learning inequality, both across birthplaces and initial skills. First, workers born in skill-dense cities have an easier access to productive interactions in the presence of migration costs (Figure 4b). For instance, workers born in Paris have an expected skill growth 3 log points higher than the average French worker. In the cross-section, an increase in the birthplace average wage by one standard deviation is associated to a lifetime skill growth 1.7 log points higher.[79] Although not directly targeted in the estimation, the same pattern holds in the data. This birthplace learning premium automatically disappears when interactions are not segmented by cities.

Second, local interactions also increase learning inequality across skills. Figure 5a displays the average skill growth experienced by workers as a function of their initial skill. When interactions are spatially segmented, there are substantive differences in learning across workers: workers in the top 1% of the young skill distribution experience a skill growth on average 17 log points higher than workers in the bottom 10% of the skill distribution. Overall, an increase in initial skill by one standard deviation is associated with a lifetime skill growth 1.6 log points larger.[80]

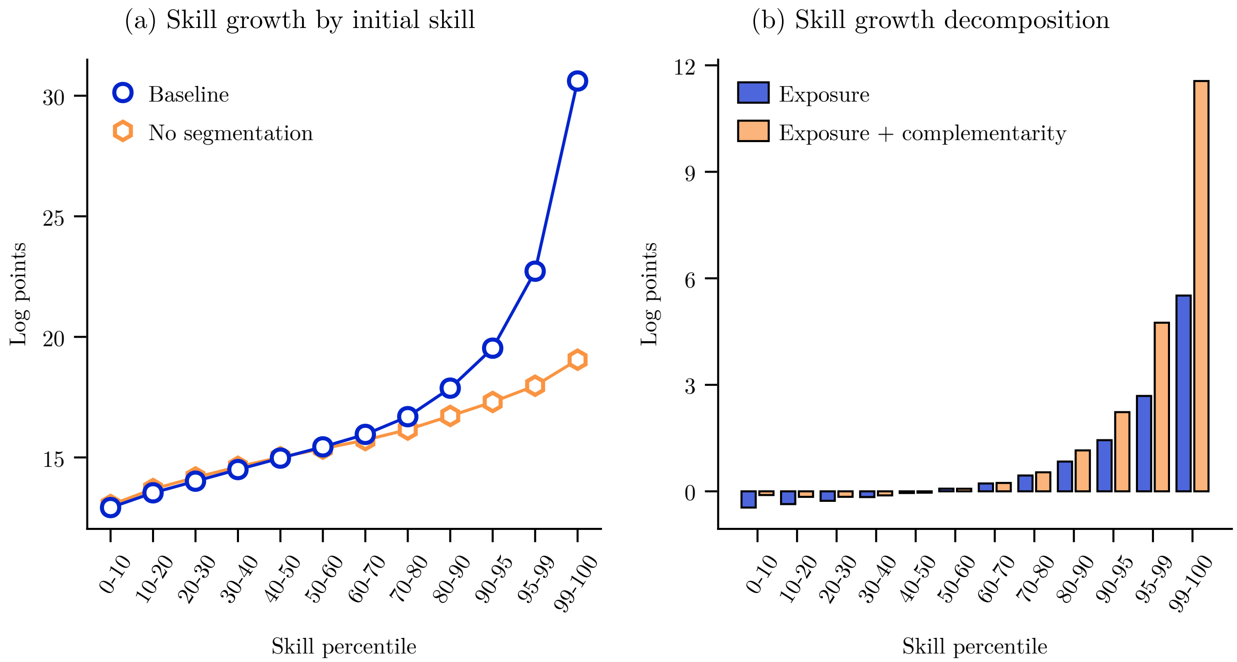
The faster accumulation of human capital for skilled workers is a robust pattern documented in the human capital literature (e.g. Lagakos et al., 2018, amongst many others). How much of this skill learning premium is due to local interactions? According to Proposition A.3, local interactions amplify the learning of skilled workers while dampening that of low-skill. Quantitatively, I find that local interactions are indeed crucial for skilled workers. When interactions are not spatially segmented, workers in the top 1% of the skill distribution experience a skill growth 12 log points (38%) lower than in the baseline equilibrium. On the contrary, low-skill workers experience very modest gains (0.1 log points) in the counterfactual equilibrium. An increase in initial skill by one standard deviation is associated with a lifetime skill growth 0.6 log points higher in the counterfactual economy, so that 63% of the skill learning premium is explained by the spatial segmentation of local interactions.

[78] There remain small spatial gaps in learning as productive cities continue to be skill-dense and skilled workers have a more productive learning technology ($g_1 > 1$).

[79] Figure 4b hides significant heterogeneity across the skill distribution. Figure F.11 shows that the learning benefits of being born in a fast-growth location are stronger for middle- and high-skill workers. For instance, workers born in Paris in the top 10% of the skill distribution enjoy a skill growth 7 log points higher than their peers. These larger differences for middle and high skill arise because their learning technology allow them to better grasp the returns from high skill interactions.

[80] A similar skill learning premium holds in the French data (see Figure F.12).

Figure 5: The aggregate consequences of local interactions on human capital accumulation



Note: panel (a) displays the average skill growth by young skill in the equilibrium with segmented interactions (blue circles) and in the equilibrium without segmented interactions (orange circles). Panel (b) displays the change in skill growth between the baseline economy and the counterfactual economy across the young skill distribution. The total change is decomposed according to (28). The blue bars plot the "average exposure" effect, and the orange bars plot the total change.

The spatial segmentation of interactions thus increases learning inequality across workers, and the variance of skill amongst old workers is 4% larger in the baseline economy. At the same time, local interactions amplify the learning of skilled workers without affecting that of low-skill. As a consequence, the aggregate stock of human capital is 1.2% higher when interactions are segmented by cities.

Why do high skill workers gain so much from local interactions? The impact of local interactions on skill growth can be decomposed into two channels. First, high skill workers may be agglomerating relatively more to skill-dense cities, and as result, may be more exposed to productive interactions when those are spatially segmented. Second, high skill workers may be better able to capitalize on skilled interactions due to the presence of within-skill learning complementarities. Formally, the change in skill growth between the baseline and the counterfactual economy can be written

$$\mathbb{E}\left[\log\left(\frac{S^o}{s}\right) \mid s\right] - \mathbb{E}\left[\log\left(\frac{\bar{S}^o}{s}\right) \mid s\right] = \underbrace{g_2 \sum_{\ell} \left(\frac{n_\ell^y(s)}{n^y(s)}\right) \Delta_\ell}_{\text{Average exposure}} + \underbrace{g_{12} \log s \sum_{\ell} \left(\frac{n_\ell^y(s)}{n^y(s)}\right) \Delta_\ell}_{\text{Learning complementarity}}, \tag{28}$$

where $\Delta_\ell \equiv \mathbb{E}[\log S_\ell] - \mathbb{E}[\log \bar{S}]$ is the difference between the average skill in city ℓ in the baseline equilibrium and the average aggregate skill in the counterfactual equilibrium.[81]

[81] The change in skill growth can further be decomposed into a partial equilibrium effect that holds the average old skill constant, $\mathbb{E}[\log S] - \sum_\ell \left(\frac{n_\ell^y(s)}{n^y(s)}\right) \mathbb{E}[\log S_\ell]$, and a general equilibrium effect, $\mathbb{E}[\log \bar{S}] - \mathbb{E}[\log S]$. The general

Figure 5b plots the skill growth decomposition. In the baseline equilibrium, high skill workers agglomerate to the most skill-dense location, with 40% of the workers in the top 10% of the skill distribution living in Paris. This greater exposure to skilled interaction explains 53% of their faster learning when interactions are segmented by cities. Said differently, half of the impact of cities on human capital accumulation for skilled workers is due to the presence of within-skill learning complementarities.

Meanwhile, the learning complementarities also minimize the learning losses experienced by low-skill workers. In the baseline equilibrium, low skill workers do not sort much across space, neither to low nor to high TFP cities. As a result, the learning opportunities they experience are already quite similar to the nationwide opportunities. Absent learning complementarities, the spatial segmentation of interactions reduces the skill growth of workers in the bottom 10% of the skill distribution by 0.5 log points. The lower returns to skilled interactions for low skill workers brings these losses further down to 0.1 log points.

By boosting the learning gains of high-skill and reducing the losses of low-skill, learning complementarities are therefore crucial to understand the impact of local interactions on aggregate human capital accumulation. Without learning complementarities, I find that cities would only augment the aggregate stock of human capital by 0.5%.

5.2 Local Interactions as a Source of Agglomeration

To what extent do local interactions shape the spatial distribution of economic activity in addition to human capital accumulation? When interactions happen between individuals who work in the same location, workers agglomerate in productive cities that feature a relatively high-share of skilled partners (Proposition 2.1). Figure 6a shows that this agglomeration force is large. When interactions are no longer spatially segmented, the population of Paris reduces from 14% to 7% of the French workforce, while the number of workers living in cities in the bottom 25% of the TFP distribution increases from 10% to 12%.[82]

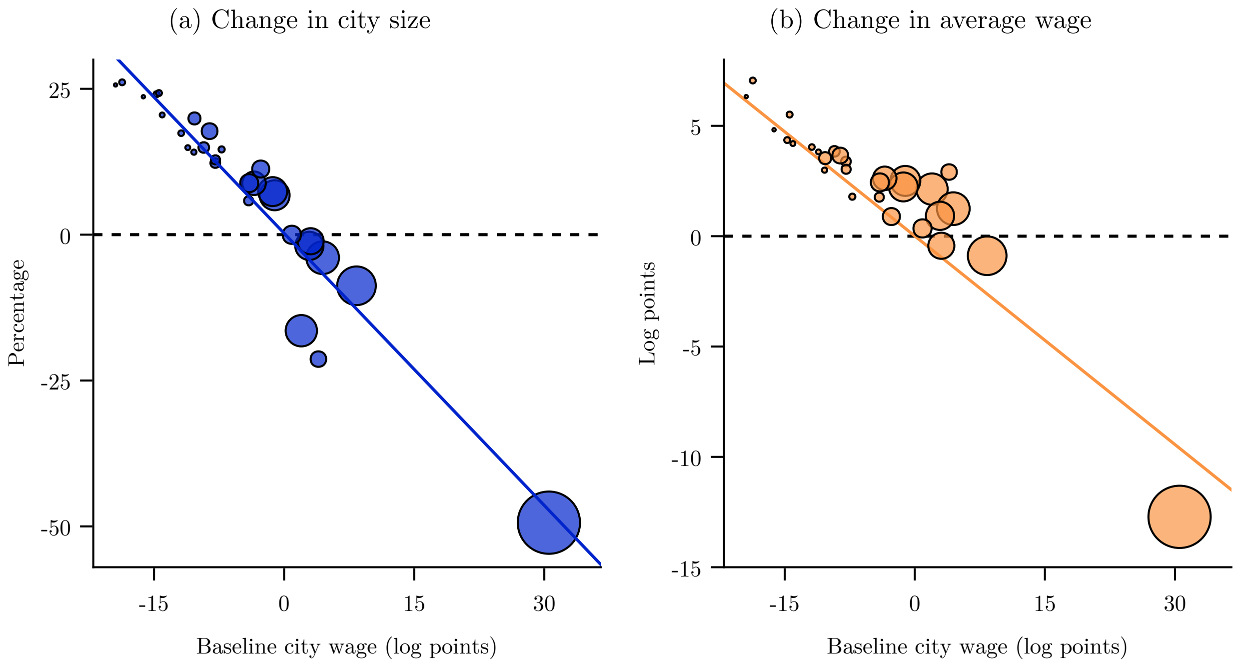
The strong within-skill learning complementarities furthermore imply that the learning opportunities of productive cities are more tailored to skilled workers. As a result, local interactions make productive cities more skill intensive than they would be in a counterfactual world without spatially segmented interactions (Proposition 2.2). In this counterfactual economy, the average skill in Paris drops by 15 log points, whereas small, low productivity cities see their local stock of human capital increase (Figure 6b).

Overall, I find that local interactions act as a strong source of agglomeration which ramps up wage inequality. Across space, the between-city variance of log wages would be 72% lower were interactions not segmented by cities. In the aggregate, the variance of log wages would be 7% lower, 36% of which is due to the smaller learning disparities across workers (Figure 5), while the remaining

equilibrium effects alone are however small (see Figure F.13).

[82] In the presence of costly migration, local interactions agglomerate both young and old workers in productive cities, as the majority of old workers stay in the locations where they lived when young (Figure F.14a).

Figure 6: Local interactions as a source of agglomeration



Note: panel (a) displays the percentage change in city size between the baseline equilibrium and the equilibrium without segmented interaction against the baseline city average wage. The blue line is the fitted line. Panel (b) plots the change in average log wage between the baseline equilibrium and the equilibrium without segmented interaction against the baseline city average wage. The orange line is the fitted line. The size of the markers is proportional to the baseline city size in both panels.

is explained by the weaker spatial agglomeration (Figure 6).[83] At the same time, by agglomerating workers in productive cities and fostering human capital accumulation, local interactions increase aggregate productivity. The average wage is 3% higher when interactions are spatially segmented; 16% of this increase is explained by the larger stock of human capital, whereas the remaining 84% are coming from the greater agglomeration of workers in productive cities.[84]

Local interactions therefore create an equity-efficiency tradeoff. On the one hand, they boost human capital accumulation and aggregate wages. On the other hand, they amplify wage and learning inequality, across space and workers. I conclude the paper by studying the implication of this equity-efficiency tradeoff for spatial policies.

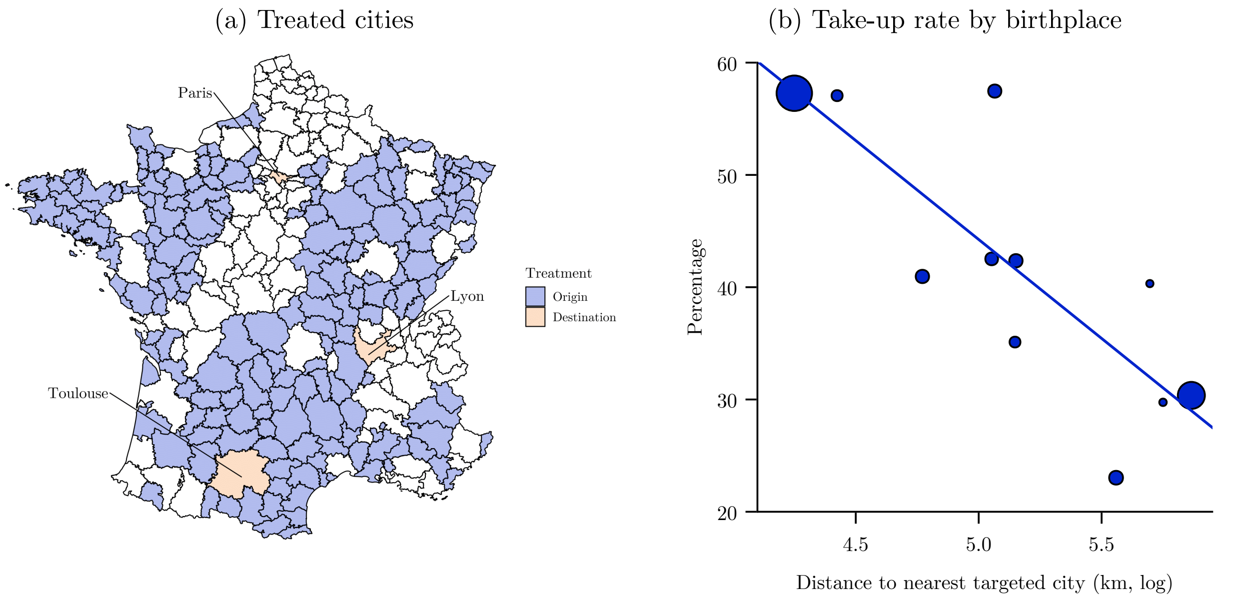
6 The Consequences of Spatial Policies

To what extent spatial policies can mitigate spatial learning inequality? And what are their consequences on aggregate human capital accumulation? To answer these questions, I study the

[83]The aggregate change in inequality can be decomposed into a TFP effect, a skill effect, and a third term that captures the sorting of skilled workers to productive locations: $\Delta\mathbb{V}\text{ar}[\log W] = \Delta\mathbb{V}\text{ar}[\log T_\ell] + \Delta\mathbb{V}\text{ar}[\log S] + 2\Delta\mathbb{C}\text{ov}[\log T_\ell, \log S]$. The first term captures the lower agglomeration to productive cities (8%), the second term reflects the learning disparities that emerge across workers as a consequence of local interactions (36%), and the last term measures the spatial sorting (56%).

[84]The change in average log wage between the baseline and the no-segmentation economies can be decomposed into a TFP tern and a human capital term, $\mathbb{E}[\log W] - \mathbb{E}[\log \bar{W}] = \sum_\ell \log T_\ell \left(N_\ell - \bar{N}_\ell\right) + \mathbb{E}[\log S] - \mathbb{E}[\log \bar{S}]$.

Figure 7: The moving voucher policy



Note: left-panel displays the cities that are offered the moving voucher (light blue) along with the three cities that are targeted by the vouchers (light orange). Right-panel shows the average take-up rate of the voucher by birthplace when the policy reaches 1.5% of GDP. The size of the circles is proportional to city size in the baseline equilibrium.

consequences of a moving voucher policy aimed at offering equal learning opportunities to workers born in remote locations (Chetty et al., 2016).[85] Specifically, the policy consists of a subsidy granted to young workers born in the bottom 25% of the city-growth distribution conditional on moving to the three largest French cities (Paris, Toulouse and Lyon). Figure 7a displays the 185 cities that are treated by the policy. The policy is self-financed within skill and age group to focus on spatial redistribution.[86]

Section 6.1 describes the consequences of the policy on spatial inequality. Section 6.2 follows by presenting its impact on human capital, and Section 6.3 wraps-up with a welfare analysis of the policy.

6.1 Spatial Inequality

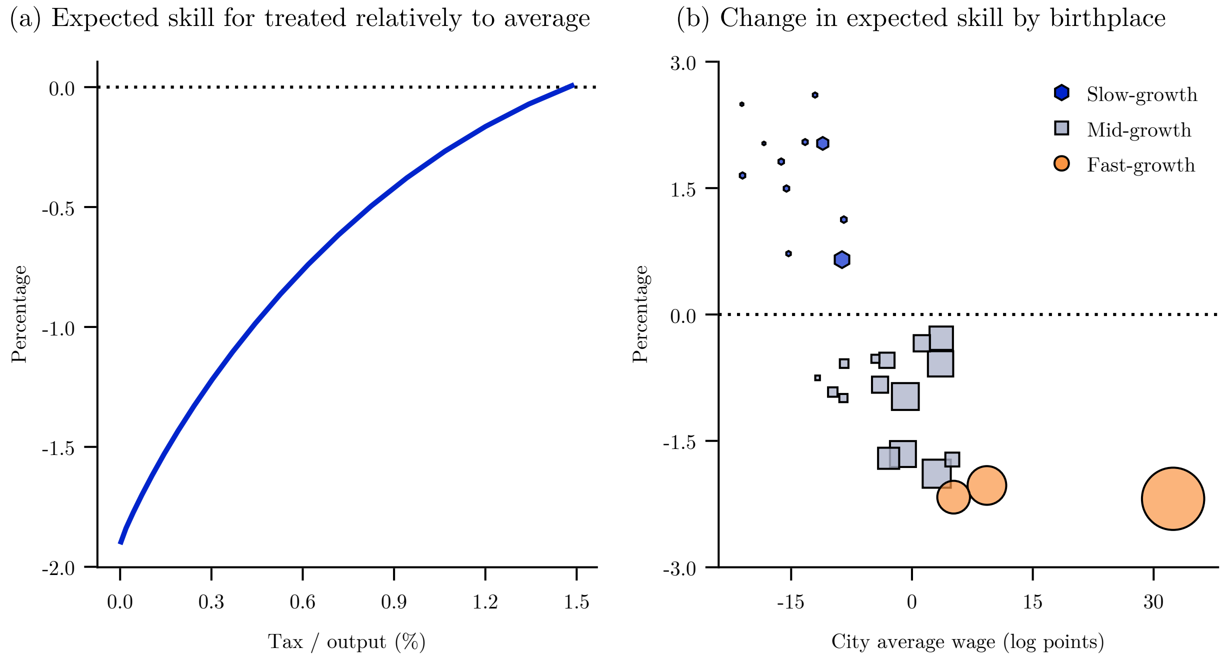
In the baseline equilibrium, workers born in slow-growth locations have lower access to productive cities and interactions, which dampens their human capital accumulation and lifetime income. Quantitatively, the expected old skill and lifetime income of workers born in slow-growth cities are respectively 2% and 4% lower than those of the average French worker.

Figure 8 presents the impact of the voucher policy on the between-city learning gap. As the moving voucher policy expands, it covers a greater share of the migration cost to productive cites.

[85] Chetty et al. (2016) evaluates the consequence of the Moving to Opportunity program in the United States, which subsidized children and teenagers born in distressed neighborhoods to move to large, productive places. In my setting, the vouchers are granted to young professionals at the beginning of their career. See Fogli et al. (2023) for a paper that evaluates the general equilibrium consequences of the MTO program.

[86] The financing matters only for the welfare calculation. As long as the financing is through a lump-sum tax that is not place-based, it has no implication on the spatial distribution of economic activity, and therefore on human capital.

Figure 8: The consequences of spatial policies on spatial learning inequality



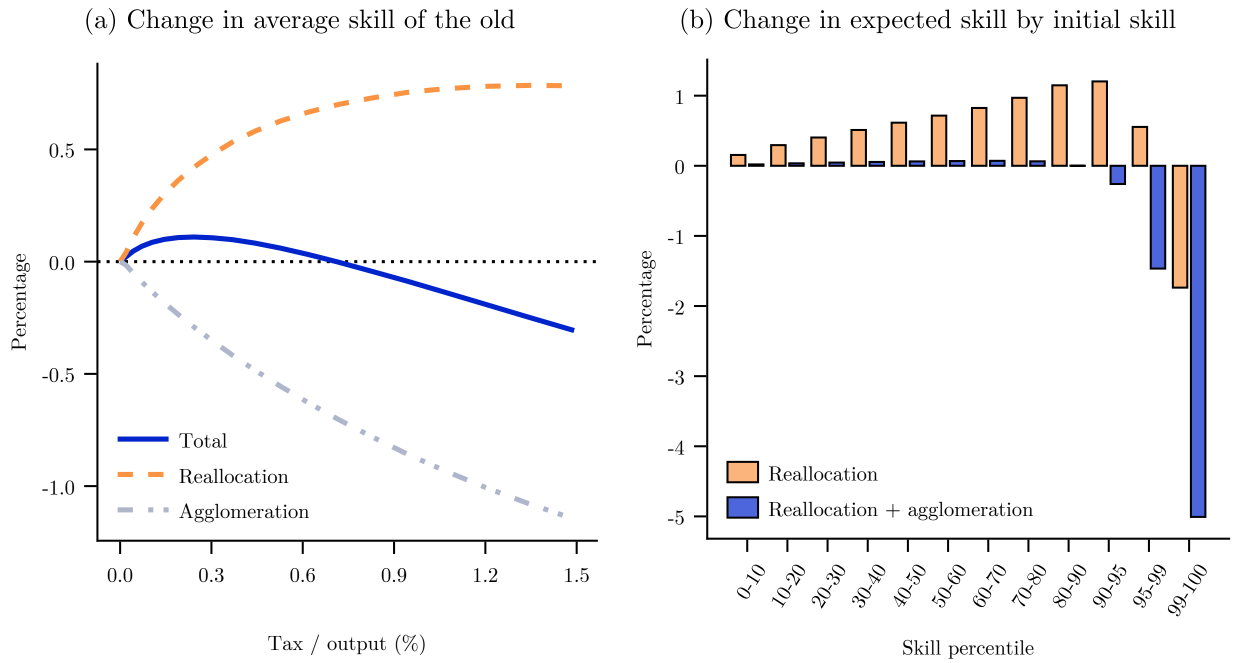
Note: panel (a) displays the expected old skill of workers born in cities in the bottom 25% of the city growth distribution relative to the nationwide expected old skill against the size of the voucher policy. The size of the policy is measured as the aggregate tax used to finance the policy as a fraction of total GDP. Panel (b) plots the change in expected old skill by birthplace between the baseline equilibrium and the policy equilibrium when the policy reaches 1.5% of GDP. The blue hexagons are the cities in the bottom 25% of the city growth distribution. The orange circles are the three largest French cities (Paris, Lyon and Toulouse). The grey rectangles are the remaining cities. The size of the markers is proportional to the baseline city size.

Workers born in remote locations have an easier access to productive learning opportunities, which reduces the spatial gap in expected skill (Figure 8a). When the voucher reaches $1,030$€, which covers 80% of the average migration cost and represents 1.5% of GDP, workers born in remote locations have the same human capital accumulation as the average French worker. Combined with the greater access to high-paying jobs, the moving voucher increases the lifetime wages of workers born in poor places by 12% on average.

The policy has heterogeneous treatment effects across workers and cities. Figure 8b plots the change in expected old skill by birthplace between the baseline equilibrium and the equilibrium when the policy reaches 1.5% of GDP. Workers born in treated cities (blue hexagons) all experience an increase in their human capital accumulation, but the gains are relatively greater for workers born near Paris, Lyon and Toulouse for whom the take-up rate of the policy is larger (Figure 7b). Across skill, treated workers in the top 10% of the skill distribution enjoy an increase of their expected old skill by 3.2%, whereas treated workers in the bottom 50% of the distribution experience a gain of 0.5% (Figure G.15a).

By granting an easier access to productive cities for workers born in poor places, moving vouchers reduce spatial wage inequality. In the policy equilibrium, the between-city variance of log wages is 26% lower than in the baseline equilibrium, and the aggregate variance of log income is 1.5% lower. The decrease in between-city wage inequality comes primarily from the reallocation of less skilled workers to productive cities (Figure G.16a).

Figure 9: The consequences of local policies on human capital accumulation



Note: panel (a) displays the average old skill against the size of the policy. The blue solid line is the change in average old skill under the voucher policy. The orange dashed line is the change in average old skill under the quasi-optimal policy. The size of the policy is measured as the aggregate tax used to finance the policy as a fraction of total GDP. Panel (b) and (c) plot the change in expected old skill across the skill distribution between the baseline equilibrium and the policy equilibrium when the policy reaches 1.5% of GDP under the moving voucher and the quasi-optimal policy respectively. The changes are decomposed according to (29). The blue bars represent the partial equilibrium effect. The orange bars display the total change.

However, the change in the skill composition of productive cities also shape the learning that takes place there. By lowering the average skill of workers in Paris, Toulouse and Lyon, the policy decreases the local quality of interactions (Figure G.16b). As a result, non-treated workers have access to relatively worse learning opportunities, and slower human capital accumulation follows (grey rectangles and orange circles in Figure 8b). These learning losses are greater for workers born in Paris, Toulouse and Lyon, who initially had the greatest exposure to skilled interactions, as well as for high-skill workers who benefit the most from the concentration of skilled individuals in Paris (Figure G.15a).[87]

The moving voucher policy therefore redistributes learning opportunities away from workers born in skilled places to those born in remote locations. In doing so, it is effective at reducing spatial wage and learning inequality. I now turn to studying the aggregate impact of the policy on human capital accumulation.

6.2 Human Capital Accumulation

What are the aggregate consequences of the moving voucher policy on human capital accumulation? Figure 9a displays the change in the average skill of old workers as a function of the size of the policy. I find that the moving voucher policy triggers only a very mild decrease in aggregate human capital. When the policy reaches 1.5% of GDP, the average old skill is 0.3% lower than in the baseline equilibrium.

Why does the moving voucher policy not have large negative consequences on human capital accumulation given the tradeoff uncovered in Section 5? The impact of the policy on human capital accumulation can be decomposed into two effects. First, a reallocation effect. Moving vouchers reallocate workers to skill-dense places, and holding constant the quality of the local interactions, this effect boosts human capital accumulation as it increases the aggregate number of interactions with skilled partners. Second, a composition effect. As marginally less skilled workers are reallocated to productive cities, the quality of the interactions worsen, which affects negatively aggregate human capital accumulation. Formally, the change in the expected skill of a young worker with skill s between the policy and the baseline equilibrium is

$$\frac{\mathbb{E}[\bar{S}^o \mid s] - \mathbb{E}[S^o \mid s]}{\mathbb{E}[S^o \mid s]} \propto \underbrace{\sum_{\ell} \left(\frac{\Delta n_\ell^y(s)}{n^y(s)} \right) \int \gamma(s, s_p) \pi_\ell(s_p) ds_p}_{\text{Reallocation effect}} + \underbrace{\sum_{\ell} \frac{\bar{n}_\ell^y(s)}{n^y(s)} \int \gamma(s, s_p) \Delta\pi_\ell(s_p) ds_p}_{\text{Composition effect}}, \quad (29)$$

where $\bar{x}$ refers to variable x in the policy equilibrium and $\Delta x \equiv \bar{x} - x$.

The dashed orange line and the dotted grey line in Figure 9a display the aggregate reallocation and composition effects. Holding constant the quality of local interactions, the moving voucher policy would increase aggregate human capital by 1%. However, in general equilibrium, the composition effect fully offsets the partial equilibrium impact of the policy, and in net, the policy neither foster nor dampen aggregate learning.

The neutral effect of the policy on human capital accumulation holds the skill distribution. Figure 9b presents the change in expected skill between the policy and the baseline equilibrium across the skill distribution. Holding constant local interactions, most worker would on average experience faster learning under the moving voucher policy. The gains would be the largest for workers between the 50th and 95th percentile of the skill distribution for whom skilled interactions are the most beneficial. However, the general equilibrium effect of the policy neutralizes those partial equilibrium gains. Overall, the mild decrease in aggregate human capital is driven by workers in the top 5% of the skill distribution, who suffer average learning losses of 2% from the policy.

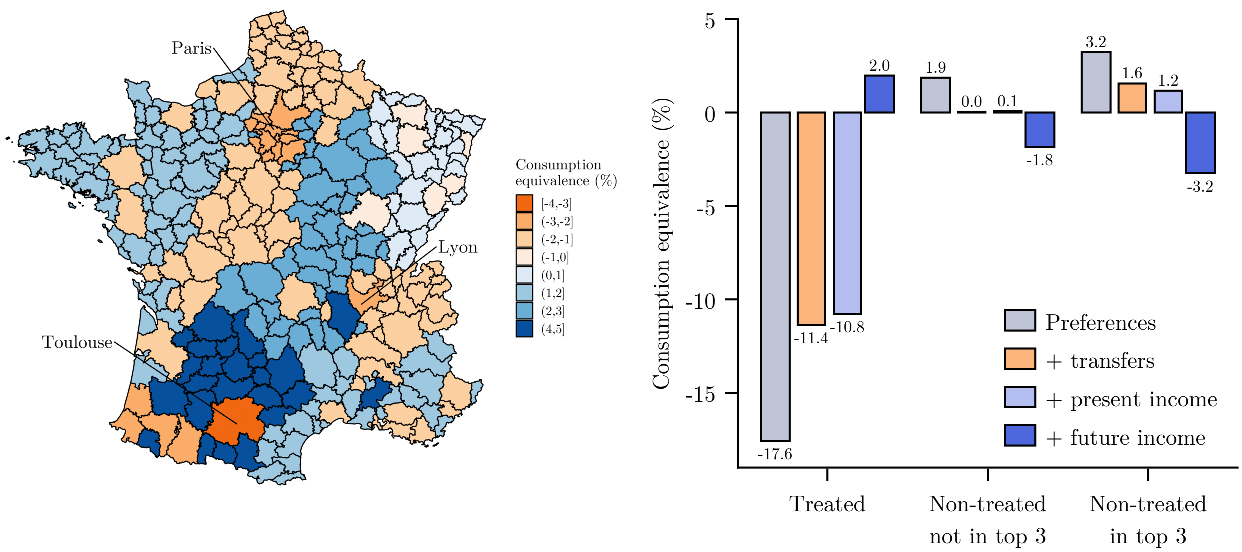
Taking stock, the moving voucher policy reduces spatial wage and learning inequality without reducing aggregate human capital accumulation. It is therefore able to achieve its equity goal without large efficiency losses. Nevertheless, the policy must be financed to be feasible. To conclude

[87] In general equilibrium, the change in the quality of the local interactions amplify the partial equilibrium effect of the policy. Holding constant the learning value of productive places, the moving voucher policies reduces spatial wage inequality only by 14%. As the composition of Paris, Lyon and Toulouse changes, skilled workers have less of an incentives to work in those cities, which contributes to further reduce spatial wage inequality.

Figure 10: The consequences of the moving voucher policy on welfare

(a) Change in welfare by birthplace

(b) Welfare decomposition by birthplaces



Note: panel (a) and (b) compares the welfare difference between the baseline equilibrium and the moving voucher equilibrium when the policy reaches 1.5% of GDP. Panel (a) displays the change in average welfare by birthplace assuming equal weights by skill within birthplaces. Panel (b) plots the change in average welfare by birthplace and initial skill, grouping cities into slow-growth cities (treated), middle-growth cities (non-treated), and fast-growth cities (non-treated but targeted by the policy).

this section, I therefore turn to the welfare consequences of the policy.

6.3 Welfare

What are the welfare impacts of the moving voucher policy? I measure welfare through the change in the expected lifetime utility of young workers between the baseline and the policy equilibrium. I express welfare changes in terms of consumption equivalence (Lucas, 1987; Alvarez and Jermann, 2004). Specifically, let $\mathcal{V}^y_\ell(s,\zeta)$ denote the lifetime utility of a young worker born in ℓ with skill s in the baseline economy when their lifetime consumption is multiplied by ζ.[88] Similarly, let $\mathcal{V}^{y\star}_\ell(s,\zeta)$ denote the lifetime utility of young workers in the policy economy. I measure the consumption-equivalent change in welfare by $\zeta_\ell(s)$ defined implicitly by $\mathcal{V}^y_\ell[s,\zeta_\ell(s)] = \tilde{\mathcal{V}}^y_\ell(s,0)$. Intuitively, $\zeta_\ell(s)$ represents the lifetime consumption gains (or losses) from the voucher policy for workers born in ℓ with skill s.

This metric can only be used to quantify the welfare impact of the policy for workers with identical skill born in the same location. In particular, it cannot be aggregated across skills or places. I therefore complement the worker-level metric with a welfare measure for the representative worker. Specifically, I measure the welfare impact of the policy on group g as the change in lifetime consumption required to make the average worker in that group indifferent between the two

[88]See Appendix I.1 for the full expression.

equilibrium,

$$\int \mathcal{V}^y_{\ell_i}(s_i,\zeta_g)dF^g_i = \int \tilde{\mathcal{V}}^y_{\ell_i}(s_i,0)d\tilde{F}^g_i,$$

where ζ_g is the welfare measure, and dF^g_i denote the distribution of workers in group g for simplicity.[89] This second welfare metric is consistent with a utilitarian social welfare function that places equal weights on workers within group g.[90] This section analyses the welfare consequences of the policy using the second metric for visual clarity. Figure G.17 reports the full distribution of welfare changes.

Figure 10a displays the welfare impact of the policy across birthplaces. Workers born in treated cities enjoy average welfare gains of 2%. Those gains are relatively larger for workers born near Paris, Lyon or Toulouse, going as high as 4.6%. Workers born in non-treated cities on the contrary suffer average welfare losses of 2.3%. These losses are relatively larger for the three cities targeted by the policy, with Toulouse experiencing losses of 3.5%.

Space, rather than skill, determines whether workers win or loses from the policy (Figure G.15b). For workers born in the treated locations, the welfare impacts of the policy are indeed homogeneous across workers. Meanwhile, all non-treated workers loose from the policy. For individuals born in non-treated cities outside of Paris, Lyon and Toulouse, the losses are slightly smaller for middle-skill workers, whereas they are the largest for the high-skill born in the three largest French cities.

In the aggregate, I find that the moving voucher policy generates aggregate welfare losses of 1.3% under a utilitarian social welfare function that places equal weights on skills and cities. Alternatively, a utilitarian planner must put Pareto weights four times larger on the treated workers for the policy to be welfare neutral.

What contributes to the welfare gains and losses from the policy? I show in Appendix I.1 that the welfare consequences of the voucher policy can be decomposed into four channels. First, its impact on workers' idiosyncratic preferences for the locations they live in. Second, the pecuniary value of the transfers net of the migration costs.[91] Third, how the vouchers affect workers' present real income, and fourth, how it affects expected future real income opportunities.

Figure 10b presents this welfare decomposition separately for treated workers, non-treated workers born outside of Paris, Lyon and Toulouse, and non-treated workers born in those three cities. The welfare impact of the policy on the treated is mainly explained by two offsetting forces. On the one hand, workers born in the treated locations that were choosing to stay there in the baseline equilibrium had on average strong preferences for these cities. On average, the policy is therefore reallocating treated workers to cities they dislike, which reduces their lifetime utility. On the other hand, the policy improves their future real income opportunities by increasing their lifetime human capital. In net, the human capital channel dominates.[92]

To understand the welfare impact of the policy on the non-treated, recall from Proposition 5

[89] Note that dF^g_i and $d\tilde{F}^g_i$ need not be the same as the policy is reallocating workers across space.

[90] Importantly however, no restrictions are placed on the social welfare function across groups. By making the groups more granular, one therefore recovers the worker-level welfare metric.

[91] For the non-treated workers, the transfers amount to the tax they have to pay to fiance the policy.

[92] Treated workers also experience welfare gains from the direct income effect of the subsidy.

that local interactions engender two externalities. First, a pulling externality: the concentration of skilled workers in a few cities requires many workers to forego their idiosyncratic spatial preferences. Second, a teaching externality: in the presence of strong within-skill complementarities, the economy features too few interactions between skilled workers. The voucher policy decreases the density of skilled workers in productive cities, and in doing so, mitigates the pulling externality while amplifying the teaching externality. Holding constant the human capital of non-treated workers, these workers would therefore also gain from the policy. However, in general equilibrium, the policy depresses their future income, which, together with the tax they have to pay to finance the policy, triggers welfare losses.

Moving vouchers are therefore an effective redistributive policy. They reduce substantially spatial wage and learning inequality with minimal aggregate productivity losses. Nevertheless, large moving vouchers are required to reduce spatial inequality, and the general equilibrium effects of the policy imposes welfare losses on non-treated workers.

7 Conclusion

This paper argues that a tradeoff exists between spatial inequality and human capital accumulation. I have shown theoretically that this tradeoff is shaped by the relative strength of within- and between-skill learning complementarities. I have documented that high-skill workers enjoy disproportionately higher wage growth when working in skill-dense cities or neighborhoods. I have argued that this pattern reveals the presence of strong within-skill learning complementarities. Using the estimated model, I concluded that local interactions increase the aggregate stock of human capital at the cost of amplified spatial inequality.

This equity-efficiency tradeoff matters for spatial policies. On the one hand, local interactions engender learning spillovers and spatial misallocation. Place-based policies that incentivizes a higher spatial concentration of skilled workers correct the local externalities. On the other hand, spatial skill concentration deteriorates the lifetime opportunities of workers born in remote cities. Vouchers that help workers relocate to productive cities have the potential to reduce these inequalities at small efficiency costs.

Skill complementarities are present everywhere. They shape our productivity. They matter for our preferences for cities. And finally, they define how we learn from each others. All these dimensions are sources of local spillovers. A natural direction along which to expand this research agenda is therefore to study optimal spatial policies in the presence of both static and dynamic complementarities. Finally, while firms may create teams that partially internalize the local learning spillovers, they are also constrained by the pool of workers they have access to. A second promising avenue for this agenda is therefore to explore how firms and space interact with each other in shaping human capital accumulation.

References

Adda, J. and Dustmann, C. (2023). Sources of wage growth. *Journal of Political Economy*, 131(2):456–503.

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

Akcigit, U., Caicedo, S., Miguelez, E., Stantcheva, S., and Sterzi, V. (2018). Dancing with the stars: Innovation through interactions. Technical report, National Bureau of Economic Research.

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

Alvarez, F. and Jermann, U. J. (2004). Using asset prices to measure the cost of business cycles. *Journal of Political economy*, 112(6):1223–1256.

Alvarez, F. E., Buera, F. J., and Lucas, R. E. (2013). Idea flows, economic growth, and trade. Technical report, National Bureau of Economic Research.

Andrews, I., Gentzkow, M., and Shapiro, J. M. (2017). Measuring the sensitivity of parameter estimates to estimation moments. *The Quarterly Journal of Economics*, 132(4):1553–1592.

Andrews, I., Stock, J. H., and Sun, L. (2019). Weak instruments in instrumental variables regression: Theory and practice. *Annual Review of Economics*, 11:727–753.

Angrist, J. D. (2014). The perils of peer effects. *Labour Economics*, 30:98–108.

Angrist, J. D. and Pischke, J.-S. (2009). *Mostly harmless econometrics: An empiricist's companion*. Princeton university press.

Arellano, M. and Bond, S. (1991). Some tests of specification for panel data: Monte carlo evidence and an application to employment equations. *The review of economic studies*, 58(2):277–297.

Atkin, D., Chen, M. K., and Popov, A. (2022). The returns to face-to-face interactions: Knowledge spillovers in silicon valley. Technical report, National Bureau of Economic Research.

Audretsch, D. B. and Feldman, M. P. (2004). Knowledge spillovers and the geography of innovation. In *Handbook of regional and urban economics*, volume 4, pages 2713–2739. Elsevier.

Baum-Snow, N., Gendron-Carrier, N., and Pavan, R. (2023). Local productivity spillovers. Technical report, Technical report.

Baum-Snow, N. and Pavan, R. (2012). Understanding the city size wage gap. *The Review of economic studies*, 79(1):88–127.

Baum-Snow, N. and Pavan, R. (2013). Inequality and city size. *Review of Economics and Statistics*, 95(5):1535–1548.

Behrens, K. and Robert-Nicoud, F. (2015). Agglomeration theory with heterogeneous agents. *Handbook of regional and urban economics*, 5:171–245.

Benabou, R. (1993). Workings of a city: location, education, and production. *The Quarterly Journal of Economics*, 108(3):619–652.

Bilal, A. (2023). The geography of unemployment. *The Quarterly Journal of Economics*, 138(3):1507–1576.

Buera, F. J. and Oberfield, E. (2020). The global diffusion of ideas. *Econometrica*, 88(1):83–114.

Cai, S., Caliendo, L., Parro, F., and Xiang, W. (2022). Mechanics of spatial growth. Technical report, National Bureau of Economic Research.

Card, D., Rothstein, J., and Yi, M. (2023). Location, location, location.

Chetty, R., Hendren, N., and Katz, L. F. (2016). The effects of exposure to better neighborhoods on children: New evidence from the moving to opportunity experiment. *American Economic Review*, 106(4):855–902.

Chetty, R., Hendren, N., Kline, P., and Saez, E. (2014). Where is the land of opportunity? the geography of intergenerational mobility in the united states. *The Quarterly Journal of Economics*, 129(4):1553–1623.

Combes, P.-P., Duranton, G., and Gobillon, L. (2008). Spatial wage disparities: Sorting matters! *Journal of urban economics*, 63(2):723–742.

Crews, L. G. (2023). A dynamic spatial knowledge economy.

Davis, D. R. and Dingel, J. I. (2019). A spatial knowledge economy. *American Economic Review*, 109(1):153–170.

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

de la Roca, J. and Puga, D. (2017). Learning by working in big cities. *The Review of Economic Studies*, 84(1):106–142.

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

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 Puga, D. (2004). Micro-foundations of urban agglomeration economies. In *Handbook of regional and urban economics*, volume 4, pages 2063–2117. Elsevier.

Duranton, G. and Puga, D. (2023). Urban growth and its aggregate implications. Working Paper 26591, National Bureau of Economic Research.

Eckert, F., Hejlesen, M., and Walsh, C. (2022). The return to big-city experience: Evidence from refugees in denmark. *Journal of Urban Economics*, 130:103454.

Eeckhout, J., Pinheiro, R., and Schmidheiny, K. (2014). Spatial sorting. *Journal of Political Economy*, 122(3):554–620.

Emanuel, N., Harrington, E., and Pallais, A. (2023). The power of proximity to coworkers: Training for tomorrow or productivity today?

Fajgelbaum, P. D. and Gaubert, C. (2020). Optimal Spatial Policies, Geography, and Sorting. *The Quarterly Journal of Economics*, 135(2):959–1036.

Fogli, A., Guerrieri, V., Ponder, M., and Prato, M. (2023). Scaling up the american dream: A dynamic analysis. Working paper.

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

Giannone, E. (2022). Skilled-biased technical change and regional convergence.

Glaeser, E. L. (1999). Learning in cities. *Journal of urban Economics*, 46(2):254–277.

Glaeser, E. L. and Gottlieb, J. D. (2008). The economics of place-making policies. Technical report, National Bureau of Economic Research.

Guvenen, F. (2009). An empirical investigation of labor income processes. *Review of Economic dynamics*, 12(1):58–79.

Guvenen, F., Kuruscu, B., Tanaka, S., and Wiczer, D. (2020). Multidimensional skill mismatch. *American Economic Journal: Macroeconomics*, 12(1):210–244.

Herkenhoff, K. F., Ohanian, L. E., and Prescott, E. C. (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 Moretti, E. (2019). Housing constraints and spatial misallocation. *American Economic Journal: Macroeconomics*, 11(2):1–39.

Huggett, M., Ventura, G., and Yaron, A. (2011). Sources of lifetime inequality. *American Economic Review*, 101(7):2923–2954.

INSEE (2020). Tableaux de l'economie française. Consommation des menages.

INSEE (2021). Emploi, chomage, revenues du travail. Salaire minimum de croissance.

Jacobs, J. (1969). *The economy of cities*. Random House.

Jaffe, A. B., Trajtenberg, M., and Henderson, R. (1993). Geographic localization of knowledge spillovers as evidenced by patent citations. *the Quarterly journal of Economics*, 108(3):577–598.

Jarosch, G., Oberfield, E., and Rossi-Hansberg, E. (2021). Learning from coworkers. *Econometrica*, 89(2):647–676.

Jovanovic, B. (2014). Misallocation and growth. *American Economic Review*, 104(4):1149–1171.

Jovanovic, B. and Rob, R. (1989). The growth and diffusion of knowledge. *The Review of Economic Studies*, 56(4):569–582.

Kennan, J. and Walker, J. R. (2011). The effect of expected income on individual migration decisions. *Econometrica*, 79(1):211–251.

Kline, P. and Moretti, E. (2014). People, places, and public policy: Some simple welfare economics of local economic development programs. *Annu. Rev. Econ.*, 6(1):629–662.

Lagakos, D., Moll, B., Porzio, T., Qian, N., and Schoellman, T. (2018). Life cycle wage growth across countries. *Journal of Political Economy*, 126(2):797–849.

Lhuillier, H. (2022). The local root of wage inequality.

Lindenlaub, I. (2017). Sorting multidimensional types: Theory and application. *The Review of Economic Studies*, 84(2):718–789.

Lise, J. and Postel-Vinay, F. (2020). Multidimensional skills, sorting, and human capital accumulation. *American Economic Review*, 110(8):2328–2376.

Lucas, R. E. (1987). *Model of business cycles*. Basil Blackwell New York.

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

Lucas, R. E. (2009b). Trade and the diffusion of the industrial revolution. *American Economic Journal: Macroeconomics*, 1(1):1–25.

Lucas, R. E. and Moll, B. (2014). Knowledge growth and the allocation of time. *Journal of Political Economy*, 122(1):1–51.

Manski, C. F. (1993). Identification of endogenous social effects: The reflection problem. *The review of economic studies*, 60(3):531–542.

Marshall, A. (1890). *The Principles of Economics*. McMaster University Archive for the History of Economic Thought.

Martellini, P. (2022). Local labor markets and aggregate productivity. Working paper.

Moretti, E. (2021). The effect of high-tech clusters on the productivity of top inventors. *American Economic Review*, 111(10):3328–3375.

Nix, E. (2020). Learning spillovers in the firm. Technical report, Working paper.

Papageorgiou, T. (2022). Occupational matching and cities. *American Economic Journal: Macroeconomics*, 14(3):82–132.

Perla, J. and Tonetti, C. (2014). Equilibrium imitation and growth. *Journal of Political Economy*, 122(1):52–76.

Perla, J., Tonetti, C., and Waugh, M. E. (2021). Equilibrium technology diffusion, trade, and growth. *American Economic Review*, 111(1):73–128.

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

Rossi-Hansberg, E., Sarte, P.-D., and Schwartzman, F. (2019). Cognitive hubs and spatial redistribution. Working Paper 26267, National Bureau of Economic Research.

Rossi-Hansberg, E. and Wright, M. L. (2007). Urban structure and growth. *The Review of Economic Studies*, 74(2):597–624.

Sampson, T. (2016). Dynamic selection: an idea flows theory of entry, trade, and growth. *The Quarterly Journal of Economics*, 131(1):315–380.

Stock, J. H. and Yogo, M. (2002). Testing for weak instruments in linear iv regression.

Topkis, D. M. (1998). *Supermodularity and complementarity*. Princeton university press.

Waldfogel, J. (2003). Preference externalities: An empirical study of who benefits whom in differentiated-product markets. *The RAND Journal of Economics*, 34(3):557–568.

Zabreĭko, P. P. (1975). *Integral equations—a reference text*. Springer.

Appendix

A Theory

A.1 Proof of Proposition 1

The problem of the young is decoupled from that of the old. The old allocation in turn depends on the aggregate old skill distribution, which follows from the young's allocation. To prove the existence of an equilibrium, I impose mild conditions on the distribution of the idiosyncratic learning shock.

Assumption 3 (Idiosyncratic learning shocks)**.**
Suppose that (i) the support of F is $\{0\} \cup \mathbb{R}_+$, and (ii) for any function $h : [\underline{x}, \bar{x}] \mapsto \mathbb{R}_+$ such that $h(x) < \infty$ for all $x \in [\underline{x}, \bar{x}]$, then $\int h(ex)dF(e) < \infty$ for all $x \in [\underline{x}, \bar{x}]$.

I start from the young spatial allocation. Existence of a solution to (4) is essentially given by Schauder fixed point theorem. I drop the y superscript for notational simplicity. Let $X \equiv [\underline{s}, \bar{s}]^L \subseteq \mathbb{R}_+^L$. Suppose for simplicity that $\underline{s} > 0$, $\bar{s} < \infty$, and $n(s) < \infty$ for all $s \in X$. Then, X is a compact convex subset of $\mathbb{R}_+^L$. Define $C_b(X, Y)$ as the set of bounded continuous functions that maps X into $Y \subset \mathbb{R}_L^+$, where $Y = [\underline{n}, \bar{n}]^L$ is also a compact convex subset of $\mathbb{R}_+^L$ with $\underline{n} > 0$ and $\bar{n} < \infty$. When equipped with the sup norm, the vector space $C_b(X, Y)$ is a Banach space. Define the operator $T : C_b(X, Y) \to Q$ by

$$T[\boldsymbol{f}](\boldsymbol{s}) = \left\{ n(s_\ell) \left(\frac{\mathrm{e}^{\vartheta(s_\ell T_\ell + \beta O(s_\ell, f_\ell))}}{\sum_{\ell'} \mathrm{e}^{\vartheta\left(s_\ell T_{\ell'} + \beta O(s_\ell, f_{\ell'})\right)}} \right) \right\}_{\ell=1}^{L}$$

for $\boldsymbol{f} \in C_b(X, Y)$, where

$$O(s, f) = \int \left(\int \mathbb{E}(V^o[e\gamma(s, s_p), \boldsymbol{\varepsilon}])dF(e) \right) \left(\frac{f(s_p)}{\int f(\sigma)d\sigma} \right) ds_p.$$

The second part of Assumption 3 guarantees that $O(s, f)$ exists. For any $\boldsymbol{f} \in C_b(X, Y)$, we know that

$$n(s) \left(\frac{\mathrm{e}^{\vartheta(sT_\ell + \beta O(s, f_\ell))}}{\sum_{\ell'} \mathrm{e}^{\vartheta\left(sT_{\ell'} + \beta O(s, f_{\ell'})\right)}} \right) \in (0, n(s)),$$

for all $s \in [\underline{s}, \bar{s}]$ and $\ell \in \{1, 2, \ldots, L\}$. Furthermore, Assumption 1 implies that $s \to O(s, n)$ is continuous, and therefore so is the above function. Therefore, $T : C_b(X, Y) \to C_b(X, Y)$. Finally, $O(s, f)$ is continuous in f since $f_\ell(s) > 0$ and $f_\ell(s) < \infty$ for all $s \in [\underline{s}, \bar{s}]$, all $\ell \in \{1, 2, \ldots, L\}$, and any $\boldsymbol{f} \in C_b(X, Y)$. It automatically follows that T is a continuous operator, and Schauder fixed point theorem implies that there exists at least one $f^\star \in C_b(X, Y)$ such that $T[\boldsymbol{f}^\star] = f^\star$.

Turning to the old allocation, (3) has a solution for all s as long as $n^o(s)$ has a solution. Using the first part of Assumption 3, the old skill distribution (5) can be rewritten

$$N^o(s^o) = \sum_\ell \int \int n_\ell^y(s_y) \pi_\ell^y(s_p) F\left(\frac{s_o}{\gamma(s_y, s_p)} \right) ds_p ds_y.$$

The two integrals always exist. Furthermore, N^o is continuous differentiable with $\lim_{s \to 0^+} N^o(s) = 0$ and $\lim_{s \to \infty} N^o(s) = 1$, and $n^o(s)$ exists.

A.2 Uniqueness and Stability of the Symmetric Equilibrium

Proposition A.1 (Stability and uniqueness)**.**
Suppose that $\boldsymbol{T} = 1$. Then, a symmetric equilibrium exists. Suppose further that $\gamma(s, s') = \mathcal{F}(s) + \mathcal{G}(s)\mathcal{H}(s')$, for $\mathcal{F}$, $\mathcal{G}$ and $\mathcal{H}$ some functions. Then, if

$$\vartheta\beta \max_\ell \max \{1 - \mathcal{B}_\ell^y, \mathcal{B}_\ell^y\} |\mathbb{C}ov[\mathcal{G}(S^y), \mathcal{H}(S^y)]| < 1,$$

the symmetric equilibrium is stable. Meanwhile, when L is large and

$$4\vartheta\beta \left(\max_s \mathcal{G}(s) - \min_s \mathcal{G}(s) \right) \left(\max_s \mathcal{H}(s) - \min_s \mathcal{H}(s) \right) < 1,$$

then the symmetric equilibrium is the unique equilibrium.

Proof. For simplicity, I do the proof assuming away the idiosyncratic learning shock. Then, under $\boldsymbol{T} = 1$, the expected utility of old workers before realizing their taste shocks is $\mathcal{V}^o(s) = \vartheta^{-1}\log s + c$ for some constant c. Since the rest of the equilibrium is fully determined by the young spatial allocation, I omit the y superscript. Given the old utility, the young expected utility before knowing what interactions they will experience is $O_\ell(s) = \mathcal{F}(s) + \mathcal{G}(s)\mathbb{E}[\mathcal{H}(S_\ell)]$. Let $\mu_\ell \equiv \mathbb{E}[\mathcal{H}(S_\ell)]$ and $\mu \equiv \mathbb{E}[\mathcal{H}(S)]$. The young spatial distribution of economic activity is given by

$$n_\ell(s) = n(s)\left(\frac{e^{\vartheta B_\ell + \vartheta\beta\mathcal{G}(s)\mu_\ell}}{\sum_{\ell'} e^{\vartheta B_{\ell'} + \vartheta\beta\mathcal{G}(s)\mu_{\ell'}}}\right).$$

Note that only the $\boldsymbol{\mu} \equiv \{\mu_\ell\}_\ell$ matters to pin down the young allocation. These local averages are in turn given by

$$\mu_\ell = \frac{\int \mathcal{H}(s)\left(\frac{e^{\vartheta\beta\mathcal{G}(s)\mu_\ell}}{\sum_{\ell'} e^{\vartheta B_{\ell'} + \vartheta\beta\mathcal{G}(s)\mu_\ell}}\right) n(s)ds}{\int\left(\frac{e^{\vartheta\beta\mathcal{G}(s')\mu_\ell}}{\sum_{\ell'} e^{\vartheta B_{\ell'} + \vartheta\beta\mathcal{G}(s')\mu_{\ell'}}}\right) n(s')ds'}.$$

Hence, we have reduced the problem to a L-dimensional fixed point. Let T be the operator that maps $\boldsymbol{\mu}$ onto itself as defined by the above L equations. Note that Assumption 1 requires $\mathcal{H}(s)$ to be bounded above and below. Let $\underline{h} \equiv \min_s \mathcal{H}(s)$ and $\bar{h} \equiv \max_s \mathcal{H}(s)$. Then, we know that $T : [\underline{h}, \bar{h}]^L \mapsto [\underline{h}, \bar{h}]^L$. Clearly, $\mu_\ell = \mu$ is a solution to this fixed point. A sufficient condition for the symmetric equilibrium to be stable is $|\partial_{\mu_{\ell'}} T_\ell(\boldsymbol{\mu})| < 1$ for all pair (ℓ, ℓ') and $\boldsymbol{\mu} = \{\mu\}_{\ell=1}^L$. A sufficient condition for the symmetric equilibrium to be unique is $|\partial_{\mu_{\ell'}} T_\ell(\boldsymbol{\mu})| < 1$ for all pair (ℓ, ℓ') and all $\boldsymbol{\mu}$, in which case T is a contraction.[93] For any $\boldsymbol{\mu}$, the partial derivative of T_ℓ w.r.t. μ_ℓ is

$$\frac{\partial T_\ell(\boldsymbol{\mu})}{\partial \mu_\ell} = \vartheta\beta \int \left(\mathcal{H}(s) - \mu_\ell\right)\mathcal{G}(s)\pi_\ell(s)\left(1 - \frac{n_\ell(s)}{n(s)}\right) ds,$$

and, for $\ell \neq \ell'$, it is

$$\frac{\partial T_\ell(\boldsymbol{\mu})}{\partial \mu_{\ell'}} = -\vartheta\beta \int (\mathcal{H}(s) - \mu_\ell)\mathcal{G}(s)\pi_\ell(s)\left(\frac{n_{\ell'}(s)}{n(s)}\right) ds.$$

Stability. When the equilibrium is symmetric, $\boldsymbol{\mu} = \{\mu\}_\ell$. Then, the derivatives simplify to

$$\frac{\partial T_\ell(\boldsymbol{\mu})}{\partial \mu_\ell} = \vartheta\beta\left(1 - \mathcal{B}_\ell\right)\int \left(\mathcal{H}(s) - \mu\right)\mathcal{G}(s)n(s)ds = \vartheta\beta\left(1 - \mathcal{B}_\ell\right)\mathbb{C}\text{ov}[\mathcal{G}(S), \mathcal{H}(S),]$$

and

$$\frac{\partial T_\ell(\boldsymbol{\mu})}{\partial \mu_{\ell'}} = -\vartheta\beta\mathcal{B}_{\ell'}\mathbb{C}\text{ov}[\mathcal{G}(S), \mathcal{H}(S)]$$

for $\ell' \neq \ell$, where

$$\mathcal{B}_\ell \equiv \frac{e^{\vartheta B_\ell}}{\sum_{\ell'} e^{\vartheta B_{\ell'}}}. \tag{30}$$

Imposing $|\partial_{\mu_{\ell'}} T_\ell(\boldsymbol{\mu})| < 1$ yields the first condition of Proposition A.1.

Uniqueness. For uniqueness, the above condition is necessary but not sufficient. To derive a sufficient condition, suppose that there are many cities. Since $\mathcal{S}^y$ is bounded, we know that $|\mu_\ell - \mu_{\ell'}| \leq \bar{s} - \underline{s} < \infty$ for all (ℓ, ℓ'). That is, no city will ever have learning opportunities so attractive as to capture the entire employment share. As a consequence,

[93] To see this, note that the condition implies $\|DT(\boldsymbol{\mu})\| \leq K < 1$. Then, take two points $\boldsymbol{\mu}$ and $\boldsymbol{\mu}'$ in $[\underline{s}, \bar{s}]^L$. Define $\Delta(\lambda) \equiv \lambda\boldsymbol{\mu} + (1-\lambda)\boldsymbol{\mu}'$, for $\lambda \in [0,1]$. Clearly, $\partial_\lambda\Delta(\lambda) = \boldsymbol{\mu} - \boldsymbol{\mu}' \perp \lambda$. Hence, it follows that

$$T(\boldsymbol{\mu}) - T(\boldsymbol{\mu}') = T[\Delta(1)] - T[\Delta(0)] = \int_0^1 \frac{dT[\Delta(\lambda)]}{d\lambda} d\lambda = \int_0^1 dT[\Delta(\lambda)]\frac{\partial\Delta(\lambda)}{\partial\lambda} d\lambda,$$

where the second equality is from the fundamental theorem of calculus and the third from the chain rule. Thus, applying the norm $\|\cdot\|$ on both sides,

$$\|T(\boldsymbol{\mu}) - T(\boldsymbol{\mu}')\| = \|\int_0^1 dT[\Delta(\lambda)]\frac{\partial\Delta(\lambda)}{\partial\lambda} d\lambda\| \leq \int_0^1 \|dT[\Delta(\lambda)]\|\|\frac{\partial\Delta(\lambda)}{\partial\lambda}\| d\lambda \leq \int_0^1 K\|\boldsymbol{\mu} - \boldsymbol{\mu}'\| d\lambda = K\|\boldsymbol{\mu} - \boldsymbol{\mu}'\|,$$

where the first inequality is from the triangle inequality. Since $K < 1$, T is a contraction.

if amenities are independent of L, then $\lim_{L\to\infty} \mathcal{B}_\ell \to 0^+$ for all ℓ. Hence, for L large, the partial derivatives rewrite

$$\frac{\partial T_\ell(\boldsymbol{\mu})}{\partial \mu_\ell} \approx \vartheta\beta \int \left(\mathcal{H}(s) - \mu_\ell\right) \mathcal{G}(s)\pi_\ell(s)ds = \vartheta\beta\mathbb{C}\mathrm{ov}[\mathcal{G}(S_\ell), \mathcal{H}(S_\ell)]$$

and $\partial_{\mu_{\ell'}} T_\ell(\boldsymbol{\mu}) \approx 0$. If we knew the local distributions $\{\pi_\ell\}$, then uniqueness would be guaranteed by

$$\vartheta\beta < \frac{1}{|\mathbb{C}\mathrm{ov}[\mathcal{G}(S_\ell), \mathcal{H}(S_\ell)]|}.$$

However, the local distributions are themselves a function of $(\vartheta, \beta, \gamma)$. Yet, it is possible to bound the covariance since we know that the $\mathcal{G}(S_\ell)$ and $\mathcal{H}(S_\ell)]$ are bounded. This yields the second condition of Proposition A.1. $\square$

A.3 The Sorting of Old Workers

Proposition A.2 (Sorting)**.**
The old allocation satisfies SPAM.

Proof. For the entire proof I omit the old superscript for notational simplicity. I first derive an expression for the difference in density between ℓ and ℓ':

$$\pi_\ell(s) - \pi_{\ell'}(s) = \left(\frac{N_{\ell'}}{N_\ell}\right) \pi_{\ell'}(s) \mathrm{e}^{\vartheta\left(A_\ell - A_{\ell'}\right)} \left(\mathrm{e}^{\vartheta s(T_\ell - T_{\ell'})} - \mathbb{E}\left[\mathrm{e}^{\vartheta S_{\ell'}(T_\ell - T_{\ell'})}\right]\right). \tag{31}$$

This follows from the definition of $\pi_\ell(s)$, N_ℓ and (3). Let $f : \mathcal{S}^o \mapsto \mathbb{R}$ be any measurable function. Using (31), the difference in the f-transformed mean skill between city ℓ and ℓ' is

$$\mathbb{E}[f(S_\ell)] - \mathbb{E}[f(S_{\ell'})] = \left(\frac{N^o_{\ell'}}{N^o_\ell}\right) \mathrm{e}^{\vartheta\left(A_\ell - A_{\ell'}\right)} \mathbb{C}\mathrm{ov}\left(f(S_{\ell'}), \mathrm{e}^{\vartheta S_{\ell'}(T_\ell - T_{\ell'})}\right).$$

Take any non-decreasing function f. Then, the above expression implies $\mathbb{E}[f(S_\ell)] > \mathbb{E}[f(S_{\ell'})]$ iff $T_\ell > T_{\ell'}$. That is, the local skill densities of old workers can be FOSD-ordered by T_ℓ. $\square$

A.4 The First Order Approximation

A.4.1 The general learning technology

Old workers Linearizing the old allocation is straightforward. From (3), we know that

$$\left.\frac{\partial n^o_\ell(s)}{\partial T_{\ell'}}\right|_{\boldsymbol{T}=1} = \mathcal{B}^o_\ell \left(\frac{\partial n^o(s)}{\partial T_{\ell'}} + \vartheta s \mathrm{n}^o(s) I_{\ell\ell'}\right),$$

for $I_{\ell\ell'} \equiv \mathbb{1}\{\ell' = \ell\} - \mathcal{B}^o_{\ell'}$ and $\mathcal{B}^o_\ell$ defined in (30). The first term is the GE effect of increasing the TFP of city ℓ' on the density of skill s. This term will be nil in equilibrium (see (41)). The second term is the reallocation of workers over space. Hence, to a first order, the spatial allocation of old workers is

$$n^o_\ell(s) \approx \mathrm{n}^o(s)\mathcal{B}^o_\ell + \vartheta \mathrm{n}^o(s)\mathcal{B}^o_\ell \mathcal{T}^o_\ell s, \tag{32}$$

where $\mathcal{T}^o_\ell \equiv T_\ell - \sum_{\ell'} \mathcal{B}^o_{\ell'} T_{\ell'}$. Second, the partial derivative of the within-city skill density is

$$\left.\frac{\partial \pi^o_\ell(s)}{\partial T_{\ell'}}\right|_{\boldsymbol{T}=1} = \vartheta \mathrm{n}^o(s)\left(s - \mathbb{E}[\mathcal{S}^o]\right) I_{\ell\ell'},$$

since $\int \partial_{T_{\ell'}} n^o(s)ds = 0$. It follows that

$$\pi^o_\ell(s) \approx \mathrm{n}^o(s) + \vartheta \mathrm{n}^o(s)\mathcal{T}^o_\ell \left(s - \mathbb{E}[\bar{S}^o]\right).$$

The average wage of old workers in city ℓ is then given by

$$\mathbb{E}_\ell[\mathcal{W}^o] = T_\ell \int s\pi^o_\ell(s)ds \approx \mathbb{E}[\mathcal{S}^o] + \vartheta \mathcal{T}^o_\ell \mathbb{V}\mathrm{ar}[\mathcal{S}^o].$$

Young workers For the young, given (4), we have

$$\left.\frac{\partial n_\ell^y(s)}{\partial T_{\ell'}}\right|_{\boldsymbol{T}=1} = \vartheta \mathcal{B}_\ell^y n^y(s)\left[sI_{\ell\ell'} + \beta\left(\frac{\partial O_\ell}{\partial T_{\ell'}} - \sum_l \mathcal{B}_l^y \frac{\partial O_l(s)}{\partial T_{\ell'}}\right)\right].$$

The first term is identical to the old allocation. The second term captures how the local learning varies with $\boldsymbol{T}$. Recall that

$$O_\ell(s) = \int\int \mathcal{V}^o[e\gamma(s,s_p)]\pi_\ell^y(s_p)dF(e)ds_p.$$

When $\boldsymbol{T}=1$, $\mathcal{V}^o(s) = s + c$ for some constant c. In addition, $\partial_{T_\ell}\mathcal{V}^o(s) = s\mathcal{B}_\ell^o$. Hence,

$$\left.\frac{\partial O_\ell(s)}{\partial T_{\ell'}}\right|_{\boldsymbol{T}=1} = \mu_e \mathcal{B}_{\ell'}^o \int \gamma(s,s_p)n^y(s_p)ds_p + \mu_e \int \gamma(s,s_p)\partial_{T_{\ell'}}\pi_\ell^y(s_p)ds_p + c,$$

where μ_e is the mean of the learning shock, $\mu_e \equiv \int edF(e)$. Using $\sum_\ell n_\ell^y(s) = n^y(s)$ for all $\boldsymbol{T}$, it follows that

$$\frac{\partial O_\ell}{\partial T_{\ell'}} - \sum_l \mathcal{B}_l^y \frac{\partial O_l}{\partial T_{\ell'}} = \int \frac{\partial \pi_\ell^y(s_p)}{\partial T_{\ell'}}\gamma(s,s_p)ds_p = \frac{1}{\mathcal{B}_\ell^y}\int \frac{\partial n_\ell^y(s_p)}{\partial T_{\ell'}}\left(\gamma(s,s_p) - \mathbb{E}[\gamma(s,\mathcal{S}^y)]\right)ds_p,$$

where the second equality follows from the expression for $\partial_{T_{\ell'}}\pi_\ell^y(s_p)$ and I have redefined $\gamma = \mu_e\gamma$ wlog.[94] Plugged in the initial expression, we can conclude that

$$\left.\frac{\partial n_\ell^y(s)}{\partial T_{\ell'}}\right|_{\boldsymbol{T}=1} = \vartheta n^y(s)\left(\mathcal{B}_\ell^y sI_{\ell\ell'} + \beta\int \frac{\partial n_\ell^y(s_p)}{\partial T_{\ell'}}\left(\gamma(s,s_p) - \mathbb{E}[\gamma(s,S^y)]\right)ds_p\right).$$

The above expression constitutes a fixed point over the functions $\partial_{T_{\ell'}}n_\ell^y(s)$. I solve this fixed point in two steps. First, I guess and verify the general form of the solution:

$$\frac{\partial n_\ell^y(s)}{\partial T_{\ell'}} = \vartheta n^y(s)\mathcal{B}_\ell^y I_{\ell\ell'}\eta(s), \tag{33}$$

for η some city-independent function. From this guess, we also know that the partial derivative of the local skill density is

$$\frac{\partial \pi_\ell^y(s)}{\partial T_{\ell'}} = \vartheta n^y(s)I_{\ell\ell'}\left(\eta(s) - \mathbb{E}[\eta(S^y)]\right). \tag{34}$$

Second, plugging the guess in the expression for $\partial_{T_{\ell'}}n_\ell^y(s)$ verifies it and yields the implicit expression for η:

$$\eta(s) = s + \vartheta\beta\int\left(\gamma(s,s_p) - \mathbb{E}[\gamma(s,S^y)]\right)n^y(s_p)\eta(s_p)ds_p = s + \vartheta\beta\Theta(s),$$

where $\Theta(s)$ is defined in (6). The above expression is a Fredholm integral equation of the second type. Standard results can be used to show that a solution exists when the dispersion forces, summarized here by ϑ, are large enough.

Lemma A.1 (Existence)**.**
Suppose that

$$\vartheta\beta\left(\int\int |K(s,s_p)|^2 dN(s_p)dN(s)\right)^{1/2} < 1.$$

Then, a solution to (6) *exists, is unique, and is continuous and bounded in* $[\underline{s},\bar{s}]$.

Proof. I omit the y superscript for simplicity. Define $K(s,s_p) \equiv \gamma(s,s_p) - \mathbb{E}[\gamma(s,S)]$ so as to rewrite (6) as

$$\eta(s) = s + \vartheta\beta\int K(s,s_p)\eta(s_p)dN(s).$$

This constitutes a Fredholm equation of the second kind. I restrict my attention to solutions to (6) that belong to

[94] Clearly the mean of the learning shock and the average learning γ cannot be separately identified from this theory. From the estimation, I assume that e is log-normally distributed and I normalize the mean of e to zero.

$L_2(dN, [\underline{s}, \bar{s}])$. Under Assumption 1 and $\bar{s} < \infty$, we have[95]

$$\int |K(s, s_p)|^2 dN(s_p) \leq C \quad \text{and} \quad \int |s|^2 dN(s) < \infty,$$

for C some constant. In addition, K is globally continuous with respect to x. It follows that any solution to (6) is bounded and continuous in $[\underline{s}, \bar{s}]$ (Zabreĭko, 1975, Chapter 2, Theorems 1.7 and 1.8).

Uniqueness of a solution to (6) implies existence (Zabreĭko, 1975, Theorem 1.2). Equation (6) has a unique solution when $\vartheta\beta < |\bar{\lambda}|$, where $\bar{\lambda}$ is the characteristic value of K with smallest magnitude. This characteristic value can be solved for when imposing a separability condition on γ (see Section A.4.2). Without this restriction, solving for $|\bar{\lambda}|$ is non-trivial. However, if the Neumann series

$$\eta^f(s) = f(s) + \sum_{j=1}^{\infty} (\vartheta\beta)^j \int K_j(x,t) t dN(t) \tag{35}$$

converges for any function f satisfying $\int |f(s)|^2 dN(s) < \infty$, where K_j is the j-th iterated kernel of K, then $\vartheta\beta < |\bar{\lambda}|$ (e.g. Zabreĭko, 1975, Chapter 2 §2). The inequality $\vartheta\beta\|K\| < 1$ provides a sufficient condition for (35) to converge. □

From there, the young spatial allocation is given to a first order by

$$n_\ell^y(s) \approx n^y(s)\mathcal{A}_\ell + \vartheta n^y(s)\mathcal{A}_\ell \mathcal{T}_\ell^y s + \vartheta^2 \beta n^y(s) \mathcal{A}_\ell \mathcal{T}_\ell^y \Theta(s).$$

The within-city young skill distribution is

$$\pi_\ell^y(s) \approx n^y(s) + \vartheta n^y(s)\mathcal{T}_\ell^y \left(s - \mathbb{E}[S^y]\right) + \vartheta^2 \beta n^y(s) \mathcal{T}_\ell^y \left(\Theta(s) - \mathbb{E}[\Theta(S^y)]\right). \tag{36}$$

Accordingly, the young average wage in city ℓ is

$$\mathbb{E}[\mathcal{W}^y] \approx T_\ell \mathbb{E}[S^y] + \vartheta \mathcal{T}_\ell^y \mathbb{V}\mathrm{ar}[S^y] + \vartheta^2 \beta \mathcal{T}_\ell^y \mathbb{C}\mathrm{ov}[S^y, \Theta(S^y)].$$

It only remains to derive the expression for the expected local skill growth and lifetime earnings. The expected skill growth for young workers with skill s living in city ℓ is

$$\mathbb{E}_\ell \left[\frac{S'}{s} \mid s\right] = \int g(s, s_p) \pi_\ell^y(s_p) ds_p, \tag{37}$$

where $g(s, s_p) \equiv \gamma(s, s_p)/s$, and, as before, I have integrated the average learning shock μ_e in γ. This statistic only depends on π_ℓ^y. Hence, the partial derivative of skill growth w.r.t. the TFP of city ℓ' is

$$\frac{\partial}{\partial_{T_{\ell'}}} \mathbb{E}\left[\frac{S'}{s} \mid s, \ell\right]\Bigg|_{\boldsymbol{T}=1} = \vartheta I_{\ell\ell'} \left(\mathbb{C}\mathrm{ov}[g(s,S), S] + \vartheta\beta \mathbb{C}\mathrm{ov}[g(s,S), \Theta(S)]\right).$$

Accordingly, to a first order, the skill growth experienced by workers with skill s in city ℓ is

$$\mathbb{E}_\ell \left[\frac{S'}{s} \mid s\right] \approx \mathbb{E}[g(s,S)] + \vartheta \mathcal{T}_\ell^y \mathbb{C}\mathrm{ov}[g(s,S), S] + \vartheta^2 \beta \mathcal{T}_\ell^y \mathbb{C}\mathrm{ov}[g(s,S), \Theta(S)]. \tag{38}$$

To compute the unconditional skill growth at the city level, (37) has to be integrated across workers,

$$\mathbb{E}_\ell \left[\frac{S'}{s}\right] = \int \int g(s, s_p) \pi_\ell^y(s_p) \pi_\ell^y(s) ds_p ds.$$

The partial derivative of this expression w.r.t. $T_{\ell'}$ is

$$\frac{\partial}{\partial_{T_{\ell'}}} \mathbb{E}_\ell \left[\frac{s'}{s}\right]\Bigg|_{\boldsymbol{T}=1} = \int \int g(s,s') \frac{\partial \pi_\ell^y(s')}{\partial T_{\ell'}} n^y(s) ds' ds + \int \int g(s,s') n^y(s') \frac{\partial \pi_\ell^y(s)}{\partial T_{\ell'}} ds' ds.$$

Using the expression for $\partial_{T_{\ell'}} \pi_\ell^y$, we obtain

$$\begin{aligned}\mathbb{E}_\ell \left[\frac{s'}{s}\right] \approx \mathbb{E}[g(S,S_p)] + \vartheta \mathcal{T}_\ell^y \mathbb{C}\mathrm{ov}\left\{\mathbb{E}[g(S,S_p) \mid S], S\right\} + \vartheta \mathcal{T}_\ell^y \mathbb{C}\mathrm{ov}\left\{\mathbb{E}[g(S,S_p) \mid S_p], S_p\right\} + \\ \vartheta^2 \beta \mathcal{T}_\ell^y \mathbb{C}\mathrm{ov}\left\{\mathbb{E}[g(S,S_p) \mid S], \Theta(S)\right\} + \vartheta^2 \beta \mathcal{T}_\ell^y \mathbb{C}\mathrm{ov}\left\{\mathbb{E}[g(S,S_p) \mid S_p], \Theta(S_p)\right\}.\end{aligned} \tag{39}$$

[95]As mentioned in the main text, $\bar{s} = \infty$ is possible as long as $\int |s|^2 dN(s) < \infty$ is satisfied.

Finally, skill growth maps into lifetime income. The expected lifetime income of a young worker with skill s initially living in city ℓ, denoted $I_\ell(s)$, is

$$I_\ell(s) = T_\ell s + \sum_{\ell'} \int T_{\ell'} \left(\frac{n^o_{\ell'}[\gamma(s,s_p)]}{n^o[\gamma(s,s_p)]} \right) \gamma(s,s_p) \pi^y_\ell(s_p) ds_p.$$

The first term is today's wages, while the second term is workers' expected future income. The derivative of this expression w.r.t. $T_{\ell'}$ is

$$\left. \frac{\partial I_\ell(s)}{\partial T_{\ell'}} \right|_{\boldsymbol{T}=1} = \mathbb{1}\{\ell' = \ell\} s + \mathcal{B}^o_{\ell'} \mathbb{E}[\gamma(s,S_p)] + \vartheta I_{\ell\ell'} \mathbb{C}\mathrm{ov}[\gamma(s,S_p), S_p] + \vartheta^2 \beta I_{\ell\ell'} \mathbb{C}\mathrm{ov}[\gamma(s,S_p), \Theta(S_p)].$$

It follows that the expected lifetime income of a young worker with skill s is

$$I_\ell(s) \approx T_\ell s + \sum_{\ell'} T_{\ell'} \mathcal{B}^o_{\ell'} \mathbb{E}[\gamma(s,S_p)] + \vartheta \mathcal{T}^y_\ell \mathbb{C}\mathrm{ov}[\gamma(s,S_p), S_p] + \vartheta^2 \beta \mathcal{T}^y_\ell \mathbb{C}\mathrm{ov}[\gamma(s,S_p), \Theta(S_p)]. \tag{40}$$

Aggregates I have argued earlier that, to a first order, there is no aggregate effect of cities on the old skill distribution, $\partial_{T_\ell} N^o(s) = 0$. Differentiating (5) w.r.t. $T_{\ell'}$ yields

$$\left. \frac{\partial N^o(s_o)}{\partial T_{\ell'}} \right|_{\boldsymbol{T}=1} = \sum_\ell \mathcal{B}^y_\ell \int\int \frac{\partial \pi^y_\ell(s_p)}{\partial T_{\ell'}} n^y(s_y) F\left(\frac{s_o}{\gamma(s_y,s_p)} \right) ds_p ds_y + \sum_\ell \int\int n^y(s_p) \frac{\partial n^y_\ell(s_y)}{\partial T_{\ell'}} F\left(\frac{s_o}{\gamma(s_y,s_p)} \right) ds_p ds_y.$$

Using the expression for $\partial_{T_{\ell'}} \pi^y_\ell$ and $\partial_{T_{\ell'}} n^y_\ell$, this turns into

$$\begin{aligned} \left. \frac{\partial N^o(s_o)}{\partial T_{\ell'}} \right|_{\boldsymbol{T}=1} = \vartheta \sum_\ell \mathcal{B}^y_\ell I_{\ell\ell'} \int\int \left(\eta(s_p) - \mathbb{E}[\eta(S_p)]\right) F\left(\frac{s_o}{\gamma(s_y,s_p)} \right) dN^y(s_p) dN^y(s) + \\ \vartheta \sum_\ell \mathcal{B}^y_\ell I_{\ell\ell'} \int\int \eta(s_y) F\left(\frac{s_o}{\gamma(s_y,s_p)} \right) dN^y(s_p) dN^y(s_y). \end{aligned} \tag{41}$$

However, $\sum_\ell \mathcal{B}^y_\ell I_{\ell\ell'} = 0$, and therefore $\partial_{T_{\ell'}} N^o(s_o) = 0$.

A.4.2 An example

Suppose that the learning technology reads:

$$\gamma(s,s_p) = \mathcal{F}(s) + \mathcal{G}(s)\mathcal{H}(s_p),$$

for $\mathcal{F}$, $\mathcal{G}$ and $\mathcal{H}$ functions that satisfy Assumption 1. This class of learning technology contains (1), e.g. with $\mathcal{F}(s) = g_0 + g_1 s$, $\mathcal{G}(s) = g_2 + g_{12} s$ and $\mathcal{H}(s_p) = s_p$, but also many more. Then, the integral equation simplifies to

$$\eta(s) = s + \vartheta\beta\mathcal{G}(s) \int \left(\mathcal{H}(s_p) - \mathbb{E}[\mathcal{H}(S^p)]\right) n^y(s_p) \eta(s_p) ds_p. \tag{42}$$

Solving for this fixed point amounts to solving for the integral on the right-hand side. Multiplying both sides by $(\mathcal{H}(s) - \mathbb{E}[\mathcal{H}(s)]) n^y(s)$, integrating over s and re-arranging, we get

$$(1 - \vartheta\beta\mathbb{C}\mathrm{ov}[\mathcal{G}(S^y), \mathcal{H}(S^y)]) \int (\mathcal{H}(s) - \mathbb{E}[\mathcal{H}(S^y)]) n^y(s) \eta(s) ds = \mathbb{C}\mathrm{ov}[S^y, \mathcal{H}(S^y)].$$

From here, it is easy to show that the characteristic root of (42) is $|\bar{\lambda}| = |1/\mathbb{C}\mathrm{ov}[\mathcal{G}(S^y), \mathcal{H}(S^y)]|$. Hence, a solution to (42) exists and is unique when $\vartheta\beta|\mathbb{C}\mathrm{ov}[\mathcal{G}(S^y), \mathcal{H}(S^y)]| \leq 1$. This condition is tighter than the one laid out in Lemma A.1. When it is satisfied, we can plug back the expression for the integral into (42) to obtain a closed-form expression for Θ:

$$\Theta(s) = \mathcal{G}(s) \left(\frac{\mathbb{C}\mathrm{ov}[S^y, \mathcal{H}(S^y)]}{1 - \vartheta\beta\mathbb{C}\mathrm{ov}[\mathcal{G}(S^y), \mathcal{H}(S^y)]} \right).$$

If the covariance between S^y and $\mathcal{H}(S^y)$ is positive, so that, on average, workers learn more from skilled individuals, then Θ inherits the property of $\mathcal{G}(s)$. If $\mathcal{G}(s) > 0$, then worker s learns more from skilled partner, and $\Theta(s) > 0$. If

$\mathcal{G}'(s) > 0$, then so is $\Theta(s)$. Furthermore, Θ is independent from $\mathcal{F}(s)$. Using $\mathcal{G}(s) = g_2 + g_{12}s$ and $\mathcal{H}(s_p) = s_p$ yields

$$\Theta(s) = \theta\left(g_2 + g_{12}s\right) \quad \text{where} \quad \theta = \frac{\mathbb{V}\text{ar}[S^y]}{1 - \vartheta\beta g_{12}\mathbb{V}\text{ar}[S^y]},$$

as claimed in the main text.

A.5 The Second Order Approximation

The second order approximations are implemented assuming identical amenities over space.[96]

Old skill The expected old skill of a worker with initial skill s reads

$$\mathbb{E}[S^o \mid s] = \sum_\ell \left(\frac{n_\ell^y(s)}{n^y(s)}\right)\mathbb{E}_\ell[S^o \mid s] = \sum_\ell \left(\frac{n_\ell^y(s)}{n^y(s)}\right)\int \gamma(s_y, s_p)\pi_\ell^y(s_p)ds_p,$$

while the average skill of old workers is

$$\mathbb{E}[S^o] = \int \mathbb{E}[S^o \mid s]dN^y(s).$$

In both expressions I have subsumed the expected learning shock $\mu_e = \int edF(e)$ in γ wlog. The partial derivative of the expected skill for workers with skill s w.r.t. the TFP of city ℓ is

$$\frac{\partial\mathbb{E}[S^o \mid s]}{\partial T_l} = \frac{1}{n^y(s)}\left(\sum_\ell \int \gamma(s,s')\left(\frac{\partial n_\ell^y(s)}{\partial T_l}\right)\pi_\ell^y(s')ds' + \sum_\ell \int \gamma(s,s')n_\ell^y(s)\left(\frac{\partial \pi_\ell^y(s')}{\partial T_l}\right)ds'\right).$$

By a similar argument as in (41), there is no aggregate effect of local interactions on workers' expected skill at the first order. Hence, moving to the second order, differentiate the above expression w.r.t. $T_{l'}$ to get

$$\left.\frac{\partial^2\mathbb{E}\left[S^o \mid s\right]}{\partial l\partial l'}\right|_{\boldsymbol{T}=1} = \frac{1}{n^y(s)}\left(\int \gamma(s,s')\underbrace{\left(\sum_\ell \partial^2_{ll'}n_\ell^y(s)\right)}_{=0}n^y(s')ds' + \frac{1}{L}\int \gamma(s,s')n^y(s)\left(\sum_\ell \partial^2_{ll'}\pi_\ell^y(s')\right)ds'\right) +$$
$$\frac{1}{n^y(s)}\left(\int \gamma(s,s')\left(\sum_\ell \partial_l n_\ell^y(s)\partial_{l'}\pi_\ell^y(s')\right)ds' + \int \gamma(s,s')\left(\sum_\ell \partial_{l'}n_\ell^y(s)\partial_l\pi_\ell^y(s')\right)ds'\right),$$

where the equality on the first line follows from the fixed supply of young skill. Meanwhile, the second-order change in $\pi_\ell^y(s)$ reads

$$\frac{\partial^2\pi_\ell^y(s)}{\partial T_l\partial T_{l'}} = \frac{1}{N_\ell}\left\{\frac{\partial^2 n_\ell^y(s)}{\partial T_l\partial T_{l'}} - \frac{\partial n_\ell^y(s)}{\partial T_l}\frac{\partial \log N_\ell}{\partial T_{l'}} - \frac{\partial n_\ell^y(s)}{\partial T_{l'}}\frac{\partial \log N_\ell}{\partial T_l} + 2n_\ell^y(s)\frac{\partial \log N_\ell}{\partial T_{l'}}\frac{\partial \log N_\ell}{\partial T_l} - \pi_\ell^y(s)\frac{\partial^2 N_\ell}{\partial T_l\partial T_{l'}}\right\}.$$

Hence, when evaluated at $\boldsymbol{T} = 1$ and summed across ℓ, we end up with

$$\sum_\ell \left.\frac{\partial^2\pi_\ell^y(s)}{\partial T_l\partial T_{l'}}\right|_{\boldsymbol{T}=1} = -L\sum_\ell \left(\frac{\partial\pi_\ell^y(s)}{\partial T_l}\frac{\partial N_\ell}{\partial T_{l'}} + \frac{\partial\pi_\ell^y(s)}{\partial T_{l'}}\frac{\partial N_\ell}{\partial T_l}\right),$$

since $\partial_l\pi_\ell^y(s)|_{\boldsymbol{T}=1} = L\left(\partial_l n_\ell^y(s) - n^y(s)\partial_l N_\ell\right)$. Plugged in the expression for the aggregate change in old skill, we therefore obtain

$$\left.\frac{\partial^2\mathbb{E}[S^o \mid s]}{\partial T_l\partial T_{l'}}\right|_{\boldsymbol{T}=1} = \frac{1}{n^y(s)L}\int \gamma(s,s')\sum_\ell \left(\frac{\partial\pi_\ell^y(s)}{\partial T_l}\frac{\partial\pi_\ell^y(s')}{\partial T_{l'}} + \frac{\partial\pi_\ell^y(s)}{\partial T_{l'}}\frac{\partial\pi_\ell^y(s')}{\partial T_l}\right)ds',$$

[96] The aggregate effects on productivity are due to the spatial variations in local skill density, see (36). The local skill densities are independent of city amenity. Adding city amenity would therefore not change the conclusion.

which follows again from $\partial_l \pi_\ell^y(s)|_{\boldsymbol{T}=1} = L\left(\partial_l n_\ell^y(s) - n^y(s)\partial_l N_\ell\right)$. This expression relies entirely on first order micro term which we have derived in Section A.4.1. Using (34), we get

$$\sum_\ell \frac{\partial \pi_\ell^y(s)}{\partial T_l}\frac{\partial \pi_\ell^y(s')}{\partial T_{l'}} = \vartheta^2 I_{ll'}\left(\eta(s) - \mathbb{E}[\eta(S^y)]\right)\left(\eta(s') - \mathbb{E}[\eta(S^y)]\right) n^y(s')n^y(s),$$

which follows from $\sum_\ell I_{\ell l} I_{\ell l'} = I_{ll'}$. But I is symmetric function, $I_{ll'} = I_{l'l}$. Hence, $\sum_\ell \partial_l \pi_\ell^y(s)\partial_{l'}\pi_\ell^y(s') = \sum_\ell \partial_{l'}\pi_\ell^y(s)\partial_l \pi_\ell^y(s')$, and we end up with

$$\left.\frac{\partial^2 \mathbb{E}[S^o \mid s]}{\partial T_l \partial T_{l'}}\right|_{\boldsymbol{T}=1} = \frac{2}{L}\vartheta^2 I_{ll'}\int \gamma(s,s')\left(\eta(s) - \mathbb{E}[\eta(S^y)]\right)\left(\eta(s') - \mathbb{E}[\eta(S^y)]\right) dN^y(s').$$

We can therefore conclude that, to a second-order, the expected old skill workers with young skill s is

$$\mathbb{E}\left[S^o \mid s\right] \approx \mathbb{E}\left[\mathrm{S}^o \mid s\right] + \frac{1}{2}\sum_l\sum_{l'}\left(T_l - 1\right)\left(T_{l'} - 1\right)\frac{\partial^2 \mathbb{E}[S^o \mid s]}{\partial T_l \partial T_{l'}} = \mathbb{E}\left[\mathrm{S}^o \mid s\right] + \vartheta^2 \mathbb{V}\mathrm{ar}[T]\Omega(s),$$

where

$$\Omega(s) \equiv \left(\eta(s) - \mathbb{E}[\eta(S^y)]\right)\int \gamma(s,s')\left(\eta(s') - \mathbb{E}[\eta(S^y)]\right) dN^y(s').$$

Accordingly, to a second order, the average skill of old workers is

$$\mathbb{E}\left[S^o\right] \approx \mathbb{E}\left[\mathrm{S}^o\right] + \vartheta^2 \mathbb{V}\mathrm{ar}[T]\int \Omega(s) dN^y(s).$$

Between-city wage inequality Let $\bar{W}_\ell^y \equiv \mathbb{E}[W_\ell^y]$ be the average wage in city ℓ. The between-city wage inequality is

$$\mathbb{V}\mathrm{ar}[\bar{W}_\ell^y] = \sum_\ell \left(\bar{W}_\ell^y - \bar{W}^y\right)^2 N_\ell^y,$$

for $\bar{W}^y \equiv \mathbb{E}[W^y]$ the aggregate average wage. As expected, there is no first order effect on the variance. The partial derivative of $\mathbb{V}\mathrm{ar}[\bar{W}_\ell^y]$ w.r.t. T_l reads

$$\partial_l \mathbb{V}\mathrm{ar}[\bar{W}_\ell^y] = 2\sum_\ell \left(\bar{W}_\ell^y - \bar{W}^y\right)\left(\partial_l \bar{W}_\ell^y - \partial_l \bar{W}^y\right) N_\ell^y + \sum_\ell \left(\bar{W}_\ell^y - \bar{W}^y\right)^2 \partial_l N_\ell^y,$$

which, when evaluated at $\boldsymbol{T} = \bar{T}$, simplifies to $\partial_l \mathbb{V}\mathrm{ar}[\bar{W}_\ell^y] = 0$. Hence, turning to the second order effect, we have that the second order partial derivative of $\mathbb{V}\mathrm{ar}[\bar{W}_\ell^y]$ w.r.t. $(T_l, T_{l'})$ evaluated at $\boldsymbol{T} = \bar{T}$ is

$$\partial^2_{ll'}\mathbb{V}\mathrm{ar}[\bar{W}_\ell^y] = \frac{2}{L}\sum_\ell \left(\partial_{l'}\bar{W}_\ell^y - \partial_{l'}\bar{W}^y\right)\left(\partial_l \bar{W}_\ell^y - \partial_l \bar{W}^y\right). \tag{43}$$

All that is required is therefore the marginal, first order effect of cities' TFP on the cities' average wages and the aggregate average wage. Using (34), the first marginal effect reads

$$\partial_l \bar{W}_\ell^y = \mathbb{1}\{\ell = l\}\mathbb{E}[S^y] + \vartheta I_{l\ell}\bar{T}\mathbb{C}\mathrm{ov}[S^y, \eta(S^y)].$$

Using (33), the second marginal effect is

$$\partial_l \bar{W}^y = \frac{1}{L}\mathbb{E}[S^y].$$

Combined, (43) rewrites

$$\partial^2_{ll'}\mathbb{V}\mathrm{ar}[\bar{W}_\ell^y] = \frac{2}{L}I_{ll'}\left(\mathbb{E}[S^y] + \vartheta\bar{T}\mathbb{C}\mathrm{ov}[S^y, \eta(S^y)]\right)^2,$$

where I have used $\sum_\ell I_{l'\ell}I_{l\ell} = I_{ll'}$. I therefore conclude that, to a second order, the between-city variance of wages is

$$\mathbb{V}\mathrm{ar}[\bar{W}_\ell^y] \approx \mathbb{V}\mathrm{ar}[T_\ell]\left(\mathbb{E}[S^y] + \vartheta\bar{T}\mathbb{V}\mathrm{ar}[S^y] + \vartheta^2\beta\bar{T}\mathbb{C}\mathrm{ov}[S^y, \Theta(S^y)]\right)^2$$

since the between-city variance of wages is nil when city TFPs are homogeneous.

A.6 Proof of Lemma 1

Suppose that the learning technology is either supermodular or submodular, $\gamma_{12}(s, s_p) > 0$ or $\gamma_{12}(s, s_p) < 0$ a.e. I then guess and verify that $\mathbb{C}\mathrm{ov}[\gamma_1(s, S^y), S^y] \geq -1$ implies $\eta'(s) \geq 0$. So guess that $\eta'(s) > 0$. I first sign Θ' under this guess. Note from (6) that Θ can be written $\Theta(s) = \mathbb{C}\mathrm{ov}[\gamma(s, S^y), \eta(S^y)]$. Differentiate Θ to obtain $\Theta'(s) = \mathbb{C}\mathrm{ov}[\gamma_1'(s, S), \eta(S)]$. The covariance of two increasing function is positive while the covariance of an increasing and a decreasing function is negative. Hence, since $\eta' > 0$, we have that $\gamma_{12} > 0$ implies $\Theta'(s) > 0$ while $\gamma_{12} < 0$ implies $\Theta'(s) < 0$. We can now verify the guess. Differentiate η and plug in the expression for η in Θ to find

$$\eta'(s) = 1 + \vartheta\beta \mathbb{C}\mathrm{ov}[\gamma_1(s, S^y), S^y] + (\vartheta\beta)^2 \, \mathbb{C}\mathrm{ov}[\gamma_1(s, S^y), \Theta(S^y)].$$

I distinguish between two cases. First, suppose that $\gamma_{12} > 0$. Then, $\mathbb{C}\mathrm{ov}[\gamma_1(s, S^y), S^y] > 0$. Furthermore, since $\Theta' > 0$, we also have $\mathbb{C}\mathrm{ov}[\gamma_1(s, S^y), \Theta(S^y)] > 0$. Hence, $\eta' > 0$. Second, suppose that $\sigma_{12} < 0$. Then, $\mathbb{C}\mathrm{ov}[\gamma_1(s, S^y), S^y] < 0$. Second, since $\Theta' < 0$ and the covariance of two decreasing functions is positive, we have $\mathbb{C}\mathrm{ov}[\gamma_1(s, S), \Theta(S)] > 0$. Hence, $\eta'(s) > 1 + \vartheta\beta\mathbb{C}\mathrm{ov}[\gamma_1(s, S), S] \geq 0$, where the second inequality is from the assumption on γ. This verifies our guess.

A.7 Proof of Proposition 2

[Redo.]

Proving Proposition 2 is then immediate when we know that η is an increasing function. Since the covariance of two increasing functions is positive, $\gamma(s, s_p)$ increasing in s_p implies $\Theta(s) > 0$. Conversely, since the covariance of a decreasing and a positive function is negative, $s_p \to \gamma(s, s_p)$ decreasing implies that $\Theta(s) < 0$. Similar arguments can be used to sign $\Theta'(s) = \mathbb{C}\mathrm{ov}[\sigma_1'(s, S), \eta(S)]$.

A.8 Proof of Proposition 4

The expression for Ω is derive in Section A.5. Define $\chi(s) \equiv \eta(s) - \mathbb{E}[\eta(S^y)]$. From the fundamental theorem of calculus,

$$\gamma(s, s') = \int_{\underline{s}}^{s} \int_{\underline{s}}^{s'} \gamma_{12}(x, y) dy dx + \gamma(s, \underline{s}) - \gamma(\underline{s}, \underline{s}) + \gamma(\underline{s}, s').$$

From the definition of χ, we know that $\int \gamma(s, \underline{s}) \chi(s) ds = \int \gamma(\underline{s}, s') \chi(s') ds' = \int \gamma(\underline{s}, \underline{s}) \chi(s) ds = 0$. Therefore, Ω can be rewritten

$$\Omega = \int \int \int_{\underline{s}}^{s} \int_{\underline{s}}^{s'} \gamma_{12}(x, y) \chi(s) \chi(s') dy dx dN^y(s') dN^y(s) = \int \int \gamma_{12}(x, y) \Psi(x) \Psi(y) dx dy,$$

where $\Psi(x) \equiv \int_x \chi(s) dN^y(s)$ and the second equality follows from Fubini's. Note that Ψ rewrites $\Psi(x) = (1 - N(x)) \left(\mathbb{E}[\eta(S^y) \mid S^y \geq x] - \mathbb{E}[\eta(S^y)] \right)$. Hence, $\Psi(x) > 0$ iff $\mathbb{E}[\eta(S^y) \mid S^y \geq x] > \mathbb{E}[\eta(S^y)]$. But Lemma 1 guarantees that η is an increasing function. It follows that $s \to \mathbb{E}[\eta(S^y) \mid S^y \geq s]$ is increasing, and $\mathbb{E}[\eta(S^y) \mid S^y \geq s] \geq \mathbb{E}[\eta(S^y) \mid S^y \geq \underline{s}] = \mathbb{E}[\eta(S^y)]$. Hence, $\Psi(x)\Psi(y) > 0$ and Ω preserves the sign of γ_{12}.

A.9 The Learning Consequences of Local Interactions across Skills

In a similar fashion as 10, the expected old skill of workers endowed with an initial skill s is, to a second order, given by (see Section A.5)

$$\mathbb{E}[S^o \mid s] \approx \mathbb{E}[\bar{S}^o \mid s] + \vartheta^2 \mathbb{V}\mathrm{ar}[T] \Omega(s),$$

where

$$\Omega(s) = (\eta(s) - \mathbb{E}[\eta(S^y)]) \int \gamma(s, s_p) \left(\eta(s_p) - \mathbb{E}[\eta(S^y)] \right) dN^y(s_p).$$

The term $\Omega(s)$ has a similar interpretation as Ω but at the skill level. If $\Omega(s) > 0$, the spatial segmentation of learning interactions boost the average learning of workers initially endowed with a skill s. The converse happen if $\Omega(s) < 0$. The following proposition characterizes $\Omega(s)$ as a function of the learning technology.

Proposition A.3 (Learning inequality)**.**
Define $s^\star$ as the skill with the average willingness to sort, $\eta(s^\star) \equiv \mathbb{E}[\eta(S^y)]$.

1. *If $\gamma(s,\cdot)$ is increasing for all s, then $\Omega(s)>0$ if and only if $s>s^\star$;*
2. *If $\gamma(s,\cdot)$ is decreasing for all s, then $\Omega(s)>0$ if and only if $s<s^\star$.*

Proof. I use similar notation as for the proof of Proposition 4. From the fundamental theorem of calculus, $\Omega(s)$ can be rewritten

$$\gamma(s,s')=\int_{\underline{s}}^{s'}\gamma_2(s,y)dydx+\gamma(s,\underline{s}).$$

From the definition of χ, we know that $\int\gamma(s,\underline{s})\chi(s)ds'=0$, and therefore $\Omega(s)$ reads

$$\Omega(s)=\chi(s)\int\int_{\underline{s}}^{s'}\gamma_2(s,y)dy\chi(s')dN^y(s')=\chi(s)\int\gamma_2(s,y)\Psi(y)dy.$$

The proof of Proposition 4 has shown that $\Psi(y)>0$. The sign of $\Omega(s)$ therefore depends on the product between the sign of $\chi(s)$ and that of γ_2. Let $s^\star$ be such that $\eta(s^\star)=\mathbb{E}[\eta(S^y)]$ and $\eta(s)>\mathbb{E}[\eta(S^y)]$ iff $s>s^\star$. Lemma 1 guarantees that such $s^\star$ exists. For $s>s^\star$, if $\gamma_2>0$ for all y, then $\Omega(s)>0$; if $\gamma_2<0$ for all y, then $\Omega(s)<0$. Opposite equality exists for $s<s^\star$. □

In particular, when every worker learns relatively more from skilled interactions, Proposition A.3 implies that the segmentation of learning interactions boost the learning of relatively skilled workers and dampens that of relatively low-skill individuals. That is, local interactions spur learning inequality across skills.

B Efficiency

B.1 *Planner's problem*

Suppose that the planner is utilitarian. The planner places unitary weights on skill and cities and discounts the future of future generations at rate β_S. Let $\mathbb{E}_\ell[V_t^{y\star}(s,\boldsymbol{\varepsilon})\mid s]$ denote the expected utility of young workers with skill s working in city ℓ at time t, where the expectation is taken across the unobserved idiosyncratic preferences $\boldsymbol{\varepsilon}$. Define $\mathbb{E}_\ell[V_t^{o\star}(s,\boldsymbol{\varepsilon})\mid s]$ for old workers. The planner's social welfare function,

$$\int\sum_\ell\mathbb{E}_\ell[V_0^{o\star}(s,\boldsymbol{\varepsilon})\mid s]n_{0\ell}^{o\star}(s)ds+\sum_{t>0}\beta_S^t\int\sum_\ell\mathbb{E}_\ell[V_t^{y\star}(s,\boldsymbol{\varepsilon})\mid s]n_{t\ell}^{y\star}(s)ds.$$

The expected utility of workers must be consistent with workers' location choices. By the property of extreme-value type one distribution, the expected utility of old workers working in city ℓ is identical to the *ex-ante* expected utility:

$$\begin{aligned}\mathbb{E}_\ell[V_t^{o\star}(s,\boldsymbol{\varepsilon})\mid s]&=\mathbb{E}[V_t^{o\star}(s,\boldsymbol{\varepsilon})\mid s]\\&=\frac{1}{\vartheta}\log\left(\sum_{\ell'}\mathrm{e}^{\vartheta c_{t\ell'}^{o\star}(s)}\right)\equiv\mathcal{V}_t^{o\star}(s).\end{aligned}$$

Similarly for the young,

$$\mathcal{V}_t^{y\star}(s)=\frac{1}{\vartheta}\log\left(\sum_\ell\mathrm{e}^{\vartheta\left(c_\ell^{y\star}(s)+\beta O_\ell^\star(s)\right)}\right),$$

where

$$O_\ell^\star(s)=\int\int\mathcal{V}^{o\star}[e\gamma(s,s_p)]dF(e)\pi_{t\ell}^{y\star}(s_p)ds_p.$$

Hence, the social welfare function simplifies to

$$\int\mathcal{V}_0^{o\star}(s)\sum_\ell n_{0\ell}^{o\star}(s)ds+\sum_{t>0}\beta_S^t\int\mathcal{V}_t^{y\star}(s)\sum_\ell n_{t\ell}^{y\star}(s)ds.$$

The problem of the planner is to maximize this objective function by choosing a sequence $\{\boldsymbol{c}_t^{y\star},\boldsymbol{c}_t^{o\star},\boldsymbol{n}_t^{y\star},\boldsymbol{n}_t^{o\star},n_t^\star\}_{t=0}^\infty$.

Next, I note that between-skill transfers are not pined down. Intuitively, workers' utility is linear in income. The idiosyncratic location preferences generate some concavity across cities but not across skills. Hence, any between-skill transfer can be sustained to the extent that these transfers are feasible. To see this, suppose that the planner considers two consumption plans at time t for young workers: $\boldsymbol{c}_t^y$ on the one hand, and $\tilde{\boldsymbol{c}}_t^y$ on the other hand. The two plans differ in that $\tilde{c}_{t\ell}^{y\star}(s) = c_{t\ell}^{y\star}(s) + q(s)$ for some function q. Since q is not city-specific, it does not affect the spatial allocation, and the learning value of cities is the same in the two plans. Similarly, total output is constant. Hence, these two plans are jointly feasible if $\int q(s)dN^y(s)ds = 0$. Meanwhile, the difference in the aggregate utility of cohort t between the two consumption plans is

$$\frac{1}{\vartheta}\int\left(\log\left(\sum_\ell \mathrm{e}^{\vartheta\left(c_\ell^{y\star}(s)+\beta O_\ell^\star(s)\right)}\right) - \log\left(\sum_\ell \mathrm{e}^{\vartheta\left(\tilde{c}_\ell^{y\star}(s)+\beta O_\ell^\star(s)\right)}\right)\right) n^y(s)ds = \int q(s)n^y(s)ds = 0.$$

That is, any function q yields the same aggregate welfare, and between-skill transfers are not determinate. I therefore suppose that between-skill transfers are not allowed. This restriction is without loss of generality for the optimal spatial allocation since between-skill transfers do not distort workers' incentives to live in particular cities. I also suppose that between-age transfers are not allowed to focus on the learning externalities.

The planner thus faces the following conditions:

1. Young and old consumption feasibility, $\kappa_t^y(s)$ and $\kappa_t^o(s)$ respectively:

$$\sum_\ell c_{t\ell}^{y\star}(s)n_{t\ell}^{y\star}(s) \le \sum_\ell T_\ell s n_{t\ell}^{y\star}(s), \quad \text{and} \quad \sum_\ell c_{t\ell}^{o\star}(s)n_{t\ell}^{o\star}(s) \le \sum_\ell T_\ell s n_{t\ell}^{o\star}(s), \quad \forall(s,t)$$

2. Young labor mobility conditional on workers' unobserved idiosyncratic preferences, $\lambda_{t\ell}^y(s)$:

$$n_{t\ell}^{y\star}(s) \le n^y(s)\left(\frac{\mathrm{e}^{\vartheta\left(c_{t\ell}^{y\star}(s)+\beta O_{t\ell}^\star(s)\right)}}{\sum_{\ell'}\mathrm{e}^{\vartheta\left(c_{t\ell'}^{y\star}(s)+\beta O_{t\ell'}^\star(s)\right)}}\right), \quad \forall(s,t,\ell).$$

3. Old labor mobility conditional on workers' (unobserved) idiosyncratic preferences, $\lambda_{t\ell}^o(s)$:

$$n_{t\ell}^{o\star}(s) \le n_t^{o\star}(s)\left(\frac{\mathrm{e}^{\vartheta c_{t\ell}^{o\star}(s)}}{\sum_{\ell'}\mathrm{e}^{\vartheta c_{t\ell'}^{o\star}(s)}}\right), \quad \forall(s,t,\ell).$$

4. The aggregate old skill law of motion, $\nu_t(s)$:

$$n_{t+1}^{o\star}(s_o) = \sum_\ell\int\int\frac{1}{\gamma(s_y,s_p)}f\left(\frac{s_o}{\gamma(s_y,s_p)}\right)n_{t\ell}^{y\star}(s_y)\pi_{t\ell}^{y\star}(s_p)ds_p ds_y, \quad \forall(s,t).$$

Given these preliminary remarks, the Lagrangian of the planner's problem reads

$$\begin{aligned}
\mathcal{L} = \frac{1}{\vartheta}\int\log\left(\sum_\ell \mathrm{e}^{\vartheta c_{0\ell}^{o\star}(s)}\right)\sum_\ell n_{0\ell}^{o\star}(s)ds + \frac{1}{\vartheta}\sum_{t>0}\beta_S^t\int\log\left(\sum_\ell \mathrm{e}^{\vartheta\left(c_{t\ell}^{y\star}(s)+\beta O_{t\ell}^\star(s)\right)}\right)\sum_\ell n_{t\ell}^{y\star}(s)ds - \\
\sum_t\sum_\ell\int\lambda_{t\ell}^y(s)\left(n_{t\ell}^{y\star}(s) - n^y(s)\left(\frac{\mathrm{e}^{\vartheta\left(c_{t\ell}^{y\star}(s)+\beta O_{t\ell}^\star(s)\right)}}{\sum_{\ell'}\mathrm{e}^{\vartheta\left(c_{t\ell'}^{y\star}(s)+\beta O_{t\ell'}^\star(s)\right)}}\right)\right)ds - \\
\sum_t\sum_\ell\int\lambda_{t\ell}^o(s)\left(n_{t\ell}^{o\star}(s) - n_t^{o\star}(s)\left(\frac{\mathrm{e}^{\vartheta c_{t\ell}^{o\star}(s)}}{\sum_{\ell'}\mathrm{e}^{\vartheta c_{t\ell'}^{o\star}(s)}}\right)\right)ds - \\
\sum_t\int\nu_t(s_o)\left(n_{t+1}^{o\star}(s_o) - \sum_\ell\int\int\frac{1}{\gamma(s_y,s_p)}f\left(\frac{s_o}{\gamma(s_y,s_p)}\right)n_{t\ell}^{y\star}(s_y)\pi_{t\ell}^{y\star}(s_p)ds_p ds_y\right)ds_o - \\
\sum_t\int\kappa_t^y(s)\sum_\ell\left(c_{t\ell}^{y\star}(s) - T_\ell s\right)n_{t\ell}^{y\star}(s)ds + \sum_t\int\kappa_t^o(s)\sum_\ell\left(c_{t\ell}^{o\star}(s) - T_\ell s\right)n_{t\ell}^{o\star}(s)ds,
\end{aligned}$$

where I have kept implicit that

$$O_{t\ell}^\star(s) = \frac{1}{\vartheta}\int\int\log\left(\sum_{\ell'}\mathrm{e}^{\vartheta c_{t+1\ell'}^{o\star}[e\gamma(s,s_p)]}\right)dF(e)\pi_{t\ell}^{y\star}(s_p)ds_p.$$

B.2 Non-linear optimal allocation

Solving this maximization problem consists of finding the function $s \to c^{y\star}_{t\ell}(s)$, $s \to c^{o\star}_{t\ell}(s)$, $s \to n^{y\star}_{t\ell}(s)$, $s \to n^{o\star}_{t\ell}(s)$, and $s \to n^{o\star}_{t}(s)$, for each $\ell \in \{1, 2, \dots, L\}$ and $t \in \{0, 1, 2, \dots\}$. This is an infinite-dimensional maximization problem that is solved using functional derivatives. I lay down the complete argument for the optimality condition regarding the initial old allocation, $c^{o\star}_{0\ell}$ and $n^{o\star}_{0\ell}(s)$. I then present more succinct derivations for the remaining optimality conditions. For the remaining of the derivation, I drop the star superscript for notation simplicity.

Initial old Suppose the planner contemplates an alternative consumption plan given by $\tilde{c}^o_{0l}(s) = c^o_{0l}(s) + \varepsilon\eta(s)$ for some l and some arbitrary function $\eta(s)$, while $\tilde{c}^o_{0\ell}(s) = c^o_{0\ell}(s)$ for $\ell \neq l$. Optimality requires that $d\mathcal{L}/d\varepsilon = 0$ at $\varepsilon = 0$, or

$$\int \eta(s) n^o_{0l}(s) ds + \vartheta \int \eta(s) n^o_{0l}(s) \left(\lambda^o_{0l}(s) - \sum_{\ell} \lambda^o_{0\ell}(s) \frac{n^o_{0\ell}(s)}{n^o_0(s)} \right) ds - \int \kappa^o_0(s)\eta(s) n^o_{0l}(s) ds = 0,$$

where I have already imposed that $\sum_\ell n^o_{0\ell}(s) = n^o_0(s)$. Optimality requires that this hold for any function $\eta(s)$. In particular, letting $\eta(s) = \delta_x(s) = \delta(s-x)$, where δ is the Dirac delta function, the above optimality condition simplifies to

$$1 + \vartheta \left(\lambda^o_{0l}(x) - \sum_{\ell} \lambda^o_{0\ell}(x) \frac{n^o_{0\ell}(x)}{n^o_0(x)} \right) = \kappa^o_0(x),$$

where I have also used that the idiosyncratic location preferences guarantee an interior allocation to simplify the $n^{o\star}_{0l}(x)$. In the above expression, the only l specific term is the KKT multiplier $\lambda^o_{0l}(x)$. Hence, $\lambda^o_{0l}(x) = \lambda^o_0(x)$ for all l, and we can conclude that $\kappa^o_0(x) = 1$.

Proceeding in a similar fashion for the optimality condition regarding $n^o_{0l}(s)$ and using $\kappa^o_0(x) = 1$, we have

$$\mathcal{V}^o_0(x) - (c^o_{0l}(x) - T_l x) = \lambda^o_0(x).$$

Multiplying by $n^o_{0l}(x)$, summing across l and using the age-skill specific feasibility condition yields $\lambda^o_0(x) = \mathcal{V}^o_0(x)$ and therefore $c^o_{0l}(x) = T_l x$.

Old workers I now derive the optimality conditions for $(\boldsymbol{n}^o_t, \boldsymbol{n}^o_{\ell t}, \boldsymbol{c}^o_{\ell t})$ for $t \geq 1$. Regarding $\boldsymbol{c}^o_{lt}$, we have

$$\begin{aligned}\int \kappa^o_t(s) n^o_{tl}(s)\eta(s) ds = \beta_S^{t-1}\beta \int \sum_{\ell} n^y_{t-1\ell}(s) \delta_{c^o_{tl}} O_{t-1\ell}(s) ds +& \\ \vartheta\beta \sum_{\ell} \int \lambda^y_{t-1\ell}(s) n^y_{t\ell}(s) \left(\delta_{c^o_{tl}} O_{t-1\ell}(s) - \sum_{\ell'} \frac{n^y_{t-1\ell'}(s)}{n^y(s)} \delta_{c^o_{tl}} O_{t-1\ell'}(s) \right) ds +& \\ \vartheta \int n^o_{tl}(s)\eta(s) \left(\lambda^o_{tl}(s) - \sum_{\ell} \lambda^o_{t\ell}(s) \frac{n^o_{t\ell}(s)}{n^o_t(s)} \right) ds,&\end{aligned}$$

for any η function, where $\delta_x F$ denote the functional derivative of F w.r.t. x. I anticipate that, at the optimum, it must be that $\lambda^y_{t-1\ell}(s) \equiv \lambda^y_{t-1}(s)$ for all ℓ and t. Hence, the second line is equal to zero. From the expression for $O_{t\ell}$, we further know that

$$\delta_{c^o_{tl}} O_{t-1\ell}(s) = \int \int \eta[e\gamma(s, s_p)] \frac{n^o_{tl}[e\gamma(s, s_p)]}{n^o_t[e\gamma(s, s_p)]} dF(e) \pi^y_{t-1\ell}(s_p) ds_p.$$

Hence, the optimality condition turns into

$$\begin{aligned}\int \kappa^o_t(s) n^o_{tl}(s)\eta(s) ds = \beta_S^{t-1}\beta \int \eta(q) \frac{n^o_{tl}(q)}{n^o_t(q)} \sum_{\ell} \int \int \frac{1}{\gamma(s, s_p)} f\left(\frac{q}{\gamma(s, s_p)} \right) n^y_{t-1\ell}(s) \pi^y_{t-1\ell}(s_p) ds_p ds dq +& \\ \vartheta \int n^o_{tl}(s)\eta(s) \left(\lambda^o_{tl}(s) - \sum_{\ell} \lambda^o_{t\ell}(s) \frac{n^o_{t\ell}(s)}{n^o_t(s)} \right) ds,&\end{aligned}$$

where I have used the change of variable $e\gamma(s, s_p) \to q$ to rewrite the integral on the first line. Using the law of motion for the aggregate old skill distribution, the two inner most integrals on the first line cancel out with $n^o_t(q)$. Setting

$\eta(s) = \delta_x(s)$, we then obtain

$$\kappa_t^o(x) = \beta_S^{t-1}\beta + \vartheta\left(\lambda_{tl}^o(x) - \sum_\ell \lambda_{t\ell}^o(x)\frac{n_{t\ell}^o(x)}{n_t^o(x)}\right),$$

where both sides have been divided by $n_{tl}^o(x)$. As for the initial old consumption allocation, there is no l-specific term in the above expression, and $\lambda_{tl}^o(x) = \lambda_t^o(x)$. Hence, $\kappa_t^o(x) = \beta_S^{t-1}\beta \equiv \kappa_t^o$ for $t \geq 1$. Setting $\beta_S = \beta$, then $\kappa_t^o = \beta^t$ for $t \geq 0$.

Turning to the optimality condition w.r.t. n_{lt}^o, note that $O_{t\ell}(s)$ is independent of n_{tl}^o. Hence, after using $\lambda_{tl}^o(x) = \lambda_t^o(x)$ and $\kappa_t^o = \beta_S^{t-1}\beta$, the optimality condition reads

$$\lambda_t^o(x) = \beta_S^{t-1}\beta\left(T_l s - c_{tl}^o(x)\right).$$

Multiplying by $n_{tl}^o(x)$, summing across l and using the feasibility condition, this implies $\lambda_t^o(x) = 0$ for all $t \geq 1$, and therefore $c_{tl}^o(x) = T_l s$.[97]

Finally, there remains the optimality condition w.r.t. n_t^o. There, we have

$$\sum_\ell \int \lambda_{\ell t}^o(s)\eta(s)\frac{n_{\ell t}^o(s)}{n_t^o(s)}ds - \int \nu_{t-1}(s)\eta(s)ds = \int \lambda_t^o(s)\eta(s)ds - \int \nu_{t-1}(s)\eta(s)ds = 0,$$

for $t \geq 1$, where the first equality follows from $\lambda_{\ell t}^o(x) = \lambda_t^o(x)$. Setting $\eta = \delta_x(s)$, we obtain the optimality condition $\nu_{t-1}(x) = \lambda_t^o(x) = 0$.

Young workers I finally derive the optimality conditions for $(\boldsymbol{n}_t^y, \boldsymbol{n}_{\ell t}^y, \boldsymbol{c}_{\ell t}^y)$ for $t \geq 0$. For consumption, note that the learning value of cities is independent of c_{tl}^y. Hence, the optimality condition w.r.t. c_{tl}^y is

$$\beta_S^t + \vartheta\left(\lambda_{tl}^y(x) - \sum_\ell \lambda_{t\ell}^y(x)\frac{n_{t\ell}^y(x)}{n^y(x)}\right) = \kappa_t^y(x).$$

Hence, here as well, this implies $\lambda_{lt}^y(x) \equiv \lambda_t^y(x)$ and $\kappa_t^y(x) = \beta_S^t \equiv \kappa_t^y$. When $\beta_S = \beta$, then $\kappa_t^y = \kappa_t^o$.

I conclude with the optimality condition for n_{tl}^y. There, and crucially, n_{tl}^y affects O_{tl}. The optimality condition is

$$\begin{aligned}
0 = \beta_S^t\beta \int \sum_\ell n_{t\ell}^y(s)\delta_{n_{tl}^y} O_{t\ell}(s)ds + \beta_S^t \int \mathcal{V}_t^y(s)\eta(s)ds - \\
\int \left(\lambda_{tl}^y(s)\eta(s) - \vartheta\beta\sum_\ell \lambda_{t\ell}^y(s)n_{t\ell}^y(s)\left[\delta_{n_{tl}^y}O_{t\ell}(s) - \sum_{\ell'}\frac{n_{t\ell'}^y(s)}{n^y(s)}\delta_{n_{tl}^y}O_{t\ell'}(s)\right]\right)ds + \\
\int \nu_t(s_o)\int\int \frac{1}{\gamma(s_y,s_p)}f\left(\frac{s_o}{\gamma(s_y,s_p)}\right)\left(\eta(s_y)\pi_{tl}^y(s_p) + \pi_{tl}^y(s_y)\left[\eta(s_p) - \pi_{tl}^y(s_p)\int\eta(\sigma)d\sigma\right]\right)ds_p ds_y ds_o - \\
\int \kappa_t^y(s)\left[c_{tl}^y(s) - sT_l\right]\eta(s)ds.
\end{aligned}$$

where I have already imposed on the first line that $\sum_\ell \phi_\ell(x)\frac{n_{t\ell}^y(x)}{n^y(x)} = 1$ by setting $\phi_\ell(s) = 1$. Using the spatial equalization of λ_{lt}^y, the second line boils down to $\int \lambda_t^y(s)\eta(s)ds$. Using $\nu_t^o(s) = 0$, the third line vanishes. Finally, from the expression for $O_{tl}(s)$, we have

$$\delta_{n_{tl}^y}O_{t\ell}(s) = \frac{1}{N_{tl}^y}\mathbb{1}\{\ell = l\}\left(\int\int \mathcal{V}^o[e\gamma(s,s_p)]dF(e)\eta(s_p)ds_p - O_{tl}(s)\int\eta(\sigma)d\sigma\right).$$

Using $\eta(s) = \delta_x(s)$ and $\int \delta_x(s)ds = 1$, this simplifies to

$$\partial_{n_{tl}(x)^y}O_{t\ell}(s) = \frac{1}{N_{tl}^y}\mathbb{1}\{\ell = l\}\left(\int \mathcal{V}^o[e\gamma(s,x)]dF(e) - O_{tl}(s)\right).$$

[97]The same would hold if we were to allow for transfers within age across skills. Then, feasibility would require $\int \lambda_t^o(x)n_t^o(x)dx = \beta_S^{t-1}\beta\sum_l \int n_{tl}^o(x)\left(T_l s - c_{tl}^o(x)\right)dx = 0$. But $\lambda_t^o(x) \geq 0$ since it is a KKT multiplier, and therefore $\lambda_t^o(x) = 0$ for all x. Hence, on the contrary of the young allocation, the old allocation is uniquely pin down even with across-skill transfers. Why? Between-skill transfers for old workers affect the incentives to learn, and thus the spatial allocation of workers.

When the planner increases the measure of workers x in ℓ, it has two effects on the opportunities of the city. On the one hand, it increases the likelihood of interactions with x. The gains from these extra interactions are captured by the first term. On the other hand, it decreases the likelihood of interactions with every other workers in the city. When there are a continuum of skills, this congestion cost is captured by the average value of interactions in l, which is the second term. Combining this expression in the optimality condition, and setting $\kappa_t^y = \beta_S^t$, we obtain

$$c_{tl}^y(x) = xT_l + \beta \int \left(\int \mathcal{V}^o[e\gamma(s,x)]dF(e) - O_{tl}(s) \right) \pi_{tl}^y(s)ds + \mathcal{V}_t^y(x) - \lambda_t^y(x)/\beta_S^t.$$

In steady state, it must be that $\lambda_t^y(s)/\beta_S^t \equiv \lambda^y(s)$ is constant over time. Then, the optimal allocation is given by

$$c_l^y(x) = xT_l + \tau_l(x) + \mathcal{V}^y(x) - \lambda^y(x),$$

for

$$\tau_l(x) \equiv \beta \int \int \left(\mathcal{V}^o[e\gamma(s,x)] - \mathbb{E}[\mathcal{V}^o[e\gamma(s,S_\ell^y)]]\right) dF(e)\pi_l^y(s)ds$$

after substituting the expression for $O_l(s)$. This expression corresponds to (13) in the main text. Using the feasibility condition, it must be that $\mathcal{V}^y(x) - \lambda^y(x) = -\sum_l n_l^y(x)/n^y(x)\tau_l(x)$, and we conclude that[98]

$$c_l^y(x) = xT_l + \tau_l(x) - \sum_\ell \left(\frac{n_\ell^y(x)}{n^y(x)}\right)\tau_\ell(x). \tag{44}$$

In the competitive equilibrium, workers' consumption equate their wage, i.e. $c_l^y(x) = xT_l$. (44) thus yields (12) in the main text. Alternatively, (44) can be rewritten

$$c_l^y(x) = w_l(x) + t_l(x),$$

where $t_l(x) \equiv \tau_l(x) - \sum_\ell \left(\frac{n_\ell^y(x)}{n^y(x)}\right)\tau_\ell(x)$. Hence, the optimal allocation can be decentralized through the transfers $t_l(x)$.

B.3 Linearizing the social optimum

The social optimum is given by the spatial consumption allocation (44) and the young spatial allocation

$$n_\ell^{y\star}(s) = n^y(s)\left(\frac{\mathrm{e}^{\vartheta\left(c_\ell^{y\star}(s)+\beta O_\ell^\star(s)\right)}}{\sum_{\ell'}\mathrm{e}^{\vartheta\left(c_{\ell'}^{y\star}(s)+\beta O_{\ell'}^\star(s)\right)}}\right). \tag{45}$$

Solving for the perturbed equilibrium around $\boldsymbol{T} \approx T$ requires to solve for the joint approximation of (44) and (45). I drop the star superscript for notational simplicity. When $\boldsymbol{T} = T$, (44) implies that $c_l^y(x) = x\bar{T}$ and $n_l^y(x) = n^y(x)/L$, as in the decentralized equilibrium.

Spatial allocation From (45), we have

$$\left.\frac{\partial n_\ell^y(s)}{\partial T_l}\right|_{\boldsymbol{T}=T} = \vartheta\left(\frac{n^y(s)}{L}\right)\left\{\frac{\partial c_\ell^y(s)}{\partial T_l} + \beta\frac{\partial O_\ell(s)}{\partial T_l} - \frac{1}{L}\sum_{\ell'}\left(\frac{\partial c_{\ell'}^y(s)}{\partial T_l} + \beta\frac{\partial O_{\ell'}(s)}{\partial T_l}\right)\right\}\Bigg|_{\boldsymbol{T}=1}.$$

Following similar steps as in Section A.4.1, the change in the city option value is

$$\frac{\partial O_\ell(s)}{\partial T_l} - \frac{1}{L}\sum_{\ell'}\frac{\partial O_{\ell'}(s)}{\partial T_l} = L\int \frac{\partial n_\ell^y(s_p)}{\partial T_l}\left(\gamma(s,s_p) - \mathbb{E}[\gamma(s,S^y)]\right)ds_p.$$

Consumption allocation To continue, we need to derive the consumption perturbation. From (44), the perturbed transfers read

$$\partial_l\tau_\ell(x) = \beta\int\left(\partial_l\pi_\ell^y(s) + \frac{n^y(s)}{L}\right)\left(\gamma(s,x) - \mathbb{E}[\gamma(s,S^y)]\right)ds - \beta\int \mathbb{E}[\gamma(S^y,s_p)]\partial_l\pi_\ell^y(s_p)ds_p.$$

[98] If we were to allow for transfers across skill within the young, then the optimal allocation would be $c_l^y(x) = xT_l + \tau_l(x) + t(x)$, for any function $t(x)$ that satisfies $\int t(x)n^y(x)dx = -\int\sum_l n_l^y(x)\tau_l(x)dx$.

Hence,

$$\partial_l\left(\tau_\ell(x)-t(x)\right)=\beta\int\partial_l\pi_\ell^y(s)\left(\gamma(s,x)-\mathbb{E}[\gamma(s,S^y)]\right)ds-\beta\int\mathbb{E}[\gamma(S^y,s_p)]\partial_l\pi_\ell^y(s_p)ds_p.$$

where I have used $\tau_\ell(s)=0$ when $\boldsymbol{T}=T$ and $\sum_\ell\partial_l\pi_\ell(s)=0$. It follows that the perturbed consumption allocation is

$$\partial_l c_\ell^y(x)=x\mathbb{1}\{l=\ell\}+\beta\int\partial_l\pi_\ell^y(s)\left(\gamma(s,x)-\mathbb{E}[\gamma(s,S^y)]\right)ds-\beta\int\mathbb{E}[\gamma(S^y,s_p)]\partial_l\pi_\ell^y(s_p)ds_p. \tag{46}$$

Using $\sum_\ell\partial_l c_\ell^y(x)=x$ and the expression for $\partial_l\pi_\ell^y(s)$ returns

$$\begin{aligned}\partial_l c_\ell^y(x)-\frac{1}{L}\sum_\ell\partial_l c_\ell^y(x)=xI_{\ell l}+\beta L\int\left(\partial_l n_\ell^y(s)-n(s)\int\partial_l n_\ell^y(q)dq\right)\left(\gamma(s,x)-\mathbb{E}[\gamma(s,S^y)]\right)ds-\\ \beta L\int\left(\partial_l n_\ell^y(s_p)-n(s_p)\int\partial_l n_\ell^y(q)dq\right)\mathbb{E}[\gamma(S^y,s_p)]ds_p.\end{aligned}$$

Optimal integral equation Combining the expression for $O_\ell(s)$ and $c_\ell(s)$, we obtain

$$\begin{aligned}\partial_l\partial n_\ell^y(x)=\vartheta n^y(x)\Bigg\{&\frac{xI_{\ell l}}{L}+\\ &\beta\int\left(\partial_l n_\ell^y(s)-n(s)\int\partial_l n_\ell^y(q)dq\right)\left(\gamma(s,x)-\mathbb{E}[\gamma(s,S^y)]\right)ds-\\ &\beta\int\left(\partial_l n_\ell^y(s_p)-n(s_p)\int\partial_l n_\ell^y(q)dq\right)\mathbb{E}[\gamma(S^y,s_p)]ds_p+\\ &\beta\int\partial_l n_\ell^y(s_p)\left(\gamma(x,s_p)-\mathbb{E}[\gamma(x,S^y)]\right)ds_p\Bigg\}.\end{aligned}$$

Guessing and verifying the form of the solution,

$$\partial_l n^y(s)=\vartheta n^y(s)\left(\frac{I_{\ell l}}{L}\right)\eta^\star(s),$$

we end up with the optimal integral equation

$$\begin{aligned}\eta^\star(x)=x+\vartheta\beta\int\left(\{\gamma(x,s)-\mathbb{E}[\gamma(x,S^y)]\}-\{\mathbb{E}[\gamma(S^y,s)]-\mathbb{E}[\gamma(S^y,S^y)]\}\right)\eta^\star(s)n^y(s)ds+\\ \vartheta\beta\int\left(\{\gamma(s,x)-\mathbb{E}[\gamma(S^y,x)]\}-\{\mathbb{E}[\gamma(s,S^y)]-\mathbb{E}[\gamma(S^y,S^y)]\}\right)\eta^\star(s)n^y(s)ds.\end{aligned} \tag{47}$$

Finally, using (46) and the fact that $\partial_l\pi_\ell^y(s)=\vartheta n^y(s)I_{l\ell}\left(\eta^\star(s)-\mathbb{E}[\eta^\star(S^y)]\right)$, the skill-by-city transfers $t_l(x)$ can be decomposed into a city-specific tax, T, and additional skill-by-city subsidies, $d_\ell(x)$. Specifically,

$$t_l^y(x)\approx T+d_l^y(x), \tag{48}$$

where

$$T_\ell=-\vartheta\beta\left(T_\ell-\bar{T}\right)\int\mathbb{E}[\gamma(S^y,s_p)]\left(\eta^\star(s_p)-\mathbb{E}[\eta^\star(S^y)]\right)dN^y(s_p)$$

and

$$d_\ell(x)=\vartheta\beta\left(T_\ell-\bar{T}\right)\int\left(\eta^\star(s)-\mathbb{E}[\eta^\star(S^y)]\right)\left(\gamma(s,x)-\mathbb{E}[\gamma(s,S^y)]\right)dN^y(s).$$

B.4 Proof of Proposition 5

Using covariance, (47) rewrites

$$\begin{aligned}\eta^\star(x)=x+\vartheta\beta\mathrm{Cov}[\gamma(x,S^y),\eta^\star(S^y)]+\vartheta\beta\mathrm{Cov}[\gamma(S^y,x),\eta^\star(S^y)]-\\ \vartheta\beta\mathrm{Cov}[\mathbb{E}[\gamma(s,S^y)\mid s],\eta^\star(s)]-\vartheta\beta\mathrm{Cov}[\mathbb{E}[\gamma(S^y,s)\mid s],\eta^\star(s)].\end{aligned}$$

The next lemma provides condition to ensure positive SPAM in the social optimum.

Lemma B.1 (Positive sorting)**.**
Suppose that γ *is either (weakly) supermodular or submodular. If*

$$\vartheta\beta\left(\mathbb{C}ov[\gamma_1(x,S^y),S^y]+\mathbb{C}ov[\gamma_2(S^y,x),S^y]\right)>-1,$$

then $\eta'^{\star}(x)>0$.

Proof. For notational simplicity, define $\tilde{\Theta}^{\star}(s)\equiv\mathbb{C}\mathrm{ov}[\gamma(s,S^y),\eta^{\star}(S^y)]$ and $\tilde{\Upsilon}^{\star}(s)\equiv\mathbb{C}\mathrm{ov}[\gamma(S^y,x),\eta^{\star}(S^y)]$, so that the expression for $\eta^{\star}$ reads

$$\eta^{\star}(x)=x+\vartheta\beta\tilde{\Theta}^{\star}(x)+\vartheta\beta\tilde{\Upsilon}^{\star}(x)-\vartheta\beta\mathbb{C}\mathrm{ov}[\mathbb{E}[\gamma(s,S^y)\mid s],\eta^{\star}(s)]-\vartheta\beta\mathbb{C}\mathrm{ov}[\mathbb{E}[\gamma(S^y,s)\mid s],\eta^{\star}(s)]$$

Differentiate w.r.t. x and use the expression for $\tilde{\Theta}$ and $\tilde{\Upsilon}$ to obtain

$$\eta'^{\star}(x)=1+\vartheta\beta\mathbb{C}\mathrm{ov}[\gamma_1(x,S^y),\eta^{\star}(S^y)]+\vartheta\beta\mathbb{C}\mathrm{ov}[\gamma_2(S^y,x),\eta^{\star}(S^y)].$$

Plug the expressions for $\eta^{\star}(x)$ back in the above,

$$\begin{aligned}\eta'^{\star}(x)=1&+\vartheta\beta\mathbb{C}\mathrm{ov}[\gamma_1(x,S^y),S^y]+(\vartheta\beta)^2\mathbb{C}\mathrm{ov}[\gamma_1(x,S^y),\tilde{\Theta}^{\star}(S^y)]+(\vartheta\beta)^2\mathbb{C}\mathrm{ov}[\gamma_1(x,S^y),\tilde{\Upsilon}^{\star}(S^y)]+\\&\vartheta\beta\mathbb{C}\mathrm{ov}[\gamma_2(S^y,x),S^y]+(\vartheta\beta)^2\mathbb{C}\mathrm{ov}[\gamma_2(S^y,x),\tilde{\Theta}^{\star}(S^y)]+(\vartheta\beta)^2\mathbb{C}\mathrm{ov}[\gamma_2(S^y,x),\tilde{\Upsilon}^{\star}(S^y)].\end{aligned}$$

Now, guess that η is increasing. From the expression for $\tilde{\Theta}$ and $\tilde{\Upsilon}$, if γ is supermodular (submodular), then $\tilde{\Theta}$ and $\tilde{\Upsilon}$ are increasing. Hence, if γ is supermodular, then all the covariances in the expression above are positive, and $\eta'^{\star}(x>1>0$, which verifies the guess. Suppose now that γ is submodular. Then, the covariances between γ_1, γ_2, $\tilde{\Theta}$ and $\tilde{\Upsilon}$ are positive since the covariance between two decreasing function is positive. Hence,

$$\eta'^{\star}(x)>1+\vartheta\beta\mathbb{C}\mathrm{ov}[\gamma_1(x,S^y),S^y]+\vartheta\beta\mathbb{C}\mathrm{ov}[\gamma_2(S^y,x),S^y].$$

It follows that $\vartheta\beta\left(\mathbb{C}\mathrm{ov}[\gamma_1(x,S^y),S^y]+\mathbb{C}\mathrm{ov}[\gamma_2(S^y,x),S^y]\right)>-1$ implies $\eta'^{\star}(x)>0$. □

This lemma is sufficient to prove the two elements of Proposition 5.

Proposition 5.1 The optimal spatial allocation solves

$$n_\ell^{y\star}(s)\approx\frac{n^y(s)}{L}+\vartheta\left(\frac{n^y(s)}{L}\right)\mathcal{T}_\ell\eta^{\star}(s).$$

Note that $\mathbb{E}[\eta^{\star}(S^y)]=\int\eta^{\star}(x)n^y(x)dx=\mathbb{E}[S^y]$. Hence, the optimal city sizes are

$$N_\ell^{\star}\approx\frac{1}{L}+\frac{1}{L}\vartheta\mathcal{T}_\ell\int\eta^{\star}(s)n^y(s)ds=\frac{1}{L}+\frac{1}{L}\vartheta\mathcal{T}_\ell\mathbb{E}[S^y].$$

In the decentralized equilibrium, city size was given by

$$N_\ell\approx\frac{1}{L}+\vartheta\frac{1}{L}\mathcal{T}_\ell\mathbb{E}[S^y]+\vartheta^2\beta\frac{1}{L}\mathcal{T}_\ell\mathbb{E}[\Theta(S^y)].$$

Hence, it is immediate to conclude that, for productive cities, $N_\ell>N_\ell^{\star}\iff\mathbb{E}[\Theta(S^y)]>0$ (and the converse holds for unproductive cities). But, we also know from Proposition 2 that $s_p\to\gamma(s,s_p)$ increasing implies $\Theta(s)>0$. Hence, if $s_p\to\gamma(s,s_p)$, there is too much agglomeration in productive cities. Conversely, if $s_p\to\gamma(s,s_p)$ is decreasing, there is too little agglomeration in productive cities.

Proposition 5.2 The optimal skill density is

$$\pi_\ell^{y\star}(s)\approx n^y(s)+\vartheta n^y(s)\mathcal{T}_\ell\left(\eta^{\star}(s)-\mathbb{E}[S^y]\right).$$

In particular, the new average skill in city ℓ reads

$$\mathbb{E}_\ell[S^{y\star}]\approx\mathbb{E}[S^y]+\vartheta\mathcal{T}_\ell\mathbb{C}\mathrm{ov}[\eta^{\star}(S^y),S^y]$$

In the decentralized equilibrium, we had

$$\mathbb{E}_\ell[S^y] \approx \mathbb{E}[S^y] + \vartheta \mathcal{T}_\ell \mathbb{C}\text{ov}[\eta(S^y), S^y].$$

Hence, for productive cities, $\mathbb{E}_\ell[S^y] > \mathbb{E}_\ell[S^{y\star}] \iff \mathbb{C}\text{ov}[\eta(S^y), S^y] > \mathbb{C}\text{ov}[\eta^\star(S^y), S^y]$. A sufficient condition for $\mathbb{C}\text{ov}[\eta(S^y), S^y] > \mathbb{C}\text{ov}[\eta^\star(S^y), S^y]$ is $\eta' > \eta'^\star$ a.e.[99] I now show that γ supermodular (submodular) implies $\eta'^\star > \eta'$ ($\eta'^\star < \eta'$).

Differentiate the expression for $\eta^\star$ w.r.t. x and use covariances to simplify the integrals,

$$\eta'^\star(x) = 1 + \vartheta\beta \mathbb{C}\text{ov}[\gamma_1(x, S^y), \eta^\star(S^y)] + \vartheta\beta \mathbb{C}\text{ov}[\gamma_2(S^y, x), \eta^\star(S^y)].$$

Taking a similar derivative in the decentralized equilbrium returns

$$\eta'(x) = 1 + \vartheta\beta \mathbb{C}\text{ov}[\gamma_1(x, S^y), \eta(S^y)].$$

The two functions η and $\eta^\star$ are the solution to integrals equations. To prove the propositions, I therefore guess and verify the statement. First, suppose that γ is supermodular, and guess that $\eta'^\star > \eta'$ a.e. Since γ is supermodular, $s \to \gamma_2(s, x)$ is an increasing function, and $\mathbb{C}\text{ov}[\gamma_2(S^y, x), \eta^\star(S^y)] > 0$ from Lemma B.1. Hence,

$$\eta'^\star(x) > 1 + \vartheta\beta \mathbb{C}\text{ov}[\gamma_1(x, S^y), \eta^\star(S^y)] > 1 + \vartheta\beta \mathbb{C}\text{ov}[\gamma_1(x, S^y), \eta(S^y)] = \eta'(x),$$

where the second inequality follows from the guess that $\eta'^\star(x) > \eta'(x)$. Therefore, $\eta'^\star(x) > \eta'(x)$ for all x, which verifies the guess. Suppose now that γ is submodular and guess that $\eta'^\star(x) < \eta'(x)$. Then, by a similar argument, $\mathbb{C}\text{ov}[\gamma_2(S^y, x), \eta^\star(S^y)] < 0$, and

$$\eta'^\star(x) < 1 + \vartheta\beta \mathbb{C}\text{ov}[\gamma_1(x, S^y), \eta^\star(S^y)] < 1 + \vartheta\beta \mathbb{C}\text{ov}[\gamma_1(x, S^y), \eta(S^y)] = \eta'(x),$$

where $\mathbb{C}\text{ov}[\gamma_1(x, S^y), \eta^\star(S^y)] < \mathbb{C}\text{ov}[\gamma_1(x, S^y), \eta(S^y)]$ follows from the guess. This verifies the guess, and proves Proposition 5.2.

C Model extensions

C.1 Adding Skill Complementarities in Production

To do.

D Quantitative Model

D.1 Workers' problem

Consumption Young and old workers solve a standard consumption choice problem that maximize their static utility. A worker with income y thus consumes

$$c_\ell(y) = \alpha y \quad ; \quad h_\ell(y) = \left(\frac{1-\alpha}{p_\ell}\right) y$$

on the numeraire and the housing good respectively. Accordingly, the local price index is $P_\ell = p_\ell^{1-\alpha}$ and indirect utility in city ℓ for a worker with income y is y/P_ℓ.

Old workers Worker' income are made of three sources: their wage, the subsidy they might receive, and the flat tax. Hence, the income of an old worker with skill s living in city l when young and choosing to live in city ℓ when old is $y^o_{l\ell}(s) = w_\ell(s) + \tau^o_{l\ell}(s) - t$. Given this income, old workers solve a static location choice problem,

$$V^o_l(s, \varepsilon) = \max_\ell \; \frac{y_{l\ell}(s)}{P_\ell} + \varepsilon_\ell - \kappa^o_{l\ell}.$$

[99]Take three increasing function, f, g and h. Suppose that $g' > h'$. We want to show that $\mathbb{C}\text{ov}[f(S), g(S)] > \mathbb{C}\text{ov}[f(S), h(S)]$. Since the covariance is a linear operator, we have $\mathbb{C}\text{ov}[f(S), g(S)] - \mathbb{C}\text{ov}[f(S), h(S)] = \mathbb{C}\text{ov}[f(S), g(S) - h(S)] > 0$, which follows from the covariance of two increasing functions being positive and $g - h$ increasing from $g' > h'$.

Letting $m_l^o(s)$ denote the spatial distribution of old workers before their new location choice, the spatial allocation of workers is given by

$$n_{l\ell}^o(s) = m_l^o(s)\left(\frac{\mathrm{e}^{\vartheta\left(y_{l\ell}^o(s)/P_\ell + B_\ell^o - \kappa_{l\ell}^o\right)}}{\sum_{\ell'} \mathrm{e}^{\vartheta\left(y_{l\ell'}^o(s)/P_{\ell'} + B_{\ell'}^o - \kappa_{l\ell'}^o\right)}}\right), \tag{49}$$

where $n_{l\ell}^o(s)$ is the measure of workers moving from l to ℓ with skill s when old. Consistently, the (unconditional) measure of old workers in city ℓ is

$$n_\ell^o(s) = \sum_l n_{l\ell}^o(s),$$

and the mass of old workers in ℓ is $N_\ell^o = \int n_\ell^o(s)ds$. Finally, let $\mathcal{V}_l^o(s) \equiv \mathbb{E}[V_l^o(s,\boldsymbol{\varepsilon}) \mid s, l]$ be the expected utility before the realization of the idiosyncratic preferences $\boldsymbol{\varepsilon}$, given by

$$\mathcal{V}_l^o(s) = \frac{1}{\vartheta}\log\left(\sum_\ell \mathrm{e}^{\vartheta\left(y_{l\ell}^o(s)/P_\ell + B_\ell^o - \kappa_{l\ell}^o\right)}\right).$$

Young workers Young workers solve a dynamic location choice problem which reads

$$V_l^y(s,\boldsymbol{\varepsilon}) = \max_\ell \frac{y_{l\ell}^y(s)}{P_\ell} + \varepsilon_\ell - \kappa_{l\ell}^y + \beta \int\int \mathcal{V}_\ell^o\left[e\gamma(s,s_p)\right] dF(e)\pi_\ell(s_p)ds_p,$$

where $y_{l\ell}^y(s) = w_\ell(s) + \tau_{l\ell}^y(s) - t$. In a similar fashion as for the old workers, the spatial allocation of young workers is given by

$$n_{l\ell}^y(s) = m_l^y(s)\left(\frac{\mathrm{e}^{\vartheta\left(y_{l\ell}^y(s)/P_\ell + B_\ell^y - \kappa_{l\ell}^y + \beta O_\ell(s)\right)}}{\sum_{\ell'} \mathrm{e}^{\vartheta\left(y_{l\ell'}^y(s)/P_{\ell'} + B_{\ell'}^y - \kappa_{l\ell'}^y + \beta O_{\ell'}(s)\right)}}\right). \tag{50}$$

I define $n_\ell^y(s)$ and N_ℓ^y in the same manner as $n_\ell^o(s)$ and N_ℓ^o.

D.2 General equilibrium

In equilibrium, the markets must clear, the expected spatial distribution of workers has to be consistent with workers' location choices, and the aggregate skill distribution must be consistent with workers' learning.

Housing prices Housing prices must clear the housing market in each location, and therefore they must satisfy[100]

$$p_\ell = \left(\frac{(1-\alpha)\,Y_\ell}{\mathcal{H}}\right)^{\frac{1}{1+\delta}}, \tag{51}$$

where $Y_\ell = T_\ell \int s \sum_a n_\ell^a(s)ds = \mathbb{E}_\ell[W]N_\ell$ is the output produced in city ℓ. Taking logs give (20).

Skill density The within-city skill density of location ℓ must be consistent with workers' location choices, given by

$$\pi_\ell(s) = \frac{\sum_a n_\ell^a(s)}{\sum_a N_\ell^a}. \tag{52}$$

Skill distributions There are two distributions to be solved for. First, the location-specific distribution of old workers. Define $M_\ell^o(s)$ as the mass of old workers that lived in city ℓ when young and that obtained a skill less than s. Given the learning process, this is given by

$$M_\ell^o(s) = \int\int n_\ell^y(x)\pi_\ell(y)F\left(\frac{s}{\gamma(x,y)}\right)dydx. \tag{53}$$

[100]Note that this implies $\Pi_\ell = p_\ell^{1+\delta} = (1-\alpha)Y_\ell$, where Π_ℓ are the revenues earned by the representative land owner.

From here, one can define the location-specific measure as $m^o_\ell(s) = \partial M^o_\ell(s)/\partial s$ and the aggregate old skill density as $n^o(s) = \sum_\ell m^o_\ell(s)$. Finally, given the birth process, the location-specific distribution of young workers is $m^y_\ell(s) = N^y_\ell n^y(s)$.

Definition D.1 (Steady state equilibrium)**.**
A steady state equilibrium is a collection of young and old spatial allocation, $s \to n^y_{l\ell}(s)$ and $s \to n^o_{l\ell}(s)$ for each $l, \ell \in \{1, 2, \dots, L\}$, city-specific skill distribution, $s \to m^y_\ell(s)$ and $s \to m^o_\ell(s)$ for each $\ell \in \{1, 2, \dots, L\}$, and housing prices $\{p_\ell\}_{\ell=1}^L$, such that

1. *Taking as given the location decisions of other workers and housing prices, the young and old allocation satisfy* (50) *and* (49) *respectively;*
2. *The local skill densities are consistent with workers' location decisions, given by* (52)*;*
3. *Given young workers' location decisions, the city-specific young and old skill distribution are given by $m^y_\ell(s) = N^y_\ell n^y(s)$ and* (53) *respectively;*
4. *Given workers' location decisions, housing prices solve* (51)*.*

D.3 Algorithm

I use two different algorithms to solve for the steady state equilibrium depending on whether I am estimating the model or computing counterfactuals. Both algorithms are constituted of an inner loop and an outer loop. The inner loops are identical. Given city fundamentals, $\boldsymbol{T}$ and $\boldsymbol{B}$, housing prices $\boldsymbol{p}$, and skill distributions $\boldsymbol{m}^y$ and $\boldsymbol{m}^o$, they iterate on (50) and (49) to solve for the local skill densities, $\boldsymbol{\pi}$.

Estimation To estimate the model, I need to solve for the steady state equilibrium and recover the city characteristics, $\boldsymbol{T}$ and $\boldsymbol{B}$. The targeted moments for $\boldsymbol{T}$ and $\boldsymbol{B}$ imply that, when TFP and amenities are estimated, I match the city average wage and city size, and therefore cities' output, $\mathbb{E}_\ell[W]N_\ell$. Given estimates for $(\alpha, \delta, \mathcal{H})$, I can therefore read off (51) the housing prices that must hold in the baseline steady state equilibrium. To solve for the baseline steady state equilibrium, the outer loop therefore takes $\boldsymbol{p}$ as given and iterate on the skill distributions, $\boldsymbol{m}^y$ and $\boldsymbol{m}^o$, as well as on the city TFP and amenities through (23) and (24).

Counterfactual In the counterfactuals, city TFP and amenities are taken as given. The outer loop therefore iterates on (51) to solve for the housing prices, as well as on $m^y_\ell(s) = N^y_\ell n^y(s)$ and (52) to solve for $\boldsymbol{m}^y$ and $\boldsymbol{m}^o$.

Simulation I use the model to simulate a two-period panel datasets with $1,000,000$ workers. The simulation follows the structure of the model. All the variables, including the local skill densities, are consistent with the model's general equilibrium. As in the data, I truncate the wage distribution at the 5% and 99.9% percentile.

E Descriptive Evidence

E.1 Data description

[To-do: describe.]

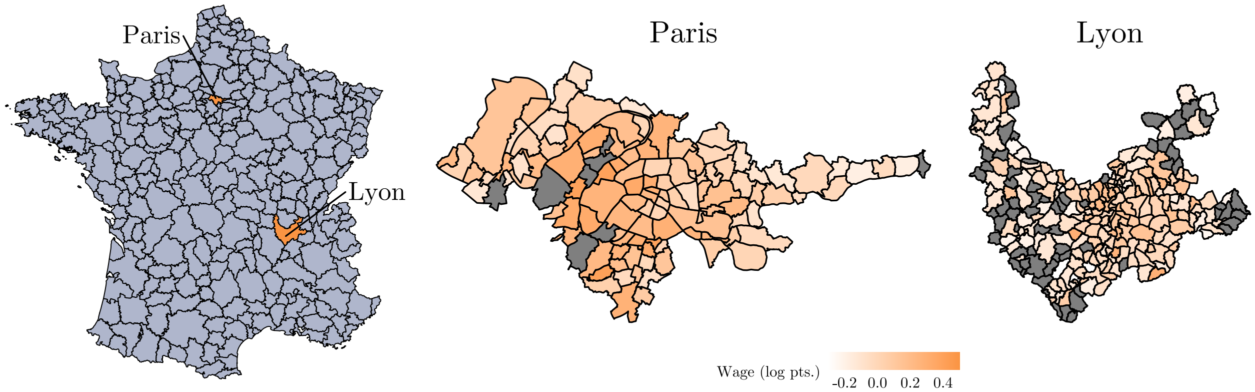
E.2 Neighborhood

As explained in Section 3.1, I define a neighborhood as a municipality, or commune in French. Commune are the most granular administrative division in France. They are also the most granular geographic unit I observe in the French matched employer-employee dataset.

Commune were introduced with the French revolution in 1789 to unify parishes (the houses and land around a church) and chartered cities (collection of parishes) under a common legal definition. By the end of the 18$^{\text{th}}$ century, 40,200 municipalities were created. Two reforms at the beginning of the 19$^{\text{th}}$ century (1831 and 1837) increased their administrative powers. Since then, both the geographic and administrative boundaries of municipalities have been stable over time. For instance, around 90% of commune are have similar geographic boundaries as when they were created during the French revolution.

In 2009, France had 36 783 communes. Their legislative power are analogous to civil townships and incorporated municipalities in the United States. In terms of geography, they are closer to ZIP codes. The three largest French

Figure E.1: The geography of commuting zones and municipalities



communes, Paris, Lyon and Marseille, are themselves sub-divided into 20, 9, and 16 arrondissements. I observe those in the French matched employer-employee dataset, and therefore refer to these arrondissements also as neighborhoods.

Figure E.1 helps visualize the granularity of communes. The left map is a map of the French commuting zones, where I have highlighted in orange the two largest French cities, Paris and Lyon. The two maps on the right are the commune included in the Paris (112 communes) and Lyon (191 communes) commuting zones. The orange shade represents the average wage of the neighborhood relative to the French average. Grey communes are communes excluded from the sample due to their small sample sizes. In total, 112 and 191 communes are kept in the sample in Paris and Lyon respectively. Throughout France, the average number of commune by commuting zones is 40. The city with the smallest number of commune is Chatillon (0.02% of the French workforce) with 4 communes. The city with the largest number of commune is Roissy (2.7% of the French workforce) with 212 communes.

E.3 Identification

In this section, I lay down how (18) identifies the learning technology. First, write wages as $\log w_{it} = \log T + \mathcal{T}_{\ell_{it}} + \log s_{it}$, where I have decomposed the city TFP into an aggregate component T and a city-specific productivity $\mathcal{T}_\ell$, $\log T_\ell = \log T + \mathcal{T}_\ell$. Alternatively, I could have decomposed $\log s_{it}$ into an aggregate component and a worker-level productivity. That is, the levels of T_ℓ and s are not separately identified. Let $\omega_{it} \equiv \log w_{it} - \log T$.[101] Then, ω_{it}, $\mathcal{T}_{\ell_{it}}$ and $\log s_{it}$ are all mean zero.

From here, substitute $\log s_{it} = \omega_{it} - \mathcal{T}_{\ell_{it}}$ and $s_{p_{it}} = s_{p_{it}} + \mathbb{E}_{\ell_{it}}[\log w_{it}] - \mathbb{E}_{\ell_{it}}[\log w_{it}]$ into (15) to obtain

$$\omega_{it+1} = g_0 + g_1\omega_{it} + g_2\mathbb{E}_{\ell_{it}}[\omega_{it}] + g_{12}\omega_{it}\mathbb{E}_{\ell_{it}}[\omega_{it}] + \log e_{it+1} + \zeta_{it} + \nu_{it}, \tag{54}$$

where

$$\zeta_{it} \equiv \mathcal{T}_{\ell_{it+1}} - (g_1 + g_2 + g_{12}\omega_{it} + g_{12}\mathbb{E}_{\ell_{it}}[\omega_{it}] - g_{12}\mathcal{T}_{\ell_{it}})\mathcal{T}_{\ell_{it}}.$$

and

$$\nu_{it} \equiv (g_2 + g_{12}\log s_{it})\left(\omega_{p_{it}} - \mathbb{E}_{\ell_{it}}[\omega_{it}]\right).$$

Specification (54) has three error terms. First, $\log e_{it+1}$ which captures learning sources other than interactions. Second, ζ_{it}, which reflects the gap between worker-level productivity and wages. And third, ν_{it}, which measures the difference between the particular interaction experienced by i and the average interaction in city ℓ_{it}. If the three error terms are uncorrelated with ω_{it} and $\mathbb{E}_{\ell_{it}}[\omega_{it}]$, then the local projection (18) yields unbiased estimates of the learning technology: $\hat{\beta} = g_1$, $\hat{\gamma} = g_2$ and $\hat{\delta} = g_{12}$.

First, Assumption 2.2 implies that $\mathbb{E}[\omega_{it}\log e_{it+1}] = \mathbb{E}[\mathbb{E}_{\ell_{it}}[\omega_{it}]\log e_{it+1}] = 0$.

Second, Assumption 2.1 implies that ν_{it} is uncorrelated with the regressors. To see why, note that

$$\mathbb{E}[\nu_{it}] = \mathbb{E}[\mathbb{E}[\nu_{it} \mid s_{it}, \ell_{it}]] = \mathbb{E}[(g_2 + g_{12}\log s_{it})\,\mathbb{E}[\omega_{p_{it}} - \mathbb{E}_{\ell_{it}}[\omega_{it}] \mid s_{it}, \ell_{it}]] = 0,$$

[101] In the data, $\log T$ is computed as the nationwide average wage.

Table E.1: Estimated learning technology (observed vs. unobserved interactions)

	Parameter	OLS: skill observed	OLS: skill unobserved
g_0	0.170	0.170	0.166
		[0.170, 0.171]	[0.166, 0.167]
g_1	0.057	0.057	0.053
		[0.055, 0.059]	[0.052, 0.055]
g_2	0.327	0.327	0.324
		[0.326, 0.329]	[0.318, 0.329]
g_{12}	0.393	0.390	0.397
		[0.386, 0.394]	[0.389, 0.406]
R^2		0.225	0.017
Obs.		955655	955655

First column reports the learning technology parameters estimated from the data and used in the model's calibration. The second and third column estimates the learning technology from simulated data. The simulation method is presented in Section D.3. The second column estimates $\log s_{it+1} = \alpha + \beta \log s_{it} + \gamma \log s_{p_{it}} + \delta \log s_{it} \log s_{p_{it}} + u_{it}$. The third column estimates $\log s_{it+1} = \alpha' + \beta' \log s_{it} + \gamma' \mathbb{E}[\log s_{p_{it}} \mid \ell] + \delta' \log s_{it} \mathbb{E}[\log s_{p_{it}} \mid \ell] + u_{it}$.

where the first equality follows from the L.I.E., the second from the definition of ν_{it}, and the third from $\mathbb{E}[\omega_{p_{it}} \mid s_{it}, \ell_{it}] = \mathbb{E}_{\ell_{it}}[\omega_{it}]$ when interactions are random within cities. A similar argument implies $\mathbb{E}[\nu_{it}\omega_{it}] = \mathbb{E}[\nu_{it}\mathbb{E}_{\ell_{it}}[\omega_{it}]] = \mathbb{E}[\nu_{it}\omega_{it}\mathbb{E}_{\ell_{it}}[\omega_{it}]] = 0$.

Third, when $\mathcal{T}_\ell$ is observed, ζ_{it} can be measured, and therefore controlled for.

These three remarks allow me to build a two-step estimator of the learning technology. First, suppose that $\mathcal{T}_\ell \approx 0$, and estimate $(\hat{\beta}^0, \hat{\gamma}^0, \hat{\delta}^0)$ from (18). Given those estimates, recover the vector of city TFP through (23), $\hat{\boldsymbol{T}}^0$. Given $\hat{\boldsymbol{T}}^0$, construct $\hat{\zeta}^1_{it}$, and obtain a new set of parameters $(\hat{\beta}^1, \hat{\gamma}^1, \hat{\delta}^1)$ from estimating (18) while controlling for $\hat{\zeta}^1_{it}$.

The fixed point defined by this iteration relies implicitly on the sorting of workers, and thus on the entire general equilibrium. I therefore cannot show that this iterative algorithm yields unbiased estimates of the learning technology for any (g_1, g_2, g_{12}) and $\boldsymbol{T}$. However, when the spatial TFP gaps are small, $\mathcal{T}_\ell \approx 0$ for all ℓ, and therefore $\zeta_{it} \approx 0$. In such a case, (18) yields unbiased estimates of the learning technology and the iterative algorithm converges to the true (g_1, g_2, g_{12}).

In practice, I estimate that the dispersion in city TFP, $\mathbb{V}\text{ar}[\mathcal{T}_\ell]$, explains 2% of the aggregate wage variance. I take this as suggestive evidence that $\zeta_{it} \approx 0$. Accordingly, when controlling for ζ_{it}, I cannot statically differentiate the new sets of estimates from the baseline estimates at the 5%.

[To-do: discuss the reflection problem here.]

E.4 Instrument variable

In Section 3.3, I argue that I can isolate spatial variation in skill density across neighborhoods within cities that are orthogonal to neighborhood-level wage shocks. To do so, I instrument the contemporaneous neighborhood average wage by the change in the white-collar employment share between 1993 and 2000. I define white-collar workers through their occupations. I define a white-collar occupation as the occupations with one-digit occupation code 3 in the French classification. This includes licensed professional (e.g. lawyers and doctors), professors and scientists, managers and engineers. The instrument is relevant if past changes in the white-collar employment share correlate with the current local supply of skill. It satisfies the exclusion restriction if these past changes are not correlated with contemporary neighborhood-level wage shocks.

I provide first suggestive evidence that the instrument is relevant. First, white-collar workers are relatively more skilled than the average worker along several margins. They earn higher wages (38€ per hour compared to 22€ for the average worker), they are on average more educated (XXX% of white-collar workers went to college compared to XXX% on average), and they implement tasks that are less routine and manual and more creative. Hence, the white-collar employment-share in a neighborhood is informative about the local supply of skill. Furthermore, past changes in the white-collar employment-share have long-lasting consequences on the occupational composition of

Table E.2: White-collar employment share across time

	(1) Emp. share 2010	(2) Emp. share 2010	(3) Δ Emp. share 2010-2019	(4) Δ Emp. share 2010-2019
Δ Emp. share 1993-2000	0.330***	0.285***	0.043*	0.027
	(0.032)	(0.029)	(0.020)	(0.020)
Standardized	0.155	0.134	0.027	0.017
City F.E.		✓		✓
Obs.	8,520	8,520	8,520	8,520

Standard errors clustered at the city level. P-value: $^{***} < .001$, $^{**} < .01$, $^{*} < .05$ and $^{+} < .1$. Unit of observation is the neighborhood. Left-hand side variable is the white-collar employment share in 2010 for columns 1 and 2, and the change in the white-collar employment share between 2010 and 2019 for column 3 and 4. The right-hand side variable is the change in the white-collar employment share between 1993 and 2000. All variables truncated at the 1%.

Table E.3: The returns to local interactions (robustness)

Standard errors clustered at the city level for column 1 and at the neighborhood of work for the remaining columns. All regressions include age, job tenure, occupation tenure, and industry tenure fixed effects, as well as a dummy if the worker changed employer between the five years. Column 5 and 6 instruments for the neighborhood wage by the change in the white-collar employment share between 1993 and 2000. Sample includes all workers between 25 and 40 year old. Wage growth distribution truncated at the 5% within cities. Industry and occupation in column 3 defined at the one-digit level. Each fixed effect is interacted with year fixed effects.

neighborhoods. In Table E.2, I project the white-collar employment share in 2010 onto the change in the white-collar employment share between 1993 and 2000. Column 1 is in the cross-section and column 2 is within-city. In both cases, neighborhoods who experienced increase in their white-collar employment share in the 1990s have a higher share of white-collar today on average.

Meanwhile, the changes in the white-collar employment share between 1993 and 2000 are not correlated with the changes between 2010 and 2019 (Column 4 of Table E.2). Hence, the shocks that triggered the increase in the employment share of white collars in the 1990s have either diffused across space or faded away. Either way, I take this has suggestive evidence that the exclusion restriction is satisfied by the instrument.

E.5 Robustness

[To-do: describe the robustness exercices.]

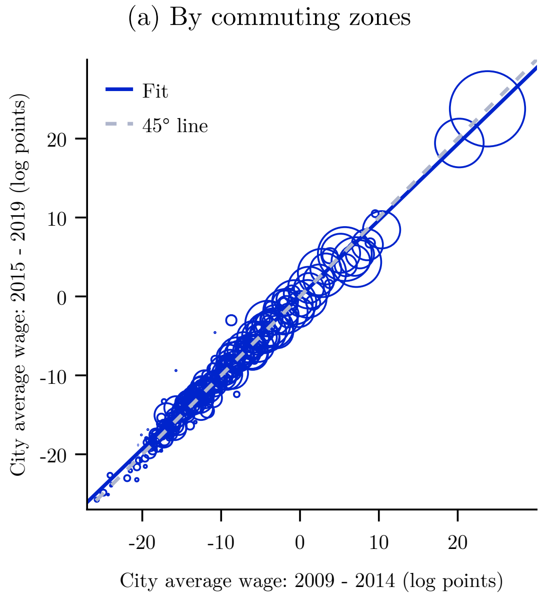
Aggregate growth 1. Plot of stability 2. Plot of real wages 3. All the robustness we are going to include.

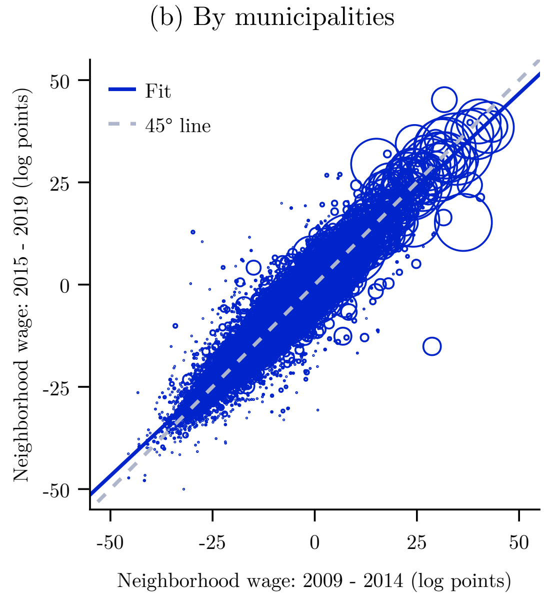
F Model Estimation

F.1 Clustering

Cities are clustered into 30 city types. The ten first types are the ten largest French cities. The twenty remaining types are constructed by a k-mean algorithm so that there are five types within each French district (North-West, North-East, South-West, and South-East). Figure F.1 shows the mapping between cities to city types within each district. The grey areas are the ten largest French cities. Each other color represents a city type within the district. While k-mean is not a spatial clustering algorithm, including the latitude and longitude of the cities in the clustering variables is sufficient to generate city types that are spatially contiguous. Table F.1 lists the number of cities by city type, the largest city in each type, the average size, as well as the average wage. Finally, Figure F.2 plots the within city-type sum of square residuals (SSR) as a fraction of the total SSR. With only 5 city types, the within city type is less than a fifth of the total SSR. A parsimonious geographic representation of France is therefore able to capture most of the variation of interest for the model.

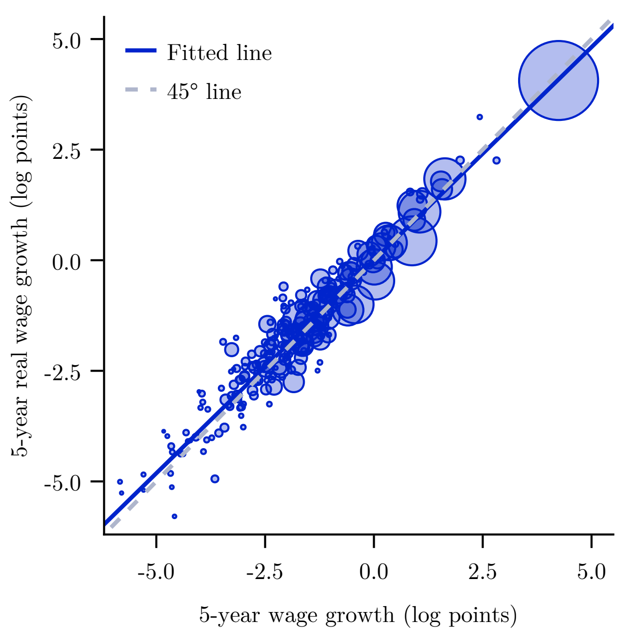
Figure E.2: Average wage at the beginning and the end of the 2010s





Note: left panel shows the average wage by commuting zone between 2009-2014 on the x-axis against the average wage in the same commuting zone between 2015-2019 on the y-axis. The marker size is proportional to the size of the commuting zone. The right-panel shows the same relationship but at the municipality level. The marker size is proportional to the size of the municipality. The aggregate average wage is normalized to zero in both periods for both panels. The blue solid line is the (unweighted) best fitted line between the 2009-2014 and 2015-2019 average wages. The grey dotted line is the 45 degree line.

Figure E.3: Nominal and real wage growth



Note: figure shows the relationship between real wage growth and nominal wage growth by commuting zones. Real wage growth are computed as $\mathbb{E}_\ell[\log w_{it+1}/w_{it}] - \alpha \log(p_{t+1\ell}/p_{t\ell})$, where the second term is the growth in rental prices. The housing expenditure share is set to 0.2 (see Table F.2). The marker size is proportional to the size of the commuting zone.

Figure E.4: The returns to local interactions by time horizon

(a) Log wage (b) City wage (c) Log wage × city wage

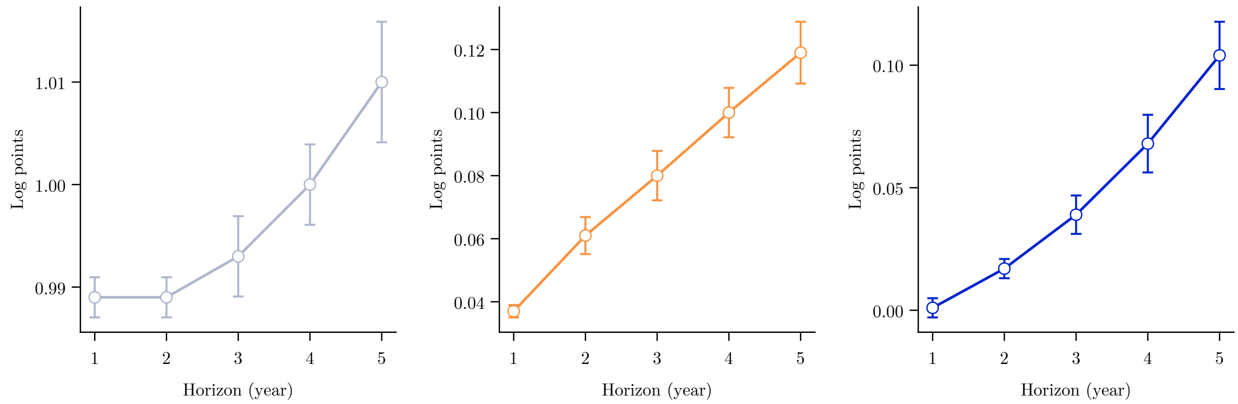


Figure E.5: The returns to local interactions by age

(a) Log wage (b) City wage (c) Log wage × city wage

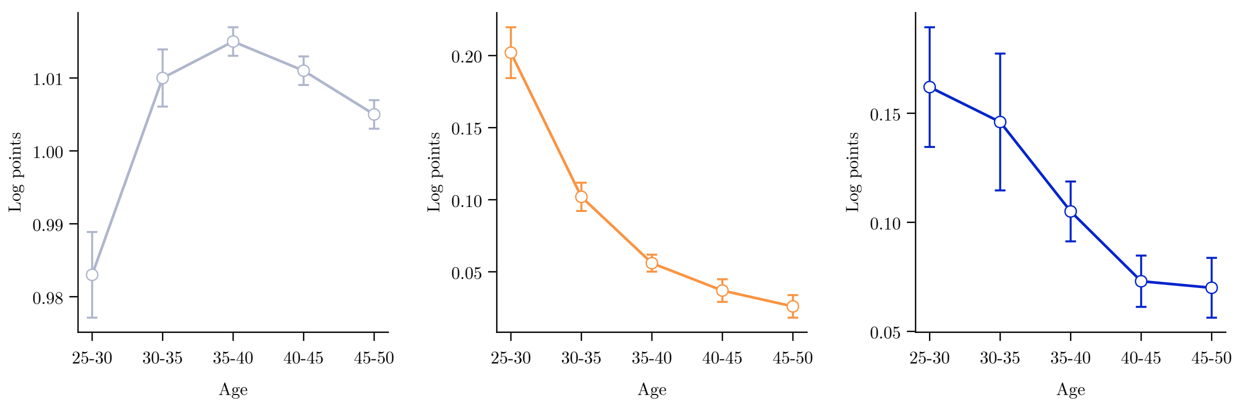


Figure E.6: The returns to local interactions by city size

(a) Log wage (b) City wage (c) Log wage × city wage

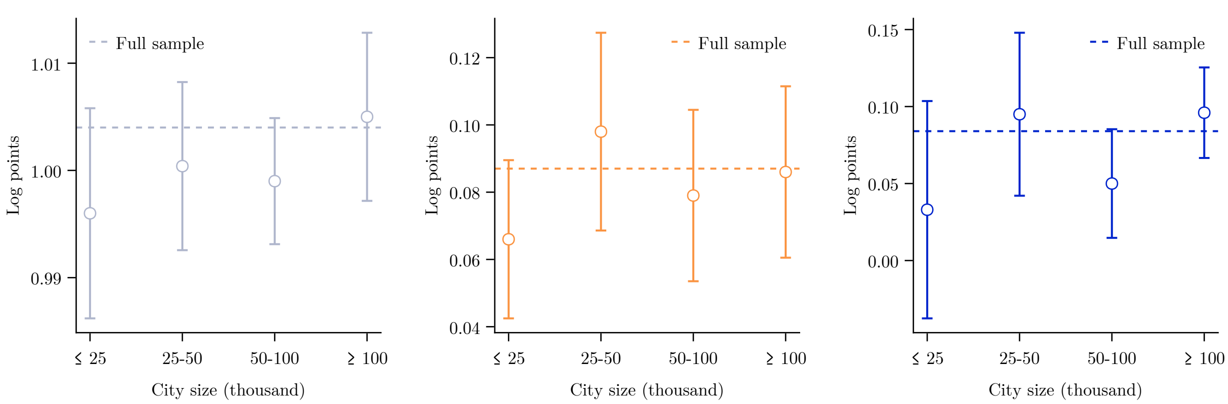


Figure E.7: The returns to local interactions by city inequality

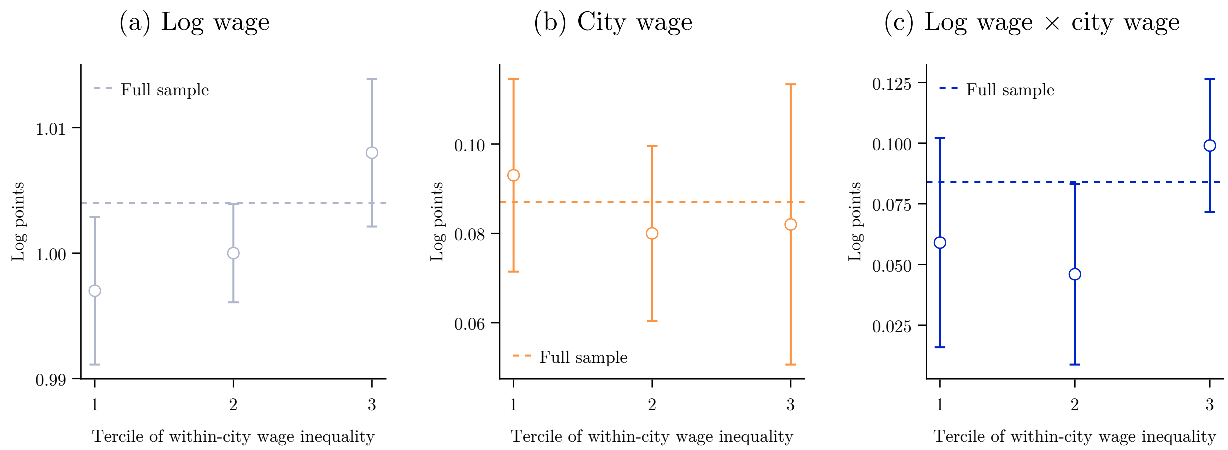
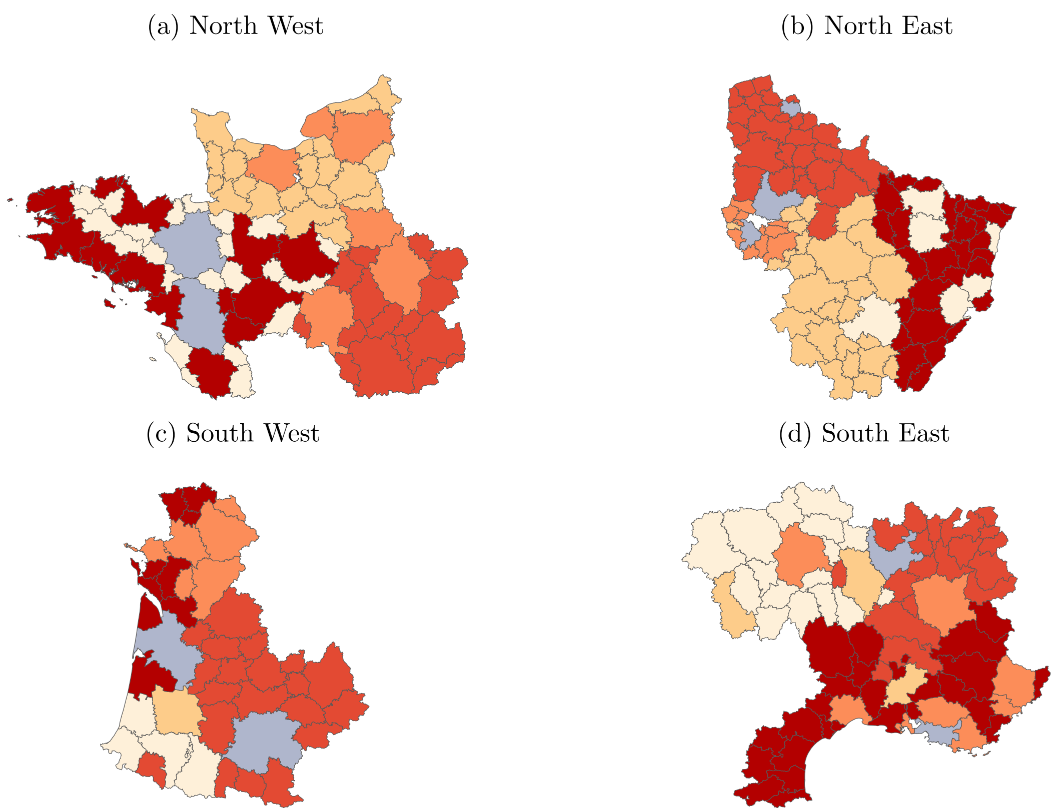


Figure F.1: City clustering



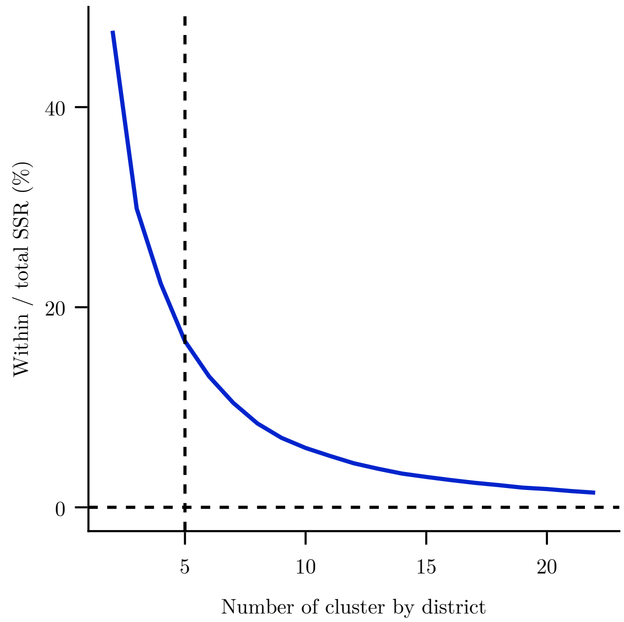
Note: each subplot is a French district. The grey areas are the ten largest French cities. The remaining commuting zones are subdivided into five groups within each district. Each city cluster is represented by a different color on the map. See Table F.1 for a list of the clusters.

Table F.1: City types characteristics

Cluster	District	# cities	Largest city	Average size (thousands)	Average wage
1	Paris	1	Paris	2800.6	30.03
2	SE	1	Lyon	609.8	23.38
3	NE	1	Roissy	443.0	22.50
4	NE	1	Saclay	414.1	27.32
5	SW	1	Toulouse	403.8	22.42
6	SW	1	Bordeaux	332.6	21.07
7	SE	1	Marseille	316.8	22.14
8	NW	1	Nantes	305.8	20.94
9	NE	1	Lille	258.4	22.06
10	NW	1	Rennes	218.4	20.54
11	NE	6	Strasbourg	131.9	20.67
12	NE	26	Troyes	32.64	18.61
13	NE	11	Orly	109.7	22.88
14	NE	27	Roubaix	58.53	19.46
15	NE	28	Besançon	32.47	19.57
16	NW	22	Challans	16.70	17.65
17	NW	21	Évreux	26.13	18.80
18	NW	6	Rouen	141.9	20.33
19	NW	15	Blois	28.05	18.86
20	NW	13	Angers	74.63	18.94
21	SW	4	Pau	65.64	19.31
22	SW	1	Mont-de-Marsan	29.40	17.29
23	SW	6	Poitiers	54.48	19.42
24	SW	20	Périgueux	25.87	18.22
25	SW	9	La Teste-de-Buch	17.74	17.26
26	SE	16	Limoges	34.67	18.08
27	SE	3	Saint-Étienne	127.5	19.56
28	SE	8	Grenoble	146.2	21.67
29	SE	19	Valence	52.71	20.31
30	SE	26	Nîmes	34.06	18.29

The tables report statistics on the city types used for the estimation of the quantitative model. The second column is the district in which the city type belongs (see Figure F.1). The third column is the number of cities included in the city city type. For the first ten types, there is a single city by definition. For the twenty remaining types, the number of cities per city type is determined by the k-mean algorithm. The fourth column reports the largest city in the city type, the fifth column the average number of employed workers (in thousands), and the sixth column the average wage.

Figure F.2: Residual of city clustering



Note: average within cluster sum of square residuals (SSR) relative to the total SSR against the number of cluster (per district) used in the k-mean algorithm. The quantitative model features five city-types per district, as represented by the vertical dashed black line.

F.2 Parameters

This section presents the parameters used in the quantitative model. Table F.2 presents the city-independent estimated parameters. Figure F.3 displays the estimated migration costs by age and city of origin as a function of the distance between the two cities. Figure F.4 plots the estimated city TFP and age-specific amenities. Figure F.5c shows the fit of the model across space by comparing the city average wage (panel a), the city employment share (panel b), and city average rents (panel c) in the data and the model.

F.3 Sensitivity

Figure F.7 and F.8 reports information to help assess the sources of identification in the indirect inference procedure. Specifically, Figure F.7 plots the value of the targeted moments in the indirect inference procedure as a function of the parameters. Each subplot is a given moment. Each line represents a particular parameter. The blue lines highlight the parameters that are meant to be identified by the moment.

Figure F.7 confirms that all parameters are well identified locally. Each targeted moment is indeed a steep function of the parameter they are meant to identify. For instance, the flattest relationship is between the wage variance for old workers and the learning shocks dispersion. Even then, a 1% increase in σ_ν leads to a 0.7% increase in the targeted moment. Figure F.7 also shows that some parameters affect mainly one moment whereas others drive several moments. For instance, the dispersion in idiosyncratic location preferences only affect the between-city share of the wage growth variance. On the contrary, the learning shocks dispersion affects both the wage variance for old workers and the old to young wage ratio.

Figure F.8 complements the analysis of Figure F.7 by reporting the sensitivity measure of Andrews et al. (2017). Intuitively, this metric is the inverse of the moment elasticities reported in Figure F.7. It quantifies how variation in targeted moments would influence estimated parameter values. Theoretically, let $\boldsymbol{h}(\boldsymbol{\theta})$ denote the correspondence that maps the vector of parameters $\boldsymbol{\theta}$ into the vector of targeted moments. Let $\boldsymbol{J}_h$ denote the Jacobian matrix of this function evaluated at the estimated parameters. Then, the sensitivity matrix is computed as $\boldsymbol{\Lambda} = (\boldsymbol{J}'_h\boldsymbol{J}_h)^{-1}\boldsymbol{J}'_h$. Figure F.8 displays the absolute value of $\boldsymbol{\Lambda}$, and the sign of $\boldsymbol{\Lambda}$ is reported in parenthesis. As in Figure F.7, the blue bar highlights which moment is supposed to identify the specific parameter.

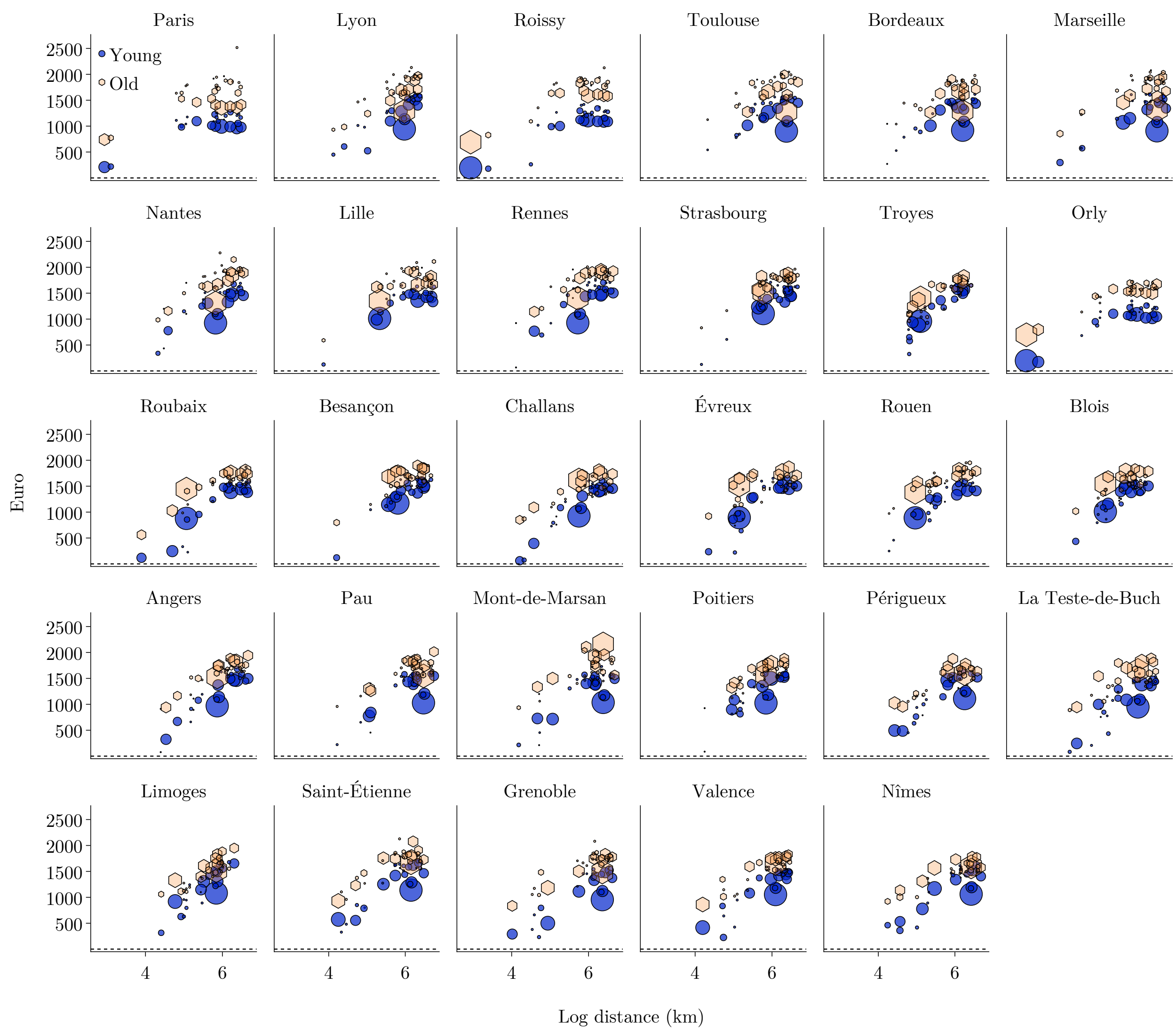
I draw two takeaways from Figure F.8. First, each moment is very informative for the parameters they are meant to identify. Second, some moments are informative for several parameters. For instance, the wage variance

Table F.2: Estimated parameters

	Parameter	Value	Moment	Data	Model
			A. Externally calibrated		
β	15-year discount factor	0.684		·	·
			B. Model inversion		
$1-\alpha$	Housing exp. share	0.200	Housing exp. share	0.200	0.200
$\{\kappa^a_{\ell\ell'}\}_{\ell,\ell'}$	Migration cost	·	Migration flows (21)	·	·
δ	Housing price elasticity	7.473	Housing price $\sim$ exp.: constant (20)	·	·
$\mathcal{H}$	Supply stock	7.2×10^{-9}	Housing price $\sim$ exp.: slope (20)	·	·
g_1	Learning technology	1.057	Wage growth $\sim$ wage (18)	·	·
g_2	Learning technology	0.327	Wage growth $\sim$ city wage (18)	·	·
g_{12}	Learning technology	0.393	Wage growth $\sim$ wage $\times$ city wage (18)	·	·
			C. Indirect inference		
$\{T_\ell\}_\ell$	City TFP	·	Average wage	·	·
$\{B_{a\ell}\}_{a,\ell}$	City amenity	·	City size	·	·
g_0	Learning technology	0.400	Old / young wage	1.250	1.248
ϑ	Taste shock dispersion	0.009	B/in city / total growth var.	0.017	0.018
μ^s_1	Mean young skill	-0.200	Average skill	0.000	0.003
σ^s_1	Variance young skill	0.309	Variance young wage	0.139	0.141
σ_ν	Var. learning shock	0.310	Variance old wage	0.234	0.240

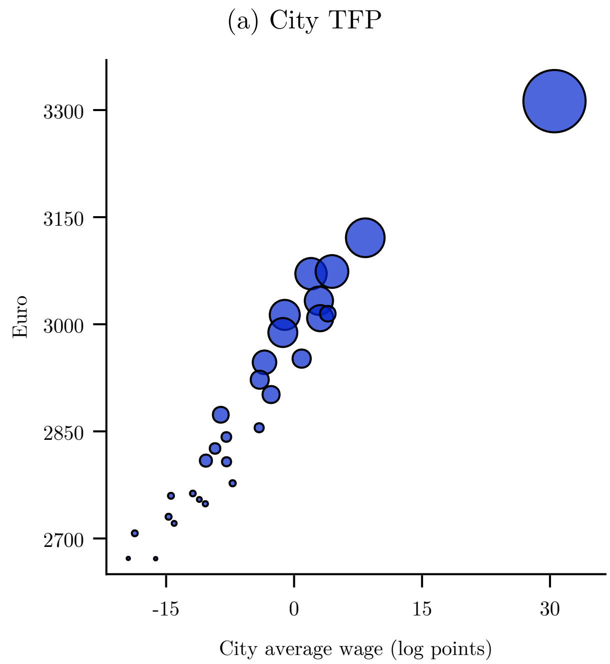
Parameter used in the quantitative model. The first two columns describe the parameter and the third column presents the value of the parameter. The fourth column describes the moment used to estimate or calibrate the parameter. The fifth and sixth columns show the value of that moment in the data and in the model. The migration costs are plotted in Figure F.3, the city TFP in Figure F.4a, and city amenities in Figure F.4b.

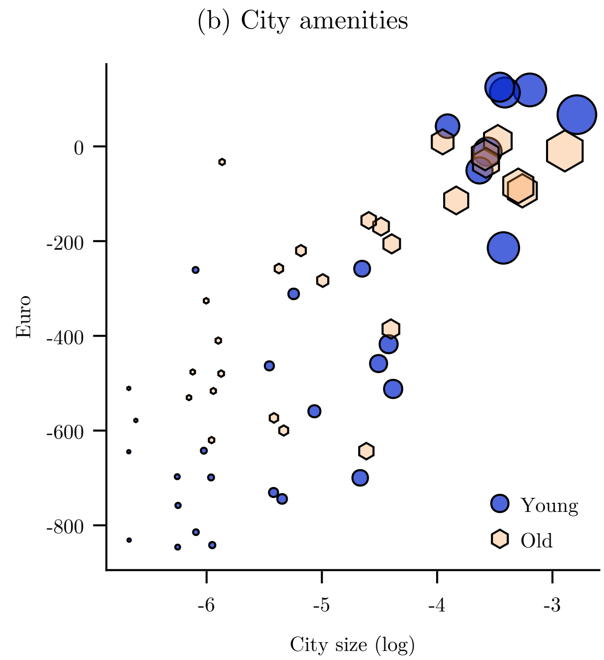
Figure F.3: Estimated age-specific migration costs by city of origin



Note: estimated age-specific migration costs by city of origin as a function of the distance between the city of origin and the destination. Each subplot corresponds to a city of origin. The migration costs are estimated according to (21). The estimated costs are in utils, and they are reported here in euros through $\kappa^{a}_{l\ell} P_{\ell}$. The blue circles and orange hexagons are the migration costs for young and old workers. Distance between two cities computed as the Haversine distance. For city types that comprise more than one city, I define the coordinates of the city type as the coordinates of the most populous city in that cluster. For reference, the average wage in the economy is 3,278€. The size of the circles is proportional to the size of the city of destination.

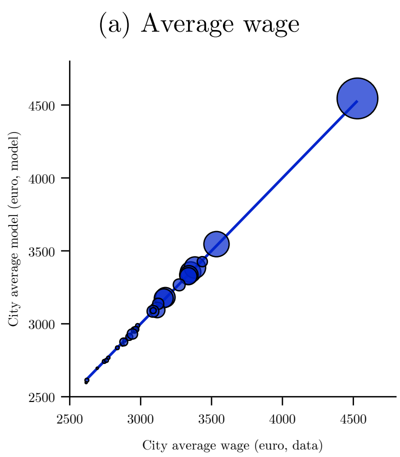
Figure F.4: Estimated city characteristics

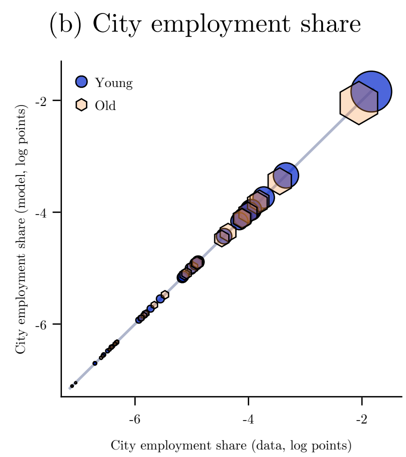


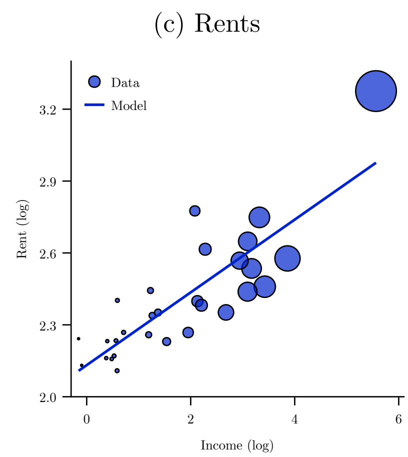


Note: panel (a) plots the estimated city TFP against the city average log wage. City TFP calibrated from (23). Panel (b) plots the estimated age-specific amenity against the age-specific city employment share in log. The blue circles and orange hexagons represent the young and old amenity. Amenities calibrated according to (24). The amenity value of Paris is normalized to zero for young and old and is omitted from panel (b) for improved visual clarity. Amenities are converted into monetary value through $B_\ell^a P_\ell$. The size of the circles is proportional to city size.

Figure F.5: Model fit

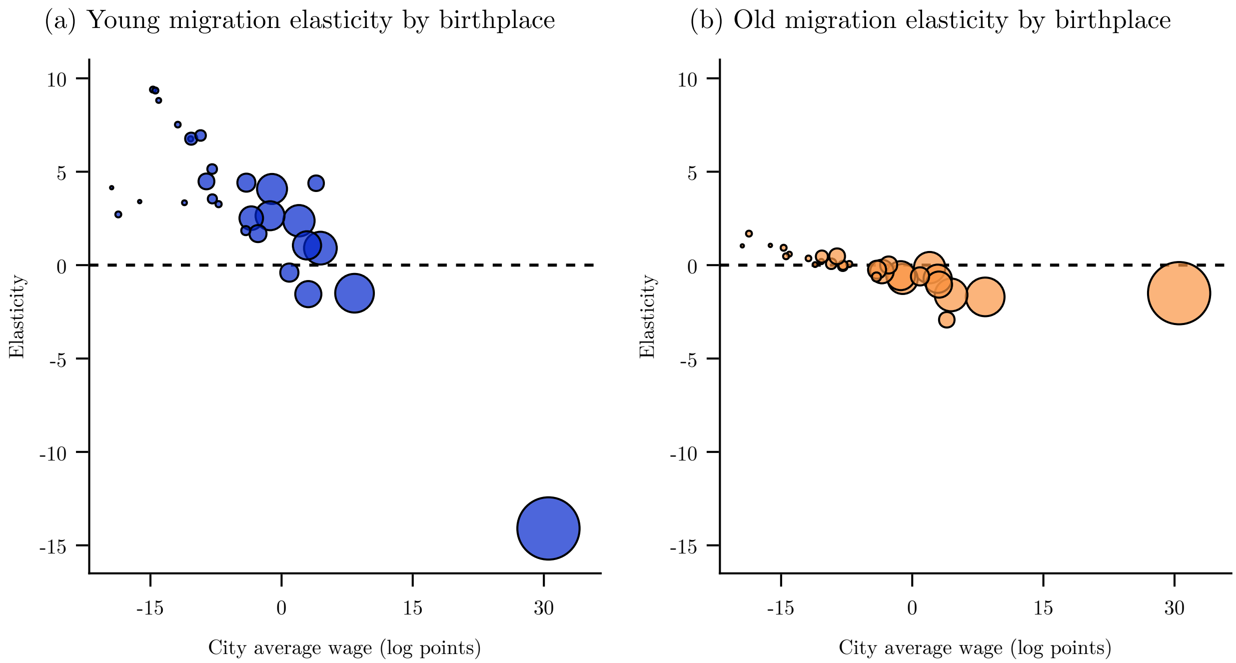






Note: panel (a) plots the average wage in the data (x-axis) against the average wage in the model (y-axis). Panel (b) plots the age-city city employment share for young (blue circles) and old workers (orange hexagons) in the data (x-axis) and in the model (y-axis). Panel (c) plots the rent in the data (blue circle) and the rent in the model (blue line) against city income. The size of the circles are proportional to the size of the cities.

Figure F.6: Migration elasticities across space



Note: panel (a) and (b) display the migration elasticities for young and old workers as estimated by (22). Skills are normalized in the regression, so that the elasticities correspond to the percentage point change in migration probability in response to an increase in skill by one standard deviation. The size of the circles are proportional to the size of the cities.

of old workers matters for the calibration of σ_s, σ_ν and ϑ. While this sensitivity measure sheds further light on the importance of the targeted moments for the calibration of the model, the magnitude of the sensitivity cannot however be interpreted.

F.4 Over-Identification Exercises

The between-city wage gaps can be decomposed into a TFP gap and a skill gap,

$$\mathbb{E}_\ell[\log W] - \mathbb{E}[\log W] = \log T_\ell - \mathbb{E}[\log T] + \mathbb{E}_\ell[\log S]. \tag{55}$$

The TFP and skill gaps are plotted in Figure 3a and 3b respectively. The blue and orange dashed lines correspond to the (employment-weighted) fitted lines of the TFP gaps on wage gaps skill gaps on wage gaps respectively. These fitted lines have a theoretical interpretation. From (55), the variance of spatial wage gaps can be broken down into three terms

$$\mathbb{V}\mathrm{ar}\left(\mathbb{E}_\ell[\log W] - \mathbb{E}[\log W]\right) = \mathbb{V}\mathrm{ar}\left(\log T_\ell - \mathbb{E}[\log T]\right) + \mathbb{V}\mathrm{ar}\left(\mathbb{E}_\ell[\log S]\right) + 2\mathbb{C}\mathrm{ov}\left(\log T_\ell - \mathbb{E}[\log T], \mathbb{E}_\ell[\log S]\right).$$

The first term is the employment-weighted dispersion in TFP. The second term captures the extent to which skills are segmented over space. Finally, the third term represents the extent to which the spatial skill segmentation is correlated with city TFPs. The first two columns of Table F.3 displays the variance decomposition in the model, for all workers and young workers only respectively. The slopes of the blue and orange dotted lines Figure 3a and 3b correspond to

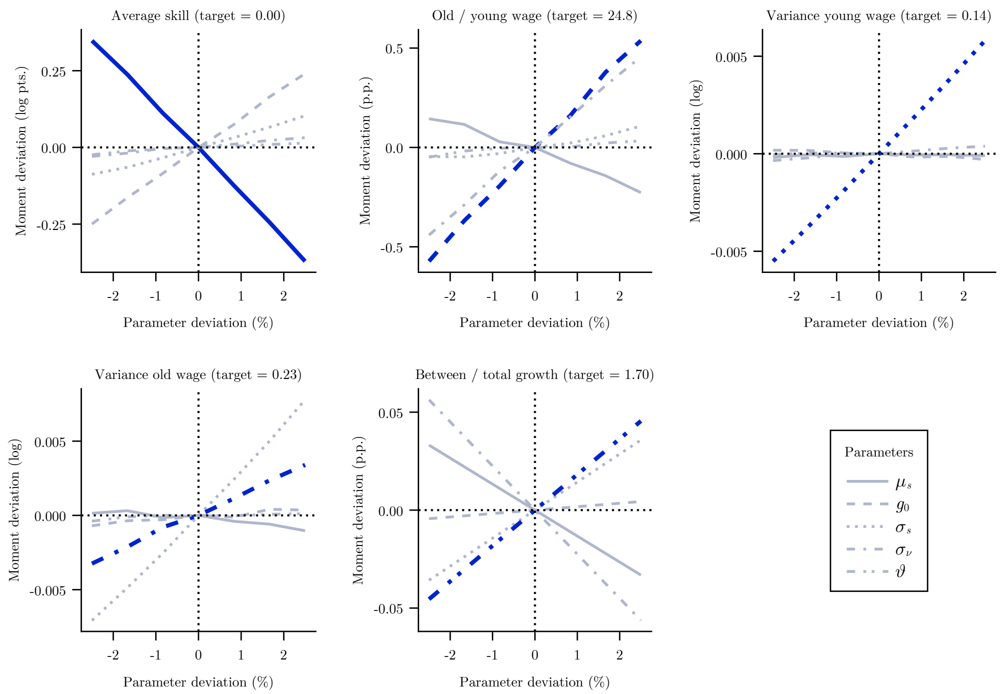
$$\frac{\mathbb{V}\mathrm{ar}\left(\log T_\ell - \mathbb{E}[\log T]\right) + \mathbb{C}\mathrm{ov}\left(\log T_\ell - \mathbb{E}[\log T], \mathbb{E}_\ell[\log S]\right)}{\mathbb{V}\mathrm{ar}\left(\mathbb{E}_\ell[\log W] - \mathbb{E}[\log W]\right)}$$

and

$$\frac{\mathbb{V}\mathrm{ar}\left(\log S\right) + \mathbb{C}\mathrm{ov}\left(\log T_\ell - \mathbb{E}[\log T], \mathbb{E}_\ell[\log S]\right)}{\mathbb{V}\mathrm{ar}\left(\mathbb{E}_\ell[\log W] - \mathbb{E}[\log W]\right)}$$

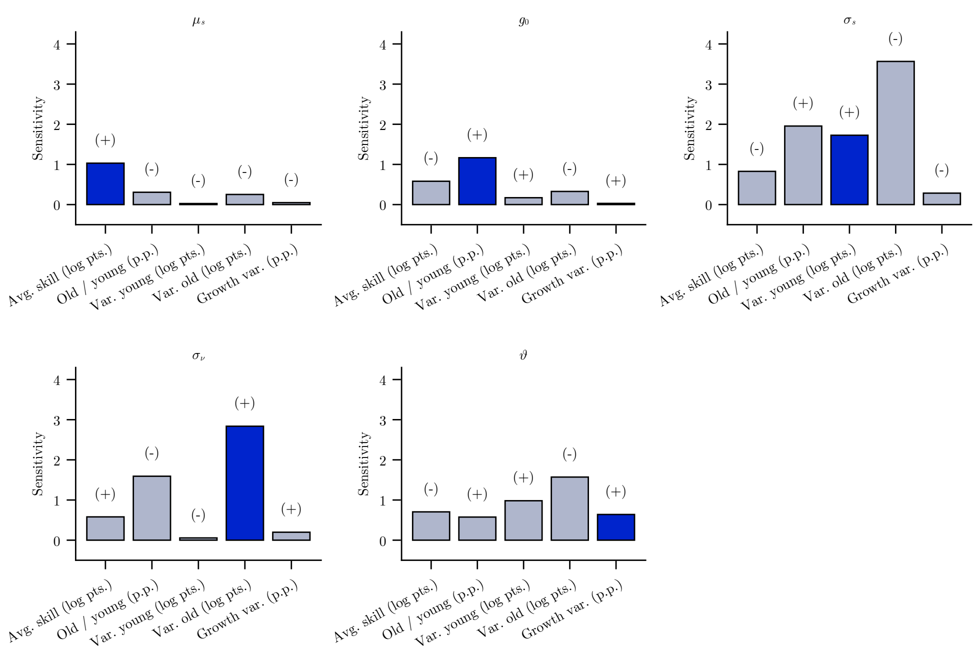
respectively.

Figure F.7: Sensitivity of targeted moments to parameters



Note: each panel is a targeted moment in the indirect inference procedure. The value of the targeted moment is reported in parenthesis. The scale of the y-axis corresponds to the scale of the targeted moment. Each line is a parameter calibrated by the indirect inference procedure. The blue lines highlight the parameter that are heuristically identified by the particular moment of interest. I consider local deviations around the estimated parameters of 2.5% in each direction. The deviations are represented in percentages on the x-axis.

Figure F.8: Sensitivity of parameters to targeted moments



Note: figure reports the sensitivity measure of Andrews et al. (2017) in absolute value. The sign of the sensitivity is reported in parenthesis. Each panel is a parameter calibrated in the indirect inference procedure. Each column is a targeted moment. The blue bar highlights which moment is supposed to identify the parameter of interest.

Table F.3: Between-city variance decomposition in the model and in the data

	Model		Data		
	(1)	(2)	(3)	(4)	(5)
$\mathbb{V}\text{ar}[\log w]$	0.020	0.019	0.020	0.019	0.020
$\mathbb{V}\text{ar}[\log T]$ (%)	19.66	19.66	19.80	19.60	16.80
$\mathbb{V}\text{ar}[\log S]$ (%)	33.54	33.54	35.20	36.30	41.00
$\mathbb{C}\text{ov}[\log T, \log S]$ (%)	46.79	46.79	45.20	44.00	42.20
R^2	1.00	1.00	0.640	0.600	0.930
Skill def.	.	.	Occ. × tenure	Occ. × tenure	Occ. × quintile
Sample	All	Young	All	Young	All

To estimate the TFP and skill gaps in the data, I run

$$\log w_{it} = \alpha_t + \beta_{\ell_{it}} + \gamma_{s_{it}} + u_{it}, \tag{56}$$

and define $\log T_\ell = \hat{\beta}_\ell$ and $\mathbb{E}_\ell[\log S] = \mathbb{E}_\ell[\hat{\gamma}_{s_{it}}]$. I use two skill definitions. First, I set s_{it} as a four-digit occupation fixed effect interacted with three-year bins of occupation tenures. I estimate (56) on the full sample of workers as well as on young workers only to minimize the amount of learning that took place within cities. Second, I set s_{it} as a four-digit occupation fixed effect interacted with the wage quintile within that occupation.

When cities differ in their learning opportunities, worker fixed effects cannot be used to estimate the productivity of a location. Indeed, when $s_{it} = i$ in (56), the city F.E. are estimated off within-worker time variation in wages for workers who switch location,

$$\log w_{it+h} - \log w_{it} = \alpha_{t+h} - \alpha_t + \beta_{\ell_{it+h}} - \beta_{\ell_{it}} + u_{it+h} - u_{it}.$$

That is, the city F.E. are estimated off wage growth variation for movers. Hence, the city F.E. conflates TFP estimates with estimates of the learning technology.

Figure F.9: The spatial distrbution of within-city wage inequality

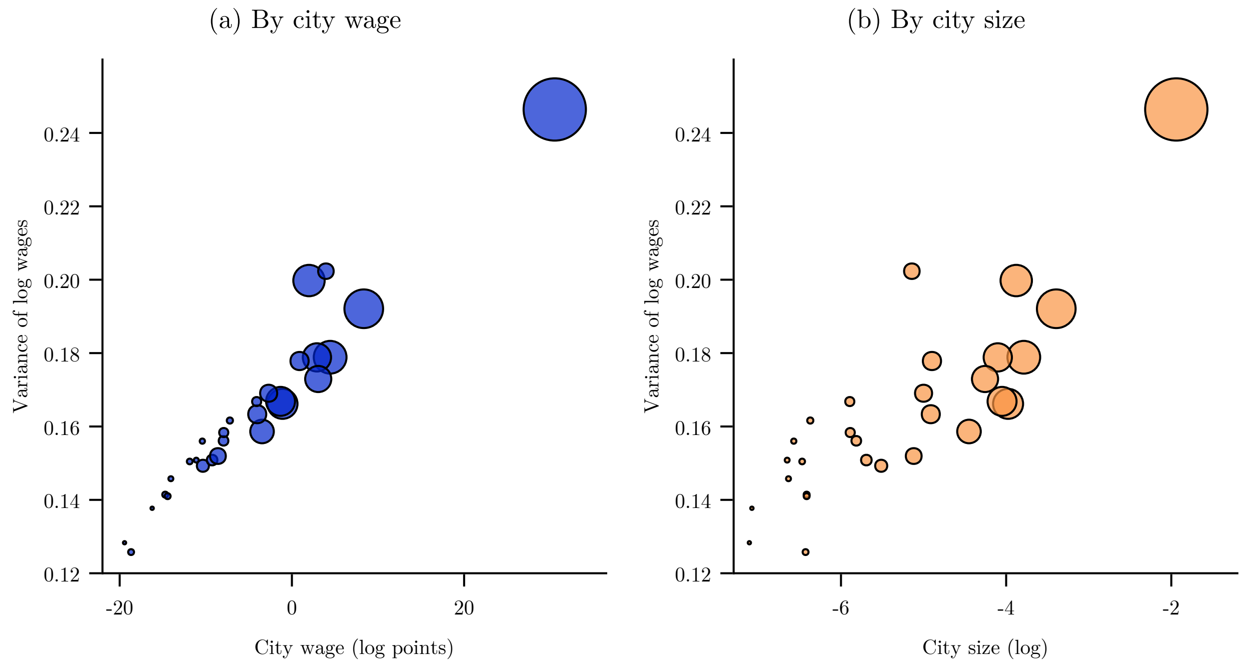
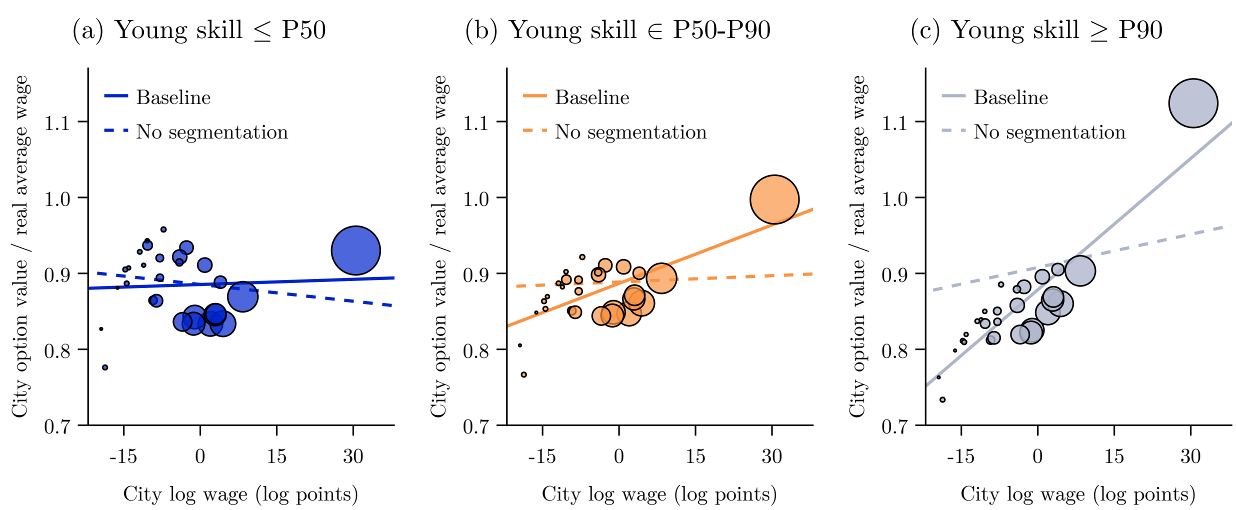


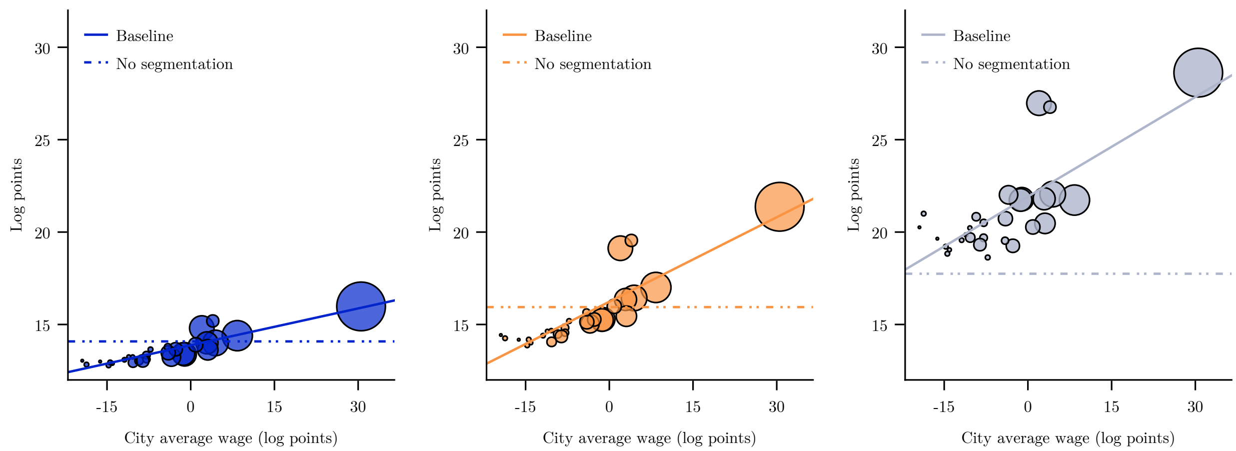
Figure F.10: The option value of cities across the skill distribution



Note: discounted option value of cities over real average wage.

Figure F.11: Skill growth by birthplace and initial skill

(a) Low skill: $\leq$ P50 (b) Middle skill: P50 - P90 (c) High skill: $\geq$ P90



F.5 *Local Interactions, Agglomeration and Human Capital Accumulation: Additional Results*

Figure F.12: This is the first figure

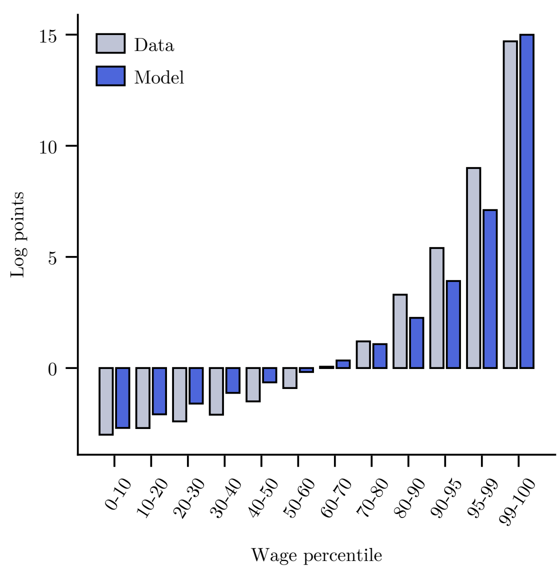
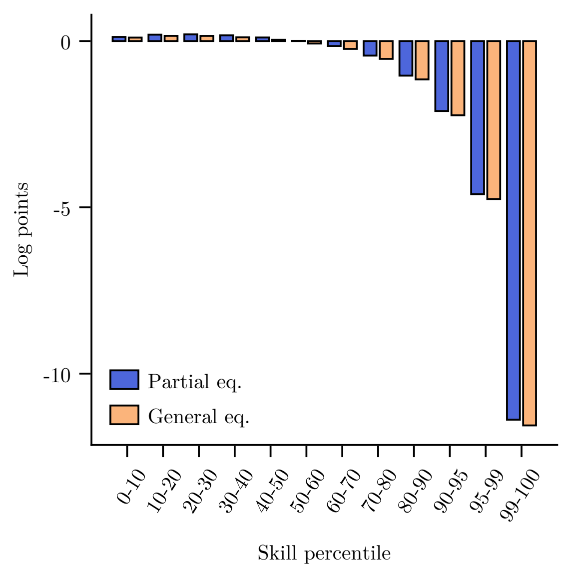


Figure F.13: This is the second figure



G The Consequences of Spatial Policies

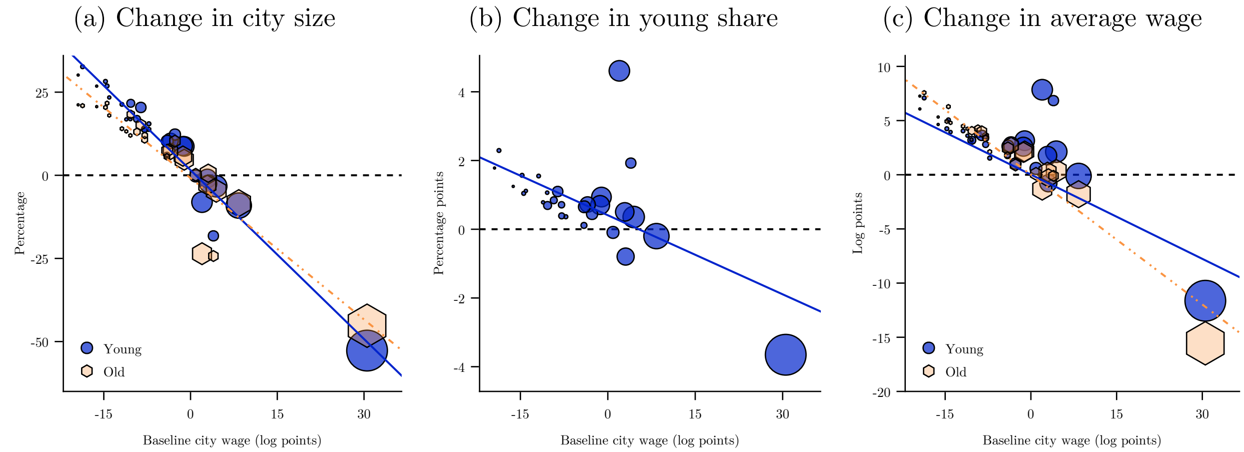
H The Tradeoff Between Human Capital and Inequality

I The Consequences of Spatial Policies

I.1 *Welfare analysis*

I first present how I compute the consumption-equivalent welfare metric of Section 6.3. In the quantitative model presented in Section 2.6, the lifetime utility of a young worker born in l with skill s if their lifetime consumption is

Figure F.14: The agglomeration effect of local interactions by age



multiplied by ζ is

$$\mathcal{V}_l^y(s,\zeta) = \frac{1}{\vartheta}\log\left(\sum_\ell \mathrm{e}^{\vartheta\left(\frac{\left[sT_\ell + \tau_{l\ell}^y(s)\right]\zeta}{P_\ell} + B_\ell^y - \kappa_{l\ell}^y + \beta O_\ell(s,\zeta)\right)}\right),$$

for $\tau_{l\ell}^y(s)$ any possible set of transfers, and[102]

$$O_\ell(s,\zeta) = \frac{1}{\vartheta}\int\int \log\left(\sum_{\ell'} \mathrm{e}^{\vartheta\left(\frac{\left[s^o T_{\ell'} + \tau_{\ell\ell'}^o(s^o)\right]\zeta}{P_{\ell'}} + B_{\ell'}^o - \kappa_{\ell\ell'}^o\right)}\right) dF\left(\frac{s^o}{\gamma(s,s_p)}\right)\pi_\ell(s_p).$$

In the baseline equilibrium, $\tau_{l\ell}^y(s) = \tau_{l\ell}^o(s) = 0$. Let the lifetime utility of young workers in this be denoted by $\mathcal{V}_l^y$. In the counterfactual equilibrium, $\tau_{l\ell}^y(s) = \tau\mathbb{1}\{l \in o \cap \ell \in d\} - t(s)$, where o are the treated origin locations, d are the targeted locations, and $t(s)$ is the within-skill tax that finances the policy. Let the lifetime utility of young workers in this policy equilibrium be denoted by $\mathcal{V}_l^{y\star}(s,\zeta)$. The consumption-equivalent welfare impact of the policy on workers born with skill s in city l, $\zeta_l(s)$, is then measured as $\mathcal{V}_l^y[s,\zeta_l(s)] = \mathcal{V}_l^{y\star}(s,0)$.

The total welfare impact of the policy can be decomposed into several margins. To see this, note that the expected lifetime utility of a young worker be written as

$$\mathcal{V}_l^y(s) = \mathbb{E}\left[\max_\ell\left\{\frac{sT_\ell + \tau_{l\ell}}{P_\ell} - \kappa_{l\ell}^y + \beta O_\ell(s) + \varepsilon_\ell\right\}\right] = \sum_\ell\left(\frac{sT_\ell + \tau_{l\ell}}{P_\ell} - \kappa_{l\ell}^y + \beta O_\ell(s)\right)\frac{n_{l\ell}^y(s)}{m_\ell^y(s)} + \chi_l^y(s).$$

where $\chi_l^y(s) \equiv \mathbb{E}\left[\varepsilon_\ell^y \mid \ell \succ \ell'\,\forall \ell' \neq \ell\right]$ and I have subsumed the amenity value of cities into $\chi_l^y(s)$.[103] The first equality follows from the definition of lifetime utility, and the second line from $n_{l\ell}^y(s)/m_l^y(s) = \Pr[\varepsilon_\ell \geq \max_{\ell' \neq \ell}\varepsilon_{\ell'} + U_{l\ell'}(s) - U_{l\ell}(s)]$. Similarly, the expected option value of city ℓ can be rewritten

$$\begin{aligned} O_\ell(s) &= \int\int \mathbb{E}\left[\max_{\ell'}\left\{\frac{e\gamma(s,s_p)T_{\ell'}}{P_{\ell'}} - \kappa_{\ell\ell'}^o + \varepsilon_{\ell'}\right\}\right]\pi_\ell(s_p)dF(e)ds_p \\ &= \int\left(\sum_{\ell'}\frac{n_{\ell\ell'}^o(s^o)}{m_\ell^o(s^o)}\left(\frac{s^o T_{\ell'}}{P_{\ell'}} - \kappa_{\ell\ell'}^o\right) + \chi_\ell^o(s^o)\right)\omega_\ell(s^o,s)ds^o. \end{aligned}$$

where the second line follows from the change of variable $e\gamma(s,s_p) \to s^o$ and the definition

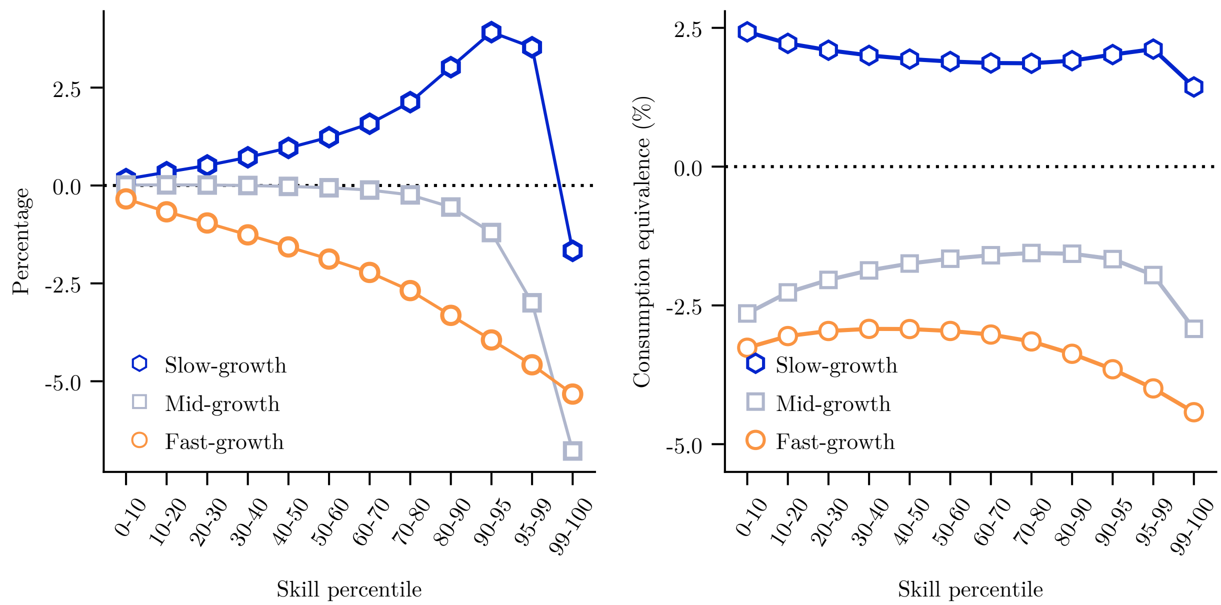
$$\omega_\ell(s^o,s) \equiv \int \frac{\pi_\ell(s_p)}{\gamma(s,s_p)} f\left(\frac{s^o}{\gamma(s,s_p)}\right) ds_p.$$

[102] I use the change of variable $e\gamma(s,s_p) \to s^o$ to derive the option value of city ℓ.

[103] No closed-form solution exists for $\chi_l^y(s)$ but it can be obtained as utility residual.

Figure G.15: The consequences of moving vouchers across workers

(a) Change in expected skill by skill and birthplace (b) Change in welfare by skill and birthplace



Combined with the expression for young workers, we can therefore write lifetime utility as

$$
\begin{aligned}
\mathcal{V}_l^y(s) = & \underbrace{\sum_{\ell}\left(\frac{\tau_{l\ell}-\kappa_{l\ell}^y P_\ell}{P_\ell}\right)\frac{n_{l\ell}^y(s)}{m_\ell^y(s)} - \beta\sum_{\ell}\frac{n_{l\ell}^y(s)}{m_\ell^y(s)}\int \omega_\ell(s^o,s)\left(\sum_{\ell'}\frac{n_{\ell\ell'}^o(s^o)}{m_\ell^o(s^o)}\kappa_{\ell\ell'}^o\right)ds^o}_{\text{Transfers net of lifetime migration costs } \equiv V_l^T(s)} \\
& \underbrace{\sum_{\ell}\left(\frac{sT_\ell}{P_\ell}\right)\frac{n_{l\ell}^y(s)}{m_\ell^y(s)}}_{\text{Present real income} \equiv V_l^I(s)} + \underbrace{\beta\sum_{\ell}\frac{n_{l\ell}^y(s)}{m_\ell^y(s)}\int \omega_\ell(s^o,s)\left(\sum_{\ell'}\frac{n_{\ell\ell'}^o(s^o)}{m_\ell^o(s^o)}\frac{s^o T_{\ell'}}{P_{\ell'}}\right)ds^o}_{\text{Future real income} \equiv V_l^{EI}(s)} + \\
& \underbrace{\chi_l(s) + \beta\sum_{\ell}\frac{n_{l\ell}^y(s)}{m_\ell y(s)}\int \chi_\ell(s^o)\omega_\ell(s^o,s)ds^o}_{\text{Preferences} \equiv V_l^A(s)},
\end{aligned}
$$

Lifetime utility can therefore be decomposed into four terms. First, the value of the transfers net of the lifetime migration costs. Second, the present real income these workers have access to. Third, their expected future real income. And third, their preferences for the location they live in. Hence, $\mathcal{V}_l^y(s) = V_l^A(s) + V_l^T(s) + V_l^I(s) + V_l^{EI}(s)$.

This utility decomposition can be used to decompose the welfare impact of the policy. The consumption-equivalent welfare change can indeed be rewritten

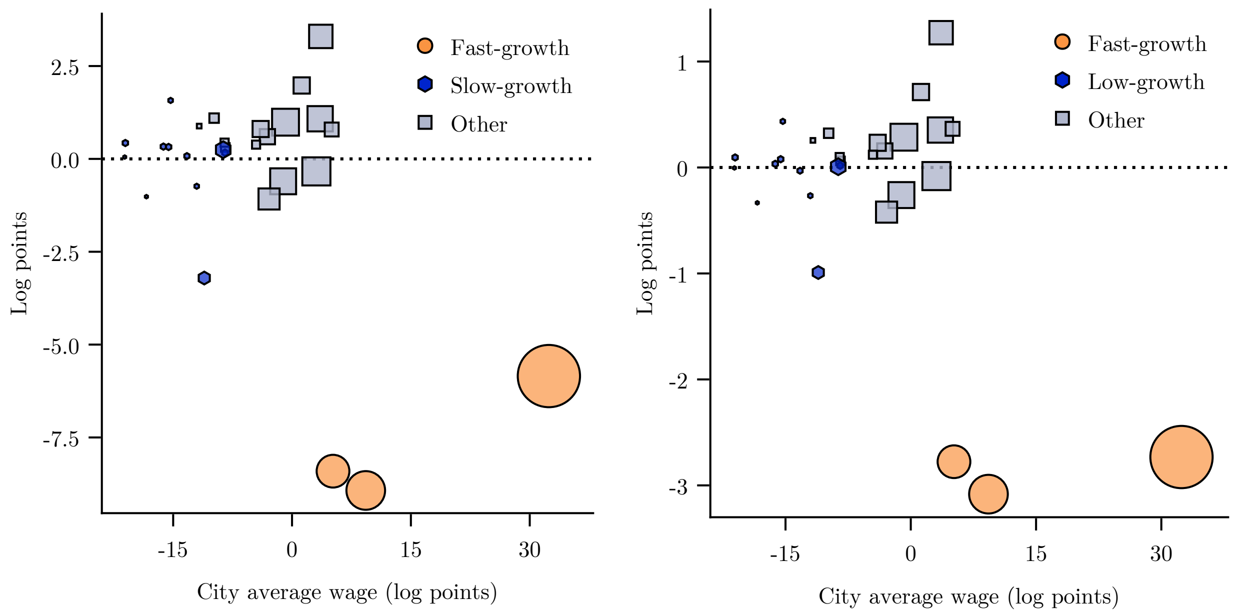
$$
\begin{aligned}
0 = \mathcal{V}_l^y[s,\zeta_l(s)] - \mathcal{V}_l^{y\star}(s) &= \mathcal{V}_l^y[s,\zeta_l(s)] - \mathcal{V}_l^y(s) + \mathcal{V}_l^y(s) - \mathcal{V}_l^{y\star}(s) \\
&= \mathcal{V}_l^y[s,\zeta_l(s)] - \mathcal{V}_l^y(s) + \left(V_l^A(s) - V_l^{A\star}(s)\right) + \left(V_l^T(s) - V_l^{T\star}(s)\right) + \\
&\quad \left(V_l^I(s) - V_l^{I\star}(s)\right) + \left(V_l^{EI}(s) - V_l^{EI\star}(s)\right),
\end{aligned}
$$

where the first equality is the definition of $\zeta_l(s)$, the second is an identity, and the third follows from the decomposition

Figure G.16: The consequences of the voucher policy on local interactions

(a) Change in average skill by workplace

(b) Change in average skill growth by workplace



I just derived. From there, we can define the four consumption-equivalent welfare metric:

$$
\begin{aligned}
\mathcal{V}_l^y[s, \zeta_l^A(s)] &= \Big(V_l^{A\star}(s) - V_l^A(s)\Big) + \mathcal{V}_\ell^y(s), \\
\mathcal{V}_l^y[s, \zeta_l^{A+T}(s)] &= \Big(V_l^{A\star}(s) - V_l^A(s)\Big) + \Big(V_l^{T\star}(s) - V_l^T(s)\Big) + \mathcal{V}_\ell^y(s), \\
\mathcal{V}_l^y[s, \zeta_l^{A+T+I}(s)] &= \Big(V_l^{A\star}(s) - V_l^A(s)\Big) + \Big(V_l^{T\star}(s) - V_l^T(s)\Big) + \Big(V_l^{I\star}(s) - V_l^I(s)\Big) + \mathcal{V}_\ell^y(s), \\
\mathcal{V}_l^y[s, \zeta_l(s)] &= \mathcal{V}_l^{y\star}(s).
\end{aligned}
$$

The term $\zeta_l^A(s)$ measure the (consumption-equivalent) change in welfare coming solely from the idiosyncratic preferences for the locations. The second term $\zeta_l^{A+T}(s)$ adds to the first term the net value of the transfer. The third term $\zeta_l^{A+T+I}(s)$ brings in the value of the present real income opportunities. Finally, $\zeta_l(s)$ adds the expected future lifetime opportunities and corresponds to the total change in welfare. These four metrics are *not* additive but allow to measure the relative importance of each term.

Figure G.17: The consequences of the moving voucher policy across skills and birthplaces