Should I Stay or Should I Grow?

How Cities Affect Learning, Inequality and Aggregate Productivity

Hugo Lhuillier*

Princeton University

March 7, 2024

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 the other workers in their cities. 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, and disproportionately so 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 human capital accumulation 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, and I am particularly grateful to Dávid Nagy for discussing the paper. This paper has also benefited from discussions with Mark Aguiar, Thomas le Barbanchon, Zach Bleemer, Juanma Castro-Vincenzi, Levi Crews, Théo Durandard, Mayara Felix, Cécile Gaubert, Gregor Jarosch, Benny Kleinman, Moritz Lenel, Adrien Matray, Eduardo Morales, Ezra Oberfield, Karthik Sastry, Andreas Schaab, John Sturm, Derek Wenning, and Owen Zidar. 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 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 does their concentration amplify regional disparities? Can policies spur higher productivity in conjunction with lower geographical inequalities?

This chaper offers answers to these questions with four contributions. First, I propose a theory that sheds light on how the concentration of skills in space affects human capital accumulation. Second, I estimate the impact of cities' skill composition on wage growth using French administrative data. Third, I use the estimated model to quantify the extent to which learning opportunities are spatially segregated, and what are the consequences for spatial disparities and aggregate productivity. Fourth, I quantify the tradeoff between spatial inequality and human capital accumulation by evaluating the general equilibrium consequences of a large-scale moving voucher policy.

In the first part of the chaper, I propose a theory of learning across space. 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 other workers in their city. The learning depends on which workers are engaged in the interaction. Importantly, I impose minimal restrictions on the learning technology. 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. Individuals are forward-looking and freely choose where to work to maximize their lifetime utility.

The sorting of workers in cities shapes the spatial distribution of learning opportunities, which in turn may amplify the spatial concentration of skills. Holding constant future income, high-skill workers sort into productive cities as these locations offer relatively higher wages. Yet, their location decisions also reflect where the local learning opportunities are. When skilled workers agglomerate in productive cities, individuals who learn relatively more from them view these locations as attractive learning centers. The stronger the within-skill learning complementarities, the more skilled workers gain over time from working next to other skilled individuals, and the further the learning motive amplifies the spatial concentration of skills.

Spatial inequality feeds into the aggregate stock of human capital. The agglomeration of high-skill next to each other increases the frequency at which they meet while reducing the opportunity for low-skill to learn from them. Whether this segregation of opportunities 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-skill learn relatively more from each other, low-skill 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 this chapter 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. I use the structure of the model to derive estimating equations that identify a first set of parameters. I then calibrate the remaining parameters in the second step to match salient features of the wage distribution in the aggregate, across cities, and across the lifecycle.

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, relatively skilled workers experience faster skill growth holding constant their interactions. 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 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 heterogeneity in wage dynamics across cities and firms. In the most restrictive specification, the skill growth 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 generate quasi-random variations in skill density across space, 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 individuals working there, and disproportionately so if they are skilled.

The estimated returns to local interactions are substantial. The average worker experiences a wage growth 4.5 log points faster over five years when working in Paris (top 10% of the city-growth distribution) compared to when working in Troyes (bottom 10%). These growth gains differ considerably across workers due to the estimated complementarities. A worker in the bottom 10% of the wage distribution sees its wage growth accelerate by 2.8 log points when moving from Troyes to Paris, whereas these dynamic gains are 7.3 log points for individuals in the top 10%. Altogether, the model explains 64% of the between-city wage growth variance.

In the third part of this chapter, I use the estimated model to quantify how much of the spatial differences in human capital accumulation are caused by the segregation of learning opportunities, and what are the implications for spatial disparities and aggregate productivity. To do so, I compare the baseline equilibrium in which cities segment learning interactions with a counterfactual economy in which individuals learn randomly from workers in the entire economy.

I find the spatial segmentation of opportunities, rather than the sorting of fast-learning workers, to explain the vast majority of the learning gaps between cities. This, in turn, engenders spatial learning inequality. Workers born in the 25% poorest French cities have a lifetime skill growth 10% lower than the average French worker. In addition, the concentration of opportunities in a few cities amplifies regional disparities. Productive cities get bigger as workers need to live there to accumulate human capital faster. 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.

The confinement of learning opportunities to productive cities also boosts aggregate human capital accumulation. The stock of human capital is 1.2% higher when learning takes place within cities. Learning complementarities explain the majority of this productivity gain: the average skill of the economy would only be 0.5% higher were all workers to learn equally from skilled workers. The learning gains are concentrated at the high end of the skill distribution due to the concentration of these workers in skill-dense cities and the presence of learning complementarities. In contrast, workers at the bottom of the skill distribution neither gain nor lose from the spatial concentration of learning opportunities.

Combined, the estimated model suggests the existence of a steep tradeoff between spatial inequality and human capital accumulation. In the fourth part of this chapter, I quantify this equity-efficiency tradeoff by evaluating the consequences of a moving voucher policy. Specifically, I study a policy subsidizing every worker born in the 25% poorest French cities to move to the three largest locations.

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 costs, 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 opportunities 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. I conclude that learning complementarities are crucial for the tradeoff between spatial inequality and human capital accumulation.

Related literature This chapter primarily adds to two strands of the literature. The first 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 relaxing land-use regulations in New York and San Francisco spurs aggregate growth.[3]

I contribute to this literature by characterizing the crucial role played by learning complementarities in shaping how cities affect 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).

The second strand of the literature estimates the dynamic returns to interactions between neighbors and coworkers. Across space, Baum-Snow and Pavan (2012), de la Roca and Puga (2017),

[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).

[3] A related literature 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).

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 some of the endogeneity concerns present in the peer effect literature (Angrist, 2014). Second, I find the effect of cities' skill composition on wage growth to be one order of magnitude larger than that of city size. Third, I show that the neighborhood where individuals work shapes their wage growth even after controlling at which establishment and firm they are employed.

This chapter also connects more broadly with several papers in economic geography. First, it adds to the literature that investigates the sources of spatial agglomeration (for a survey, see 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 can provide a motive for sorting even when cities are *ex-ante* similar. I contribute to this literature by quantifying the strength of these agglomeration forces. I find that learning within cities amplifies the skill concentration in space, an empirical pattern documented by Combes et al. (2008), Eeckhout et al. (2014), and Diamond (2016), among others. Second, this chapter 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. Related to these papers, I characterize the optimal spatial policy when spillovers are dynamic.

The rest of this chapter is organized as follows. Section 2 presents the theory and characterizes how learning complementarities shape the tradeoff between spatial inequality and human capital accumulation. Section 3 turns to the French matched employer-employee to estimate these complementarities. Section 4 estimates the rest of the model, and Section 5 quantifies the impact of local interactions on human capital accumulation and agglomeration. Section 6 proceeds to study the consequences of spatial policies on spatial inequality, productivity, and welfare. Section 7 concludes.

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).

[4] See Glaeser and Gottlieb (2008) and Kline and Moretti (2014) for reviews of the estimated consequences of place-based policies on productivity.

2.1 Setup

I model workers that sort into cities. Every period, a unit mass of young workers enters the labor market. Workers live for two periods and discount the future at rate β. Denote their first and second life period as young (y) and old (o). Old workers express no altruism towards future generations. Workers choose in which city to work at any point in time.[5] Cities are indexed by ℓ, and the total number of cities L in the economy is fixed. I study the steady state of this economy and omit time subscripts.

Preferences Individuals consume a single, freely-traded good taken to be the numéraire. I abstract from local consumption for the sake of tractability. I add housing in Section 2.6. Individuals hold linear utility over their consumption. They supply inelastically one unit of labor per period, so that wages and incomes are interchangeable. They do not have access to a saving device.

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 Gumbel 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 the opera's quality more, and I impose no relationship between B_ℓ^y and B_ℓ^o. Local amenities are taken as given. Location preferences are independent across the lifecycle.[6]

Finally, 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 $y + \varepsilon_\ell$.

Human capital Young workers are endowed with an initial skill s_y drawn from the skill distribution N^y. The young skill distribution is a primitive of the model. I suppose that skills are uni-dimensional and that the support of N^y is connected, positive, and bounded, $[\underline{s}_y, \bar{s}_y]$, with $\underline{s}_y \geq 0$ and $\bar{s}_y < \infty$.[7]

Workers learn in their youth. This learning is shaped by the interactions they encounter. If a worker with skill s_y interacts with a partner with skill s_p, they obtain new skill $s_o = e\gamma(s_y, s_p)$. The learning technology γ encapsulates how workers' human capital accumulation depends on their own skill and that of the other workers. In addition, workers experience an idiosyncratic learning shock, e, which captures the other sources of learning (e.g. on-the-job training). These shocks are drawn from the city-independent distribution F with support $\mathbb{R}_+$.[8]

I impose minimal restrictions on the learning technology.

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

---

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

[6] I add migration costs in Section 2.6. Migration costs and persistent location preferences are isomorphic up to welfare accounting.

[7] The estimation of the learning technology (Section 3) is robust to adding worker fixed effects to account for multi-dimensional skill heterogeneity.

[8] 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.

Assumption 1 allows my framework to nest prevalent theories of human capital accumulation. First, it encompasses cases where interactions are irrelevant for learning, $\gamma(s_y, s_p) = g(s_y)$ for some function g. In this case, workers' capacity to learn may still depend on their own skill, e.g. if skilled workers are better at learning-by-doing (Huggett et al., 2011). Second, all workers may learn equally from particular interactions, $\gamma(s_y, s_p) = g(s_y) + h(s_p)$, for h another function. Then, the learning technology does not feature any complementarity. In contrast, the learning complementarities may be stronger within-skill than between-skill. Formally, the learning technology may be supermodular:

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

Supermodularity demands the marginal gains from a skilled interaction to be greater for skilled individuals. This occurs when workers with similar skills are better able to learn from each other (e.g. Jovanovic, 2014). Alternatively, the learning technology may display complementarities that are stronger between than within-skill. Then, the learning technology is submodular,

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

and the marginal gains from skilled interactions are greater for low-skills. This occurs when knowledge diffuses from skilled to relatively less skilled individuals (Lucas and Moll, 2014).

The simple learning technology

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

illustrates well when do within- or between-skill complementarities prevail. While g_1 and g_2 govern the average impact of workers' and partners' skills on learning, the 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 $g_{12} < 0$.

Workers interact with one other worker in their youth. Two assumptions govern this interaction. First, interactions are spatially segmented: they occur between individuals working in the same location. In this section, I also assume for the sake of simplicity that interactions take place between young workers. I relax this assumption in Section 2.6. Second, interactions are random. Hence, while workers cannot target specific interactions, they can choose to locate in a particular city to increase their chance of meeting a particular skill-type. The likelihood of an interaction with a skill s_p in city ℓ is then given by the local density of this skill, $\pi_\ell^y(s_p)$.

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

Production Production takes place within cities. Each city has a representative competitive firm with productivity T_ℓ, which reflects the industries and firms present there. For the equilibrium to be scale invariant, I normalize the dispersion of the idiosyncratic location preferences by the (unweighted) average city TFP, $\vartheta = \theta/\mathbb{E}[T_\ell]$, for θ the scale-invariant dispersion.

The representative firm produces the tradable good using labor as the 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 properties. First, cities' TFP and workers' skills are complements: high-skill workers produce relatively more in productive cities than their low-skill counterparts. 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 isolate the impact of complementarities in the learning process. Appendix C.1 shows that the theoretical predictions are robust to adding production complementarities, and so is the estimation of the learning technology in Section 3.

The labor markets are segmented by cities and skills. The city-specific zero profit condition pins down wages, $w_\ell(s) = sT_\ell$, which inherit the complementarity between TFP and skill.

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

$$\sum_\ell n_\ell^a(s) \leq n^a(s), \quad \forall\, s,\ \forall\, a \in \{y, o\}, \tag{2}$$

where $n_\ell^a(s)$ is the measure of workers with age a and skill s employed in city ℓ, and $n^a(s) = dN^a(s)$ is the aggregate density of skill s for workers with age a. Accordingly, the local density of skill s in city ℓ amongst workers with age a is

$$\pi_\ell^a(s) = \frac{n_\ell^a(s)}{N_\ell^a}, \quad \forall\, s, \forall\, \ell, \forall\, a, \tag{3}$$

for $N_\ell^a = \int n_\ell^a(s) ds$ the mass of workers with age a employed in ℓ.

2.2 The Steady State Equilibrium

Three equations govern the steady state equilibrium.[9]

[9]The young spatial allocation is decoupled from that of the old. As a result, the economy is independent from its initial condition and converges to its steady state in one period.

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_o and their location preferences $\boldsymbol{\varepsilon}$,

$$V^o(s_o, \boldsymbol{\varepsilon}) = \max_{\ell} \{w_\ell(s_o) + \varepsilon_\ell\}, \quad \forall\, s_o, \forall\, \boldsymbol{\varepsilon}.$$

Workers' idiosyncratic preferences ensure that every skill is present in every location. In particular, the number of old workers with skill s employed in city ℓ is given by

$$n^o_\ell(s_o) = n^o(s_o) \left(\frac{\mathrm{e}^{\vartheta\left(w_\ell(s_o) + B^o_\ell\right)}}{\sum_{\ell'} \mathrm{e}^{\vartheta\left(w_{\ell'}(s_o) + B^o_{\ell'}\right)}} \right), \quad \forall\, s, \forall\, \ell. \tag{4}$$

The first term captures the aggregate supply of skill s_o amongst the old and follows from the feasibility condition (2). The second term is the probability that skill s_o chooses to work in city ℓ and is given by the Gumbel-distributed location preferences.

Young workers Young workers solve a dynamic location choice problem,

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

where $\mathcal{V}^o(s_y) \equiv \mathbb{E}[V^o(s_y, \boldsymbol{\varepsilon}') \mid s_y]$ is the expected utility of an old worker with skill s_y prior to observing their idiosyncratic preferences.[10] Young workers' location decision depends on three considerations. The first two are static and are common to old workers. The third consideration is dynamic and reflects where the best learning opportunities are located. The expected future utility of working in city ℓ is indeed given by the weighted average of the gains from particular interactions, $\mathcal{V}^o[e\gamma(s_y, s_p)]$, with weights given by the local likelihood of such interactions, $\pi^y_\ell(s_p)$. A location constitutes an attractive learning center if it provides a large density of individuals from whom workers learn relatively more. When interactions do not shape human capital accumulation, this dynamic consideration disappears.

As for the old, the number of young workers with skill s employed in city ℓ follows from the Gumbel location preferences and the feasibility condition (2),

$$n^y_\ell(s_y) = n^y(s_y) \left(\frac{\mathrm{e}^{\vartheta\left(w_\ell(s_y) + B^y_\ell + \beta \mathbb{E}[\mathcal{V}^o(E\gamma(s_y, S^y_\ell))]\right)}}{\sum_{\ell'} \mathrm{e}^{\vartheta\left(w_{\ell'}(s_y) + B^y_{\ell'} + \beta \mathbb{E}[\mathcal{V}^o(E\gamma(s_y, S^y_{\ell'}))]\right)}} \right), \quad \forall\, s, \forall\, \ell, \tag{5}$$

where $\mathbb{E}[\mathcal{V}^o(E\gamma(s_y, S^y_\ell))] = \int \int \mathcal{V}^o[e\gamma(s_y, s_p)] dF(e) \pi^y_\ell(s_p) ds_p$ is the expected future utility of young workers with skill s_y 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, individuals take the location choice of others 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

[10] Using standard extreme-value algebra, $\mathcal{V}^o(s_y) = c + \log\left(\sum_\ell \mathrm{e}^{\vartheta w_\ell(s_y)}\right)/\vartheta$ for c a constant.

with skill less than s_o is given by

$$N^o(s_o) = \sum_\ell \int\int\int \mathbb{1}\left\{e\gamma(s_y, s_p) \leq s_o\right\} dF(e) n_\ell^y(s_y)\pi_\ell^y(s_p) ds_p ds_y, \quad \forall\, s_o. \tag{6}$$

The spatial allocation of young workers determines the skill distribution of old workers, and thus the aggregate stock of human capital, by shaping the aggregate probability of two skills meeting.

Equilibrium Equations (4) to (6) describe the steady-state spatial equilibrium. Given the spatial distribution of learning opportunities, young workers choose a city where to work. In equilibrium, their expectations must be consistent with the other workers' location choices. The spatial allocation of young workers determines the aggregate skill distribution of old workers, and the spatial allocation of old workers must be feasible.

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

1. *Taking the behavior of other young workers as given, the young spatial allocation satisfies* (5)*;*
2. *Given the old workers' skill distribution, the old spatial allocation satisfies* (4)*;*
3. *The local skill densities are consistent with the location decisions of young workers* (3)*;*
4. *The skill distribution of old workers satisfies* (6)*.*

The steady state equilibrium boils down to an infinitely-dimensional, non-linear fixed point in the local skill densities. The existence of an equilibrium is guaranteed by the idiosyncratic location preferences which coordinate expectations. Furthermore, as standard in spatial models, there exists a unique steady state to the extent that the dispersion forces are larger than the agglomeration forces (Redding and Rossi-Hansberg, 2017). In the present framework, the larger the scale parameter of the taste shock distributions θ^{-1}, the larger the dispersion forces. Meanwhile, the agglomeration forces emanate from the learning interactions, and the smaller the discount factor β, the less workers value learning. Hence, when $\theta\beta$ is small enough, there exists a unique steady state.

Proposition 1 (Existence and uniqueness)**.**
A steady state equilibrium exists. If $\theta\beta$ is small enough, there exists a unique steady state.[11]

I highlight two special types of spatial allocation that may prevail in equilibrium. In the first, the skill densities are the same in every city, and therefore so are the learning opportunities. I refer to this allocation as the symmetric allocation.

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

Alternatively, productive locations may feature a higher density of skilled workers. In this case,

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

the allocation satisfies stochastic positive assortative matching (SPAM).[12]

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: $\pi^a_\ell \succ_{FOSD} \pi^a_{\ell'} \iff T_\ell > T_{\ell'}$.

In the absence of restrictions on the learning technology, the steady state equilibrium depends on the entire skill distribution of every city. To gain analytical tractability, I approximate the equilibrium around the point where cities have homogeneous TFP, $\log T_\ell \approx \log \bar{T}$ for all ℓ and some $\bar{T}$. When cities share a common TFP, the production complementarities vanish. Workers no longer have an exogenous motive to sort across cities, and there exists an equilibrium with a symmetric allocation (Proposition A.1). Proposition 1 guarantees that this symmetric equilibrium is unique when $\theta\beta$ is small enough, which I assume to be the case for the rest of this section.[13] Approximating the equilibrium around the homogeneous TFP point thus simplifies how to keep track of the local skill densities.

Sections 2.3 and 2.4 characterize the approximated equilibrium. Throughout, $\bar{x}$ refers to the value of x in the homogeneous TFP, symmetric equilibrium, and I use capital letters to denote random variables. For instance, S^y is the aggregate skill distribution of young workers, and S^y_ℓ the skill distribution of young workers in city ℓ. I return to the global solution from Section 2.5 onward.

2.3 Local Interactions as a Source of Agglomeration

I first show how the spatial segmentation of learning opportunities affects where individuals choose to work. I start by characterizing the location decisions of old workers to contrast them with those of the young. When between-city differences in TFP are small, (4) implies that the number of old workers with skill s_o employed in city ℓ is to a first order given by[14]

$$n^o_\ell(s_o) \approx \underbrace{\frac{\bar{n}^o(s_o)}{L}}_{\text{Symmetric}} + \underbrace{\theta\left(\frac{\bar{n}^o(s_o)}{L}\right)\log\left(\frac{T_\ell}{\bar{T}}\right)s_o,}_{\text{Sorting for working}} \quad \forall\, s_o, \forall\, \ell, \tag{7}$$

where $\bar{n}^o(s_o)$ is the density of old workers with skill s_o 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, the local wages are relatively higher, which attracts more workers. Furthermore, the between-city wage gaps are increasing in workers' skills due to the production complementarities. High-skill workers are therefore

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

[13]Corollary A.2 also provides a bound on $\theta\beta$ that guarantee uniqueness of the symmetric equilibrium under the assumption that $\gamma(s_y, s_p) = \mathcal{F}(s_y) + \mathcal{G}(s_y)\mathcal{H}(s_p)$ for $\mathcal{F}$, $\mathcal{G}$ and $\mathcal{H}$ some functions. For a framework where local interaction alone generate asymmetric equilibria, see Davis and Dingel (2019).

[14]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.

relatively more present in productive cities, and the old allocation satisfies SPAM.[15] The dispersion in idiosyncratic location preferences, θ, acts as the local labor supply elasticity and modulates the amount of sorting.

The location decisions of young workers differ from those of the old in that they also care about local learning opportunities. Specifically, from (5), the measure of young workers with skill s_y in city ℓ is to a first order given by

$$n_\ell^y(s_y) \approx \underbrace{\frac{n^y(s_y)}{L}}_{\text{Symmetric}} + \underbrace{\theta\left(\frac{n^y(s_y)}{L}\right)\log\left(\frac{T_\ell}{\bar{T}}\right)s_y}_{\text{Sorting for working}} + \underbrace{\theta^2\beta\left(\frac{n^y(s_y)}{L}\right)\log\left(\frac{T_\ell}{\bar{T}}\right)\Theta(s_y)}_{\text{Sorting for learning}}, \quad \forall\, s_y, \forall\, \ell.$$

The third, new term captures the learning considerations. Specifically, the function $\Theta(s_y)$ summarizes how much young workers with skill s_y value the learning opportunities of relatively productive cities. When $\Theta(s_y) > 0$, these opportunities are relatively attractive, therefore increasing the supply of s_y in these locations. When $\Theta'(s_y) > 0$, skilled workers disproportionately value productive cities' learning opportunities, intensifying their concentration in space. I define the total willingness to sort of young workers as the sum of the sorting for working and learning motives, $\eta(s_y) \equiv s_y + \theta\beta\Theta(s_y)$.

How much young workers value the learning opportunities of productive cities depends on who works there and how much individuals learn from them,

$$\Theta(s_y) = \int \underbrace{\left(\gamma\left(s_y, s_p\right) - \mathbb{E}\left[\gamma\left(s_y, S^y\right)\right]\right)}_{\substack{\text{Relative learning from}\\ s_p \text{ interaction}}} \underbrace{\left(s_p + \theta\beta\Theta(s_p)\right)}_{\substack{\text{Willingness to}\\ \text{sort of } s_p}} dN^y(s_p), \quad \forall s_y. \tag{8}$$

In particular, the probability to meet a skill s_p in a productive city depends on their willingness to work there, and therefore on how they themselves value the local interactions. Formally, $\Theta(s_y)$ is the solution to a infinitely-dimensional fixed point. The first-order approximation renders this fixed point linear, which simplifies its characterization.[16] I show in Lemma A.1 that this fixed point has a unique solution when the dispersion forces are large enough.

The willingness to sort of young workers would solely rely on the production complementarities were workers not to learn from their neighbors. The spatial allocation of young workers would then also display stochastic positive assortative matching. The learning motive could *a priori* offset this technological force if low-skill were to value the learning opportunities of productive cities disproportionately. However, a higher density of low-skill workers in productive cities would contradict the empirical evidence on sorting patterns found in this chapter and others (e.g. Combes et al., 2008; Card et al., 2023). I therefore focus my attention to the part of the parameter space that generates a SPAM allocation for young workers.

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

[15] Proposition A.3 shows that SPAM always prevail for old workers for any spatial distribution of TFP.

[16] These infinite-dimensional linear fixed points are referred to as Fredholm integral equation of the second kind. Similar integral equations can be found in Allen and Arkolakis (2014).

$\eta'(s_y) \geq 0$ *for all* s_y.

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 $\theta\beta g_{12}\mathbb{V}\text{ar}[S^y] \geq -1$, which imposes a bound on how submodular the learning technology can be.

How does the spatial concentration of skills affect the location decisions of workers when they learn from the other workers in their city? In a SPAM equilibrium, individuals who work in productive cities are more likely to interact with skilled workers. Productive cities are therefore attractive learning centers to workers who disproportionally learn from high-skill. Formally, $\gamma(s_y, \cdot)$ increasing implies $\Theta(s_y) > 0$, and the spatial segmentation of learning opportunities increases the number of young workers with skill s_y in productive cities. Two consequences follow.

First, the learning motive may constitute a source of agglomeration that endogenously increases the size of productive cities. Specifically, if every individual learns relatively more from skilled workers, the size of productive cities is larger in the presence of spatially segmented learning interactions as individuals agglomerate there to learn from the local workers.

Second, the learning motive may increase the spatial concentration of skilled workers as they agglomerate in a few cities to learn from each other. To a first order, the average skill of young workers employed in city ℓ is

$$\mathbb{E}[S_\ell^y] \approx \underbrace{\mathbb{E}[S^y]}_{\text{Symmetric}} + \underbrace{\theta \log\left(\frac{T_\ell}{\bar{T}}\right)\mathbb{V}\text{ar}[S^y]}_{\text{Sorting for working}} + \underbrace{\theta^2\beta\log\left(\frac{T_\ell}{\bar{T}}\right)\mathbb{C}\text{ov}[S^y, \Theta(S^y)]}_{\text{Sorting for learning}}, \quad \forall\, \ell, \tag{9}$$

where $\mathbb{V}\text{ar}[S^y]$ is the skill variance of young workers and $\mathbb{C}\text{ov}[S^y, \Theta(S^y)]$ is the covariance between workers' skill and how much they value the learning opportunities of productive cities. When the complementarities within-skill dominate those between-skill, young skilled workers have greater incentives to interact with their peers. They therefore place a higher value on the learning opportunities present in productive cities, and the local density of high-skill rises. Greater between-city wage inequality follows (equation (38) in Appendix A.6). In contrast, if the between-skill complementarities prevail, the learning motive reduces the concentration of skilled workers in productive cities.

I summarize the consequences of the spatial segmentation of learning opportunities on the geographic distribution of economic activity in Proposition 2.

Proposition 2 (Spatial inequality)**.**
When $\log T_\ell \approx \log \bar{T}$ *for all* ℓ, $\gamma(s_y, \cdot)$ *increasing (decreasing) implies* $\Theta(s_y) > 0$ *(*$\Theta(s_y) < 0$*), and* γ *supermodular (submodular) implies* $\Theta' > 0$ *(*$\Theta' < 0$*). As a result:*

1. *If* $\gamma(s_y, \cdot)$ *is increasing for all* s_y*, (un)productive cities are (smaller) larger when interactions are spatially segmented; the converse holds when* $\gamma(s_y, \cdot)$ *is decreasing;*
2. *If* γ *is supermodular, the average skill of young workers in (un)productive cities is (smaller) larger when interactions are spatially segmented, which leads to greater between-city wage variance amongst young workers; the converse holds when* γ *is submodular.*

2.4 The Human Capital Consequences of Local Interactions

Between-city differences in skill growth arise endogenously from the sorting of young workers. To a first order, the expected skill growth of young individuals starting their career with skill s_y in city ℓ_y is

$$\mathbb{E}\left[\frac{S^o}{s_y} \mid s_y, \ell_y\right] \approx \mathbb{E}\left[\frac{\bar{S}^o}{s_y} \mid s_y\right] + \theta \log\left(\frac{T_{\ell_y}}{\bar{T}}\right) \mathbb{C}\mathrm{ov}\left[\frac{\gamma(s_y, S^y)}{s_y}, \eta(S^y)\right], \quad \forall\, s_y.$$

In a SPAM equilibrium, interactions with skilled workers are disproportionally present in productive cities. As a result, conditional on learning more from high-skill, individuals employed in productive cities experience faster skill growth.

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

$$\mathbb{E}[S^o] = \sum_{\ell} \int\int\int e\gamma(s_y, s_p) n^y_\ell(s_y) \pi^y_\ell(s_p) dF(e) ds_p ds_y. \tag{10}$$

This equilibrium stock of human capital can be contrasted with the average skill of old workers that prevails under the symmetric allocation:

$$\mathbb{E}[\bar{S}^o] = \int\int\int e\gamma(s_y, s_p) n^y(s_y) n^y(s_p) dF(e) ds_p ds_y. \tag{11}$$

Equation (11) coincides with the stock of human capital were interactions not segmented by cities since the spatial allocation of workers is then irrelevant for human capital accumulation. Comparing (10) and (11), cities affect aggregate productivity by shaping the aggregate probability of complementarity meetings. Absent learning complementarities, i.e. $\gamma(s_y, s_p) = g(s_y) + h(s_p)$, cities indeed do not shape aggregate productivity, $\mathbb{E}[S^o] = \mathbb{E}[\bar{S}^o]$. In that case, the faster human capital accumulation of individuals working in productive cities is exactly offsetted by the slower learning of workers employed in the other locations.

More generally, cities may amplify or dampen the stock of human capital. To a second order, (10) implies that the average skill of old workers is[17]

$$\mathbb{E}[S^o] \approx \mathbb{E}[\bar{S}^o] + \theta^2 \mathbb{V}\mathrm{ar}[\log T_\ell]\Omega, \tag{12}$$

where $\mathbb{V}\mathrm{ar}[\log T_\ell]$ is the variance of cities' TFP. The constant Ω fully encapsulates the impact of cities on the stock of human capital. To a second order, this stock is larger when interactions are spatially segmented if and only if $\Omega > 0$. This aggregate effect is given by

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

[17] Cities do not have a first-order effect on the average skill of old workers as the between-city differences in learning necessarily cancel each other.

The first term captures how much skill s_y learns from s_p. The second and third terms together determine the aggregate probability that they meet, which depends on both skills' willingness to work in productive cities.

Learning complementarities shape the impact of cities on aggregate human capital accumulation. In a SPAM equilibrium, cities increase the chance that workers with similar skills interact. When within-skill complementarities are strong, the marginal gains from skilled interactions increase in workers' skills. In such cases, the agglomeration of skilled individuals in productive locations boosts 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 segregation of skilled workers in productive cities prevents low-skill workers from learning from them frequently, which lowers aggregate productivity.[18]

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

As a corollary to Proposition 3, learning complementarities determine the existence of a tradeoff between spatial inequality and human capital accumulation.[19] In a SPAM equilibrium, larger between-city wage inequality is necessarily associated with a stronger concentration of skilled workers in productive cities and, therefore, a greater segregation of learning opportunities across space. When within-skill learning complementarities prevail, an equity-efficiency tradeoff arises: greater spatial wage inequality boosts the stock of human capital. Such a tradeoff is not present when the between-skill learning complementarities are stronger. Then, a policy that reduces spatial wage inequality also increases the stock of human capital.

Corollary 4 (Equity-efficiency tradeoff)**.**
Suppose that $\log T_\ell \approx \log \bar{T}$ *for all* ℓ*. In equilibrium, a greater between-city wage variance amongst young workers leads to a greater (smaller) stock of human capital if* γ *is supermodular (submodular).*

Learning complementarities therefore shape the tradeoff between spatial inequality and aggregate productivity. They also determine whether there is too much or too little spatial skill concentration.

2.5 Optimal Policies

To study whether the spatial allocation of workers is efficient, I solve for the allocation chosen by a utilitarian planner. I assume the planner has homogeneous Pareto weights across skills and cities to focus on the learning externalities rather than redistributive concerns. I further suppose that the planner cannot condition its allocation on workers' idiosyncratic location preferences, either because

[18] Appendix A.10 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.

[19] Formally, from (12), the equilibrium average skill of old workers can be rewritten $\mathbb{E}[S^o] - \mathbb{E}[\bar{S}^o] \propto \Omega \mathbb{V}\mathrm{ar}[\mathbb{E}(\log W_\ell^y)]$, where $\mathbb{V}\mathrm{ar}[\mathbb{E}(\log W_\ell^y)]$ is the between-city wage variance. To a second-order, the equilibrium is always located on this equity-efficiency loci. In particular, the impact of any policy that affects the spatial allocation of workers on aggregate productivity is given by this technological frontier.

they are unobserved or because it is not politically feasible. 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, $\{n^{y\star}_{t\ell}, n^{o\star}_{t\ell}\}_{t,\ell}$, together with a sequence of consumption allocation across cities and skills, $\{c^{y\star}_{t\ell}, c^{o\star}_{t\ell}\}_{t,\ell}$, to maximize the discounted sum of each cohort's expected lifetime utility,[20]

$$\sum_{\ell} \int \mathbb{E}[V_0^{o\star}(s_o, \boldsymbol{\varepsilon}) \mid s] n^{o\star}_{0\ell}(s_o) ds_o + \sum_{t>0} \beta^t \sum_{\ell} \int \mathbb{E}[V_t^{y\star}(s_y, \boldsymbol{\varepsilon}) \mid s] n^{y\star}_{t\ell}(s_y) ds_y. \tag{13}$$

The function $V_0^{o\star}(s_o, \boldsymbol{\varepsilon})$ is the utility of old workers with skill s_o and preferences $\boldsymbol{\varepsilon}$ in the initial period under the planner's allocation. Likewise, $V_t^{y\star}(s_y, \boldsymbol{\varepsilon})$ 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 aggregate output. 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 must compensate them for their foregone location preferences.[21] I solve first for the global, non-linearized planner's allocation. I focus on the steady state allocation, referred to as the (constrained) optimal allocation.

Old workers do not affect the learning of young workers, and therefore do not exert externalities. As a result, the laissez-faire and the planner allocation coincide.[22]

In contrast, the allocation of young workers need not be efficient. Young workers do not internalize the impact of their location choice on others' learning. As a result, workers' incentives to live in specific cities may fall short of what is socially desirable. Specifically, there exists a wedge between the optimal and laissez-faire consumption allocation

$$c^{y\star}_{\ell}(s_y) - c^{y}_{\ell}(s_y) = \tau^{\star}_{\ell}(s_y) - \sum_{\ell'} \left(\frac{n^{y\star}_{\ell'}(s_y)}{n^y(s_y)} \right) \tau^{\star}_{\ell'}(s_y), \quad \forall\, s_y, \forall\, \ell. \tag{14}$$

When this wedge is positive, the laissez-faire equilibrium features too few workers with skill s_y 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_y in city ℓ,

$$\tau^{\star}_{\ell}(s_y) = \beta \int \int \left(\mathcal{V}^o[e\gamma(\sigma, s_y)] - \mathbb{E}[\mathcal{V}^o(e\gamma(\sigma, S^{y\star}_{\ell}))] \right) dF(e) \pi^{y\star}_{\ell}(\sigma) d\sigma, \quad \forall\, s_y, \forall\, \ell, \tag{15}$$

[20]By the property of the Gumbel distribution, the expected utility of workers who choose to work in city ℓ is identical to the unconditional expected utility, $\mathbb{E}[V_0^{o\star}(s, \boldsymbol{\varepsilon}) \mid s, \ell \succ \ell' \,\forall \ell' \neq \ell] = \mathbb{E}[V_0^{o\star}(s, \boldsymbol{\varepsilon}) \mid s]$.

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

[22]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); Donald et al. (2023); Mongey and Waugh (2024), the two sets 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.

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 σ. These local welfare gains are positive if workers in ℓ learn relatively more from skill s_y than from the average local worker. Hence, the optimal spatial allocation features a greater number of skill s_y in city ℓ to the extent that the workers employed in that city learn disproportionately from that skill.

Equation (14) implies that the laissez-faire equilibrium is inefficient as long as workers learn from others. Characterizing the spatial misallocation amounts to solving for the non-linear, infinitely dimensional fixed point consisting of (14), (15), and the local labor supplies. As for the laissez-faire equilibrium, I simplify this fixed point by studying the optimal allocation when the spatial TFP gaps are small, $\log T_\ell \approx \log \bar{T}$. Appendix B.3 presents the first-order approximation.

I find learning spillovers within cities to generate spatial misallocation through two channels. First, productive cities may be too large. Conditional on learning from skilled individuals, 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. Holding aggregate productivity constant, the planner thus desires to decentralize learning centers.[23] When spatial TFP gaps are small, this is achieved by reallocating workers from high- to low-TFP locations independently of their skills.

Second, the laissez-faire equilibrium may feature too much or too little spatial skill concentration. When within-skill complementarities dominate those between-skill, skilled workers benefit disproportionately their peers. Internalizing these spillovers, the planner increases the frequency of interactions between skilled workers by augmenting their concentration in productive cities. In contrast, relatively low-skill workers experience too few skilled interactions in the laissez-faire 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 stock of human capital.

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

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

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

Which policy instruments can decentralize the optimal allocation? Equation (15) implies that efficiency is restored through place-by-skill transfers, $t_\ell^y(s)$. These transfers complement workers' wages, and the income of young workers with skill s employed in city ℓ becomes $I_\ell^y(s) = w_\ell(s) + t_\ell^y(s)$. When the spatial TFP gaps are small, the transfers can further be decomposed into two components.

[23] In a different setting, Waldfogel (2003) coined this externality the "preference externality".

First, a city-specific tax that 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).

Taking stock, the spatial segmentation of learning opportunities may shape geographic disparities inequality, the stock of human capital accumulation, and may call for place-based policies. To evaluate quantitatively their consequences, I add several extensions to the framework that allow me to bring it to the data.

2.6 Quantitative model

I add two key elements to the model, housing and migration costs, 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. I solve for the global, non-linearized solution of the model numerically. Appendix D formally describes the quantitative model and the numerical algorithm.

Housing In addition to the freely traded numéraire, households consume a local 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.[24] 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.

Migration costs While workers remain free to change location whenever they please, they incur a migration cost upon moving. I let these migration costs vary across the lifecycle, e.g. if young workers face tighter borrowing constraints or if old workers have accumulated more social capital. I assume that workers are born in the location where their parents lived when they were young; that is, the probability of being born in city ℓ is given by N_ℓ^y.[25] A young worker born in city ℓ that moves to city ℓ' incurs a utility cost $\kappa_{\ell\ell'}^y$.[26] 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 symmetric migration costs, $\kappa_{\ell\ell'}^a = \kappa_{\ell'\ell}^a$ 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 birthplace. In particular, expected utilities are no longer equalized across space.

Learning I let young workers learn from young and old workers. Furthermore, I assume that four elasticities govern the learning technology

$$\log \gamma(s, s_p) = g_0 + g_1 \log s + g_2 \log s_p + g_{12} \log s \log s_p. \tag{16}$$

[24]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).

[25]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 $N_\ell^y \approx N_\ell^o$.

[26]Alternatively, these migration costs can represent auto-correlation in workers' location preferences. The two interpretations are isomorphic from a positive standpoint.

The first parameter, g_0, determines the average skill growth. The elasticity g_1 governs how workers' own skill shapes their human capital accumulation holding constant the skill of their learning partner. In particular, $g_1 > 1$ implies that skilled workers learn relatively faster. The two other parameters, g_2 and g_{12}, dictate the returns to interactions. Workers learn relatively more from skilled partners on average when $g_2 > 0$, and particularly so if they are themselves skilled when $g_{12} > 0$.

While parsimonious, the learning technology (16) has two advantages. First, it allows for technology that are neither supermodular nor submodular, and therefore for which the theory is ambiguous.[27] Second, Figure E.8 shows that it is not possible to reject empirically its log-linear structure.

Parametrization 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.[28]

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. Instead, I rely on numerical methods. In Section 5, I compare the estimated framework with a counterfactual economy in which interactions are not spatially segmented. The numerical results align with the qualitative predictions of Proposition 2 and 3.

3 Estimating the Returns to Local Interactions

As highlighted in the previous section, learning complementarities are crucial in determining how the spatial segmentation of learning opportunities affects productivity. In this section, I present evidence on the role of cities' skill composition in shaping workers' wage growth. I argue that these evidence can be used to estimate the learning complementarities.

3.1 Data

I estimate the model on French matched employer-employee data (DADS).[29] This dataset comes in two formats. The first format is a 4% representative panel that tracks the entire history of individuals in the labor market (DADS panel).[30] 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.

[27] The technology (16) indeed differs from (1) in that it is expressed in logs.

[28] 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.

[29] 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).

[30] 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.

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

I apply the same sample restrictions on both datasets. I focus my analysis on full-time employed workers between 25 and 55 years old. Workers employed in the public sector in France have their wages determined nationally 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 their large fraction of public servants. 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.[31] Appendix E.1 provides further details on the construction of the sample.

Over time, wages grow for two reasons. First, workers experience idiosyncratic changes. Second, the aggregate wage fluctuates.[32] Aggregate trajectories are absent from the framework laid down in Section 2. To abstract from them empirically, I normalize the average log wage to zero every year. These 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 (Figure E.2a), guaranteeing that I do not conflate faster lifecycle wage growth with faster aggregate growth.

I define a city as a commuting zone (CZ). A commuting zone is a statistical area defined by the French statistical agency (INSEE). It consists of a collection of contingent municipalities clustered together to reduce the commuting flows across them. There are 297 such areas in metropolitan France, which span the whole territory.[33] Commuting zones are the natural geographical unit when thinking of local labor markets. In particular, their relatively large size (40,000 employed workers on average) is well-suited to capture all the local interactions that a worker may experience. I use the terms commuting zones and cities interchangeably for the rest of the chapter.

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 that also span the French metropolitan territory. However, they are smaller than commuting zones: in 2019, France was split into 34,839 municipalities with an average size of 1,800 inhabitants and an average area of 15 square kilometers.[34] 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

[31] 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 for their skills. I focus on workers paid above the minimum wage by truncating the left-tail of the wage distribution at the 5%.

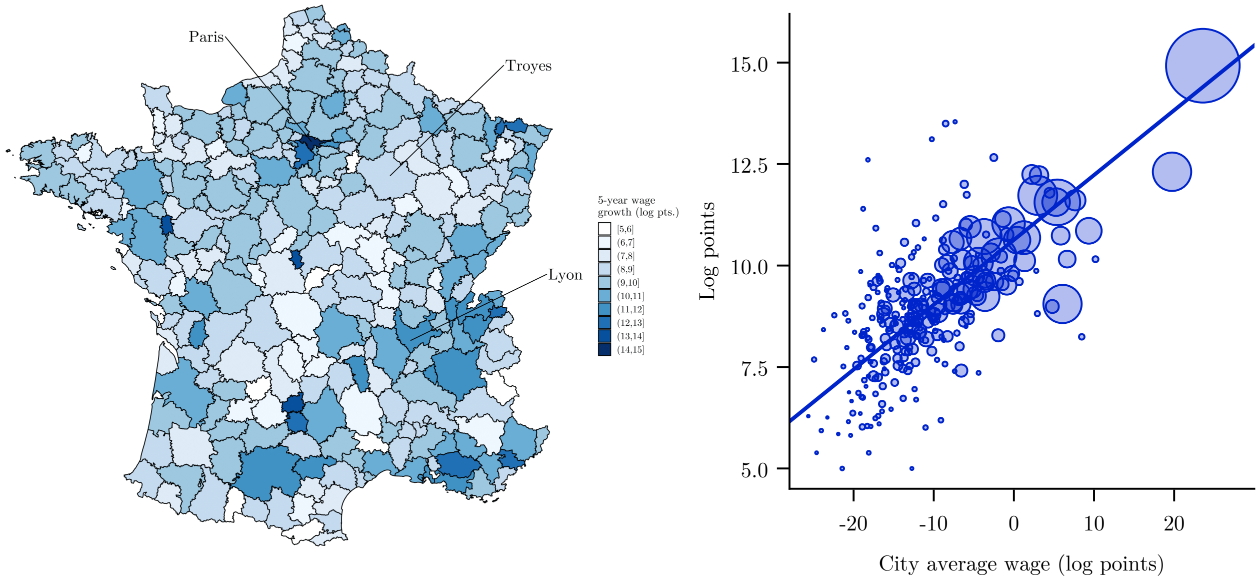
[32] Mechanically, the log wage growth of worker i can be decomposed as $\omega_{it+1} - \omega_{it} = \tilde{\omega}_{it+1} - \tilde{\omega}_{it} + \mathbb{E}[\omega_{it+1}] - \mathbb{E}[\omega_{it}]$, for $\omega_{it} = \log w_{it}$ and $\tilde{\omega}_{it} \equiv \omega_{it} - \mathbb{E}[\omega_{it}]$.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

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

[34] 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.

Figure 1: The geography of wage growth

(a) Within-worker wage growth by CZ (b) Within-worker wage growth by CZ wage



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}.

and Lyon to highlight how granular these spatial units are. I restrict the analysis to metropolitan municipalities with more than 50 employed workers to reduce local measurement errors. I am left with 12,320 municipalities, representing 99% of the French workforce.

3.2 The Geography of Wage Growth

Fast lifecycle wage growth is very spatially concentrated in France. Figure 1a plots the average five-year wage growth at the worker-level by commuting zones.[35] On average, workers under 40 years old experience a wage growth of 10.6 log points every five years. 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 9 log points.

Fast wage growth is concentrated in cities where wages are also relatively high. Figure 1b displays the average wage growth at the city level against the city average wage. In the cross-section, an increase of the city wage by one standard deviation is associated with a wage growth increase by 2 log points at the five-year horizon.[36] I now argue how this pattern can be used to identify learning complementarities.

[35] Consistent with the model, these averages are computed on "young" workers under 40 years old.

[36] This relationship is not mechanically caused by high-wage cities diverging away from the other places (Figure E.2) or by housing prices (Figure E.2a).

3.3 Identification

Given the learning technology (16), 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

$$\log s_{it+1} = g_0 + g_1 \log s_{it} + g_2 \log s_{p_{it}} + g_{12} \log s_{it} \log s_{p_{it}} + \log e_{it+1}, \tag{17}$$

where e_{it+1} is the idiosyncratic learning shock experienced by i. These shocks represent, for instance, enrollments in training programs, returns to on-the-job tenure, or strokes of genius. They are assumed to be identically and independently distributed across space, workers and partners. Through (17), 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 (17) poses two empirical challenges: how to measure skills, and how to observe learning interactions. I use the structure of the model to solve these challenges.

First, in the theory, skills and wages are related through $w_{it} = s_{it} T_{\ell_{it}}$, for $T_{\ell_{it}}$ the TFP of the city where i works. Accordingly, I measure skills as wages deflated by local TFPs, which I obtain through the structural estimation of the model (Section 4).[37]

Second, very few datasets track workers' interactions.[38]. However, when interactions are random within cities, the skill density of the location where i works constitutes a valid proxy for their interactions. Specifically, substituting $s_{p_{it}}$ in (17) for the average skill in the location where i is employed, $\mathbb{E}_{\ell_{it}}[\log s_{jt}]$, returns

$$\log s_{it+1} = g_0 + g_1 \log s_{it} + g_2 \mathbb{E}_{\ell_{it}}[\log s_{jt}] + g_{12} \log s_{it} \mathbb{E}_{\ell_{it}}[\log s_{jt}] + \log e_{it+1} + \nu_{it}. \tag{18}$$

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

The learning technology can therefore be estimated from a local projection of future skills on current skills and the average skill of the workers employed in the same city,

$$\log s_{it+1} = \alpha_t + \beta \log s_{it} + \gamma \overline{\log s}_{\ell_{it}} + \delta \log s_{it} \overline{\log s}_{\ell_{it}} + u_{it}. \tag{19}$$

In (19), α_t is a year fixed effect that controls for aggregate trends in wages, and $\overline{\log s_\ell}$ is the average log skill in city ℓ. I measure future skills at the five-year horizon to smooth out transitory wage shocks while keeping sufficient statistical power.[39] I compute local average skills on all the workers in the sample, but, consistently with the model, I only include young workers under 40 in the regression.

[37] The structural estimation of cities' TFP relies on estimates of the learning technology. To solve this fixed point, I iterate on (19) and the structural estimation of the model until convergence. In practice, I estimate a small TFP dispersion, so that $\log s_{it} \approx \log w_{it}$.

[38] For recent papers aimed at measuring face-to-face interactions, see Atkin et al. (2022) and Emanuel et al. (2023).

[39] Section 4.2 describes how the 5-year estimates produced by (19) are turned into structural 15-year estimates.

While (19) is effectively a linear-in-mean model, it does not suffer from the reflection problem of Manski (1993). The dynamic nature of the underlying learning process indeed rules out the simultaneity issue at the heart of the reflection problem: the average skill of a location is shaped by the interactions that take place there, but those must have happened in the past to have en effect on the skill composition. This argument is formally laid down in Section E.3. The two assumptions essential to the identification are summarized in the following proposition.

Proposition 6 (Identification)**.**
Suppose that

1. *Interactions are random within cities,* $(s_{it}, s_{p_{it}}) \mid \ell_{it} \perp \mathbb{E}_{\ell_{it}}[\log s_{jt}]$*,*
2. *The idiosyncratic learning shocks are i.i.d. across workers and cities,* $e_{it} \perp (s_{it}, \mathbb{E}_{\ell_{it}}[\log s_{jt}])$*.*

Then, (19) *produces unbiased estimates of the learning technology:* $\hat{\beta} = g_1$*,* $\hat{\gamma} = g_2$ *and* $\hat{\delta} = g_{12}$*.*

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 differs 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' wages and the wage of the city where they work to estimate learning complementarities.

3.4 Endogeneity concerns

The assumptions underlying Proposition 6 may not hold empirically. Skilled workers may be able to target interactions with skilled partners. The learning complementarities estimated by (19) would then be biased upward, $|\delta| > |g_{12}|$. To investigate how large this bias is, I estimate (19) separately for small and large cities and more or less unequal locations. To the extent that skilled workers can better target skilled interactions in larger or more unequal cities, the estimated complementarity should be higher in those places.

Meanwhile, the distribution of idiosyncratic learning shocks may vary across workers and cities. Young workers may learn faster for reasons other than local interactions while earning lower wages, $e_{it} \not\perp s_{it}$. Young workers may also agglomerate in large productive cities to enjoy the local amenities, $e_{it} \not\perp \mathbb{E}_{\ell_{it}}[\log s_{jt}]$. I alleviate this second set of endogeneity concerns in three ways.

First, I control for as many sources of idiosyncratic learning shocks as my data allows. I include age fixed effects to capture 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, occupation, industry, and city level. I include a dummy for employer switching to account for its effect on wage growth. Productive firms may have more frequent training programs while offering higher wages. In addition, these training programs may target particular occupations. I include firm-by-occupation fixed effects to account for these firm-specific determinants of human capital accumulation.[40] Finally, establishments may experience long-lasting productivity shocks that are passed through to wages.

[40]The firm-occupation fixed effects also allow for production complementarities across occupations at the firm level.

I account for those by controlling for the growth in average wages at the establishment level. Stacking these controls in the vector $\boldsymbol{X}_{it}$, the exogeneity assumption in Proposition 6 becomes $e_{it} \perp (s_{it}, \mathbb{E}_{\ell_{it}}[\log s_{jt}]) \mid \boldsymbol{X}_{it}$.

Second, I introduce a set of granular spatial fixed effects. The distribution of idiosyncratic learning shocks may vary across cities in ways that are complex to control for. Wealthier cities may for instance provide better public goods that influence workers' wage growth. Controlling for unobserved differences in wage growth across cities poses an identification challenge since the skill growth elasticities in (19) are also estimated from between-city variations. I slightly depart from the theory to solve this challenge and quantify how neighborhoods' skill composition shapes human capital accumulation.[41] Specifically, I estimate the augmented local projection

$$\log s_{it+1} = \alpha_{t\ell_{it}n^r_{it}} + \beta \log s_{it} + \gamma \overline{\log s}_{n^w_{it}} + \delta \log s_{it} \overline{\log s}_{n^w_{it}} + \boldsymbol{\Psi}' \boldsymbol{X}_{it} + u_{it}. \tag{20}$$

In (20), n^r_{it} and n^w_{it} are the neighborhoods where worker i resides and works. The parameters $\alpha_{t\ell n^r}$ are year by city of employment by neighborhood of residence fixed effects, which soak unobserved wage growth heterogeneity across cities (e.g. better universities) and across residence (e.g. better transportation network).[42] Meanwhile, $\overline{\log s}_{n^w}$ is the average wage of the workers employed in neighborhood n^w, and the skill growth elasticities in (20) are estimated by comparing individuals who live in the same neighborhood and work at the same firm and occupation but in different neighborhoods within the same city.[43]

Third and last, I rely on an instrumental variable strategy to generate quasi-random variations in skill density across neighborhoods. Individuals working in the same location may be exposed to common conditions that affect their present and future wages (Angrist, 2014). To address this endogeneity concern common to the peer effect literature, I instrument the average skill of neighborhoods by the change in their white-collar employment share in the 1990s,

$$Z_n = \frac{\#\text{white-collar}_{n,2000}}{\#\text{workers}_{n,2000}} - \frac{\#\text{white-collar}_{n,1993}}{\#\text{workers}_{n,1993}},$$

where Z_n is neighborhood n's instrument. I define white-collar workers through their occupations. Intuitively, the instrument uses implicit shocks in the 1990s that permanently reallocated high-skill workers across neighborhoods within cities. The exclusion restrictions are that these shocks were orthogonal to the neighborhoods' initial conditions and have faded away twenty years later.

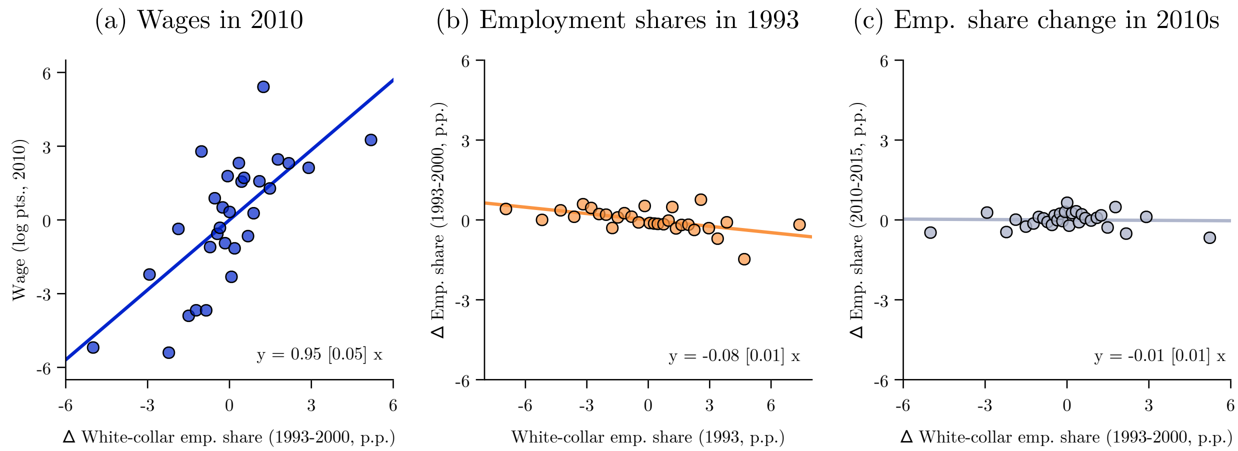
Figure 2 argues that the instrument satisfies the relevance and exclusion restrictions. Figure 2a displays the average wage of neighborhoods in the 2010s against the instrument. The two are strongly positively correlated, suggesting the relevance of the instrument: an increase by one percentage point

[41] An alternative would be to include city fixed effect in (19) and rely on temporal variations in city wages. Yet, these changes are very likely to be correlated with local productivity shocks, and the estimated skill growth elasticities would then conflate faster human capital accumulation and local aggregate growth.

[42] By comparing workers who live in the same neighborhood and work in the same firm, these fixed effects partially reduce workers' unobserved heterogeneity. Table E.1 further introduces worker fixed effects to soak the remaining unobserved heterogeneity.

[43] I control in addition for the distance between the neighborhood of residence and work.

Figure 2: The change in neighborhoods' white-collar employment share (1993-2000)



Note: panel (a) plots the average wage of neighborhoods in 2010 (y-axis) against the change in their white-collar employment share between 1993 and 2000 (x-axis). Panel (b) plots the change in neighborhoods' white-collar employment share between 1993 and 2000 (y-axis) against their employment share in 1993 (y-axis). Panel (c) plots the change in neighborhoods' white-collar employment share between 2010 and 2019 (y-axis) against the change in employment share between 1993 and 2000. All plots are binned scatter plots with 30 bins. Consistently with (20), all variables are first residualized by a city fixed effect and thus normalized to zero. The regression lines are estimated in the full sample of neighborhoods.

of the local share of white collar in the 1990s is associated with 0.95% increase in the average wage of that neighborhood in the 2010s. Figure 2b argues that the underlying shocks were uncorrelated with the neighborhoods' initial conditions by showing that the change in the employment share of white collar in the 1990s is barely correlated with their initial employment share. Finally, Figure 2c presents suggestive evidence that the shocks in the 1990s are no longer correlated with present changes in the skill composition of neighborhoods by plotting the absence of correlation between the change in the employment share of collars in the 1990s and that in the 2010s.

By isolating random variations in skill density caused by past *changes* in occupation composition, the instrument estimates the effect of neighborhoods' skill composition on wage growth net of time-invariant local characteristics. Over twenty years, changes in neighborhoods' skill composition may affect local characteristics other than the quality of the learning interactions (e.g. the physical stock of capital), and indirectly through those wage growth. My identification strategy cannot separate the "interaction" channel from these other dynamic forces. However, to the extent that both channels capture the causal impact of the local skill composition on workers' human capital accumulation, I view them as isomorphic.

How can the elasticities estimated by (20) be interpreted? The parameters obtained from the local projections (19) and (20) can be interpreted as reduced-form elasticities. In that case, γ measures the average treatment effect of the local skill composition on skill growth, and δ estimates the heterogeneity in treatment effects across the skill distribution. Interpreting those parameters structurally is challenging since interactions are modeled at the city rather than the neighborhood level. Nevertheless, to the extent that some interactions are happening within neighborhoods, I view the elasticities estimated by (20) as informative of the null hypotheses $g_2 = 0$ and $g_{12} = 0$.

Table 1: The returns to local interactions

Dep. variable: 5-year skill	(1) OLS	(2) OLS	(3) OLS	(4) OLS	(5) OLS	(6) 2SLS
Skill (g_1)	1.010	1.010	1.010	0.940	0.935	0.935
	(0.003)	(0.003)	(0.003)	(0.002)	(0.002)	(0.003)
City skill (g_2)	0.107	0.095	0.111	0.081		
	(0.015)	(0.012)	(0.005)	(0.008)		
Skill × city skill (g_{12})	0.108	0.102	0.104	0.089		
	(0.032)	(0.007)	(0.009)	(0.013)		
Log city size		0.002				
		(0.001)				
Neighborhood skill					0.032	0.053
					(0.008)	(0.021)
Skill × neighborhood skill					0.071	0.077
					(0.013)	(0.023)
Controls						
Worker-level	·	·	✓	✓	✓	✓
Establishment growth	·	·	·	✓	✓	✓
Distance work-residence	·	·	·	·	✓	✓
Fixed effects						
Year	✓	✓	✓	✓	·	·
Occupation × firm	·	·	·	✓	✓	✓
Year × residence × city	·	·	·	·	✓	✓
Obs.	971,401	971,401	971,401	974,401	971,401	974,401
F-stat.	·	·	·	·	·	53,539

Standard errors clustered at the city level for columns 1 to 4 and at the neighborhood level for columns 5 and 6. Sample includes all workers between 25 and 40 year old. Skills measured as wages deflated by city TFP. Worker-level controls include age, job, occupation, and industry tenure fixed effects, and a dummy for employer switching. Occupations and industry measured at the one-digit level. Establishment growth computed as the growth in average wages, $\mathbb{E}_e[w_{t+1}]/\mathbb{E}_e[w_t]$ for e the establishment. Column 7 instruments for the average neighborhood skill by the change in the white-collar employment share between 1993 and 2000.

3.5 Results

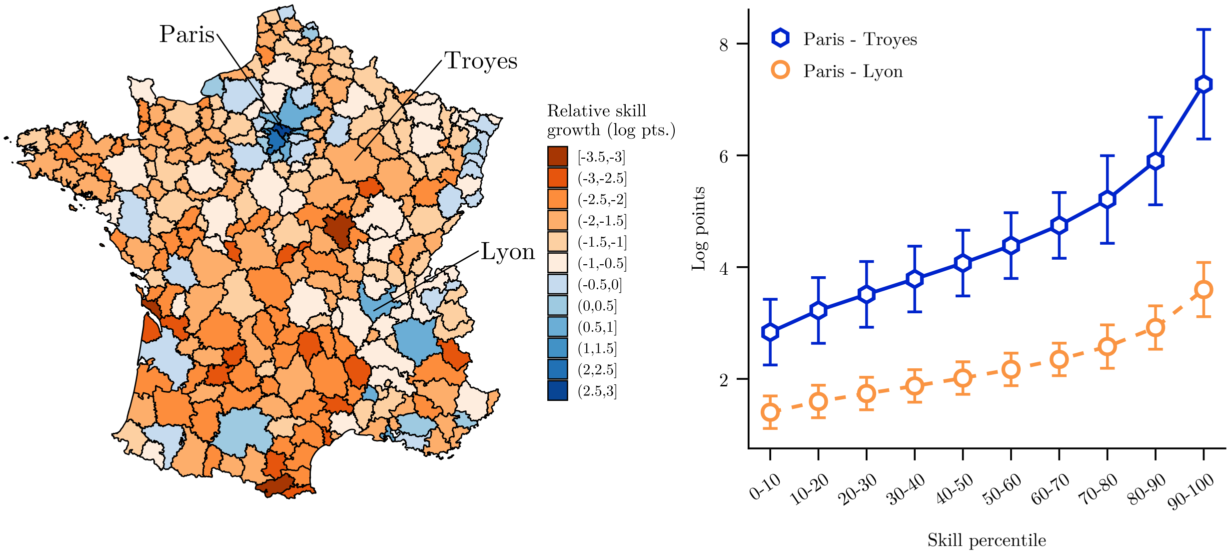
Table 1 presents the estimated skill growth elasticities. Standard errors are clustered at the same level as the spatial skill variations. The first columns presents the estimates produced by the model-consistent local projection (Proposition 6). The three elasticities are estimated to be positive and significant at the 0.1%.

First, high-skill workers enjoy faster skill growth ($\hat{\beta} > 1$): a skill increase of one standard deviation boosts skill growth by 0.4 log points over five years for workers in the average city.[44] Workers starting their career with lower skills do not catch up to high-skill individuals – although

[44]The average skill is normalized to zero, and therefore so is the skill of the average French city. Hence, $\beta - 1$ represents the marginal effect of increasing workers' skill on skill growth for workers in the average French city.

Figure 3: Predicted skill growth

(a) Average predicted skill growth by CZ (b) Predicted skill growth by CZ and initial skill



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

this effect alone is quantitatively small.[45]

Second, individuals employed in skill dense cities experience faster growth ($\hat{\gamma} > 0$). An increase by one standard deviation of the city skill implies an additional growth of 1.6 log points for the average worker. Through the lens of the model, this implies that workers learn on average more from skilled individuals.

Third and last, high-skill individuals working in skill dense cities see their skill grow faster than their low-skill counterparts in the same location ($\hat{\delta} > 0$). An increase by one standard deviation of the city average skill implies an additional growth of 1 log point for workers at the 10th percentile of the skill distribution, compared to an increase of 2.6 log points for workers at the 90th percentile. From (19), this implies that the learning technology displays strong within-skill complementarities: skilled workers learn relatively more from skilled partners than their low-skill counterparts.

The estimated returns to local interactions are substantial. Figure 3a displays the predicted relative skill growth by commuting zones for a worker with the average skill. Migrating from Troyes to Paris triggers a skill growth increase of 4.5 log points per five years for the average French worker, while moving from Lyon to Paris implies skill growth gains of 2.2 log points. Altogether, the predicted skill growth explains 64% of the between-city wage growth variance.

The between-city differences in growth are very heterogeneous across workers due to the estimated

[45]The estimated wage process is therefore non-stationary. This result rests 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.

learning complementarities. Figure 3b plots the difference in predicted growth between Paris, Lyon and Troyes across the skill distribution. All workers, both low- and high-skill, experience faster growth when employed in Paris relative to the other two cities. High-skill workers benefit however more over time from working in high-skill locations. For instance, the skill growth gap between Paris and Troyes is 2.8 log points for workers in the bottom 10% of the skill distribution. It is 7.3 log points for workers in the top 10%. The same pattern holds between Paris and Lyon, although the gaps are smaller as the two cities have a more similar skill composition.

Robustness The remaining columns of Table 1 provide robustness exercises to address the endogeneity concerns mentioned in Section 3.4.[46] Column 2 investigates the role of large cities in shaping the estimated skill growth elasticities. A vast literature, started by de la Roca and Puga (2017), finds that larger cities boost workers' wage growth. In the second column, I thus control for city size. In accordance with the previous literature, I also find that workers in larger cities experience faster growth. However, the effect of city size is one order of magnitude smaller than the effect of skill composition: an increase in city size by one standard deviation implies a 0.4 log points increase in skill growth for the average worker.[47] The difference in magnitude justifies why I abstract from city size in the quantitative model.[48]

The third column introduces the worker-level controls, and the fourth adds the occupation-by-firm fixed effects and the control for establishment wage growth. The parameters shaping the returns to local interactions, g_2 and g_{12}, are smaller than in the baseline. However, the two sets of estimates are not statistically different at the 5%. The same pattern holds when instrumenting workers' wage by their past wage to net out transitory shocks (Table E.1, column 2), or adding worker fixed effect to soak unobserved heterogeneity in abilities to learn (Table E.1, column 7). I conclude that wage growth heterogeneity across workers and firms does not bias the learning technology estimates.

The fifth column adds the neighborhood of residence by city fixed effects to control for unobserved spatial heterogeneity in wage growth.[49] The last column brings in the instrument variable. The F-statistic for the weak IV test is well above the conventional thresholds (Stock and Yogo, 2002; Andrews et al., 2019). The OLS and 2SLS estimates are not statistically different at the 5%. Furthermore, the two elasticities governing how workers' skill growth depends on the skill composition of where they work, γ and δ, remain positive. Comparing workers employed in the same city, in the same firm, at the same occupation, and living in the same neighborhood, individuals working in the relatively skill-dense neighborhood experience faster skill growth – and disproportionately so if these workers are initially relatively skilled.

In summary, Table 1 documents that the skill composition of where individuals work shapes their

[46]In all of these exercises, I keep the estimate of cities' TFP constant to their baseline values.

[47]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 (19) under the restriction that $\beta = 1$, $\gamma = 0$, and average skills are replaced by city size.

[48]I also show in Table E.1 (column 1) that the estimated elasticities are not driven by a few large cities, and in Figures E.6 and E.7 that the elasticities are constant across the city size and inequality distributions.

[49]Table E.1 (column 3) further adds city by industry by occupation fixed effects to control for spatial heterogeneity in wage dynamics that differ across occupations and industries and local production complementarities.

human capital accumulation.[50] In addition, the estimates of the learning technology are robust to controlling for numerous degrees of heterogeneity in wage growth across workers and firms. Finally, allowing for spatial heterogeneity in wage dynamics across cities, I still find that individuals learn relatively more from skilled workers, in particular if they are themselves skilled. I now turn to estimating the remaining parameters of the quantitative model.

4 Model Estimation

The remainder of the structural estimation is split into three parts. In the first part, I define the model's geography and time horizon, and accordingly externally calibrate the discount factor. In the second part, I use the structure of the model to derive estimating equations that identify the first group of parameters without the need for simulation. In the third part, I calibrate the remaining parameters. The estimated parameters are presented in Table F.1.

4.1 Definitions

I define a location in the model as a commuting zone. Solving the model for the 297 commuting zones is computationally challenging as it requires to solve for the entire skill distribution within each city. Instead, I create 30 city types and restrict all cities within the same type to be homogeneous. Importantly, there remains 297 cities in the model, and learning interactions happen within cities. I define the first 10 types as the 10 largest French cities. To construct the remaining 20 types, I divide France into four geographic districts (North-East, North-West, South-East and South-West). I then subdivide each district into five types using a k-mean algorithm that clusters cities based on variables that proxy for the geographic primitives of the model.[51] Appendix F.1 contains details on the variables used in the algorithm, and Figure F.1 plots the resulting city types.

Regarding the model's time horizon, the model features two lifecycle periods spanning workers' career.[52] 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 years 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. The model does not provide an explicit moment to estimate the discount factor. I therefore calibrate it externally, take a yearly discount factor of 0.975, and set $\beta = 0.975^{15}$.

[50] Appendix E.4 also shows that the skill growth elasticities are increasing in the time horizon considered (Figure E.4), are positive even after controlling for the skill composition of the establishment and firm where workers are employed (Table E.1, column 4), and are positive for all industries although with substantive heterogeneity (Figure E.9).

[51] While this classification is not without loss of generality, it is likely to have little impact on the counterfactuals since cities within each type are rather homogeneous. With five types, the clustering algorithm captures 83% of the total sum of square residuals (Figure F.2). In addition, clustering cities avoids the over-fitting problem present in spatial granular settings (Dingel and Tintelnot, 2020).

[52] Adding more periods would make the model closer to the learning process estimated in Table 1. However, it would also increase its computation complexity without much effects on the counterfactuals.

4.2 Model Inversion

In the second step, I invert the model to estimate the parameters dictating the housing supply and demand, the migration costs, and the the learning technology as described in Section 3. The estimating equations are derived in Appendix F.2.

Housing From the Cobb-Douglas preferences, α is the housing expenditure share. I 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 relate housing prices to local expenditures through

$$\log p_\ell = \left(\frac{1}{1+\delta}\right) \log\left(\frac{1-\alpha}{\mathcal{H}}\right) + \left(\frac{1}{1+\delta}\right) \log Y_\ell, \quad \forall\, \ell, \tag{21}$$

where p_ℓ and Y_ℓ are the housing rental price and total expenditures in city ℓ. I obtain p_ℓ from the "Carte des Loyers" (Rental Map), which estimates the average monthly rent per meter square at the municipal level.[53] Meanwhile, the expenditures in city ℓ are given by the payments to the labor inputs, $Y_\ell = N_\ell \mathbb{E}[W_\ell]$, where N_ℓ and $\mathbb{E}[W_\ell]$ are the employment share and average wage of city ℓ that I compute from the cross-sectional matched employer-employee dataset. To match the frequency of the rent data, I set wages at the monthly frequency in euros of 2018. Given $\{p_\ell, N_\ell, w_\ell\}_{\ell=1}^L$ and α, the housing supply parameters, $\mathcal{H}$ and δ, can be recovered from (21) by OLS. Figure F.5c compares the rent prices in the data with those in the model. While simple, the model captures 71% of the between-city rent variations.

Migration costs The migration costs are recovered non-parametrically from the observed migration flows. Using workers' optimal location decisions and the assumption of symmetric costs, the model predicts that the cost to move from city l to city ℓ for workers with age $a \in \{y, o\}$ is[54]

$$\mathrm{e}^{-2\vartheta\kappa^a_{l\ell}} = \int \left(\frac{n^a_{l\ell}(s)}{n^a_{ll}(s)} \bigg/ \frac{n^a_{\ell\ell}(s)}{n^a_{\ell l}(s)}\right) dN^a(s), \quad \forall\, l, \forall\, \ell. \tag{22}$$

In equation (22), $n^a_{l\ell}(s)$ is the number of workers with age a and skill s who migrates from l to ℓ. The cost to move between these two cities is thus identified from the size of the bilateral migration flows relative to the number of individuals who do not move. Measuring the right-hand side of (22) requires observing workers' skills and migration decisions. I proxy workers' skills with their wage quartile.[55] Consistently with the model, I measure migration decisions differently for young and old workers. For young workers, I compute $n^y_{l\ell}(s)$ as the number of workers under 40 born in city l and

[53] These rents are residualized and correspond to the rent paid for a typical 40m^2 apartment in France. I compute p_ℓ at the city-level by taking the population weighted average of the municipal rents.

[54] These migration costs can alternatively capture correlated location preferences – up to welfare considerations. In this case, (22) captures the autocorrelation in location preferences.

[55] 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 TFP variance explains less than 2% of the wage variance. Second, (22) 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.

currently living in city ℓ with skill s. For old workers, I compute $n^o_{l\ell}(s)$ 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. Both flows are computed from the long-panel.

Figure F.3 plots the estimated migration costs by city of origin. Empirically, 45% of young workers live in cities different from their birthplace, whereas 20% of workers change locations between their youth and their old age. Through the lens of the model, these patterns are rationalized with higher migration costs for old workers. The average migration cost for young workers is 1,255€, which corresponds to 43% of their income. The average migration costs for old workers is 1,594€ (44% of their earnings).[56]

The model predicts substantial heterogeneity in migration probabilities across workers. In particular, skilled workers born in productive cities are relatively more likely to work there, whereas skilled workers born in low-TFP locations are more likely to move out. Appendix F.3 provides more details on the migration heterogeneity.

Learning technology The learning technology parameters g_1, g_2 and g_{12} are estimated according to Proposition 6. 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 estimate (19) for three age bins: 25 to 30, 30 to 35 and 35 to 40 (Figure E.5). I then set $g_1 - 1$, g_2 and g_{12} as the sum of their respective estimated age-specific skill growth elasticities.

4.3 Indirect Inference

The remaining 95 parameters govern workers' productivity and location preferences. I calibrate them jointly to match salient features of the wage and city-size distributions. I provide first arguments that justify the identification strategy. I then show numerically how the chosen moments affect the calibration of the model.

City productivity I set the productivity of each city to match its average wage in the data. In the model, the average wage of city ℓ is given by

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

where $\int s\pi_\ell(s)ds$ is the average skill of the workers employed in ℓ. The average wage on the left-hand side can be measured from the matched employer-employee dataset. Holding constant the skill distribution, a greater TFP implies higher local wages. However, the average skill in city ℓ is not observed, and it is an equilibrium object that depends on the spatial distribution of TFPs. Equation (23) therefore cannot be directly inverted to recover T_ℓ. Nevertheless, given $\{T_\ell\}_{\ell=1}^L$ and the other parameters, the spatial distribution of skills can be recovered in equilibrium from the

[56] The estimated migration costs are in utils. The reported estimates are converted to monetary unit through $\kappa^a_{l\ell} P_\ell$. 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.

model. Equation (23) then defines a fixed point over $\{T_\ell\}_{\ell=1}^L$ that can be solved for numerically to obtain cities' TFP without taking a stance on how to measure skills empirically.

Worker-level productivity Four parameters govern the distribution of worker-level productivity. The first two, μ_y and σ_y, are the mean and the standard deviation of the skill distribution of young workers. The level of workers' skills and cities' TFP is not separately identified. I thus normalize the average log skill to zero, and set μ_y to ensure this normalization holds. Holding cities' TFP constant, a greater skill dispersion implies a greater wage variance. I therefore calibrate σ_y to match the aggregate wage variance of young workers.

The other two parameters implicitly shape the skill distribution of old workers. First, g_0 in the learning technology (16) dictates the overall amount of human capital accumulation. I set g_0 to match the wage ratio between old and young workers. Second, the variance of the idiosyncratic learning shocks, σ_ν, influences the dispersion in the skills of old workers. I calibrate σ_ν to match the variance of log wages for old workers.

Location preferences The remaining parameters – cities' amenities and the dispersion in the taste shocks – determine workers' location preferences. The higher the amenities a city offers, the more workers want to live there. I set the age-specific local amenities to match the cities' age-specific employment share.[57] The dispersion in idiosyncratic location preferences, θ, governs the amount of sorting that prevails in equilibrium. When $\theta \approx 0$, the idiosyncratic location preferences are very dispersed, and the skill distributions are homogeneous across cities. As a consequence, there are no spatial differences in within-city wage inequality. As θ increases, skilled workers concentrate in productive cities, and those places become more unequal. I therefore set θ to match the cross-sectional correlation between cities' average wage and wage variance.

Sensitivity How precise is the calibration? Figure F.8 reports how the targeted moments vary as a function of the parameters in the model. Across the board, the targeted moments are sensitive in 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%. Appendix F.4 describes in more detail this sensitivity exercise and reports the sensitivity measure of Andrews et al. (2017).

4.4 Estimation Results and Over-identification Exercises

The estimated model implies significant sorting across space (Figure F.6). 30% of the workers in the top 10% of the skill distribution choose to work in Paris, and this fraction goes as high as 46% for workers in the top 1%. In contrast, low-skill workers are much less spatially concentrated: only 10% of workers in the bottom half of the skill distribution decide to work in Paris. These differences across skills in willingness to work in productive cities also vary across the lifecycle. In Appendix F.5,

[57] As for their TFP, cities' employment share can theoretically be expressed as a function of the their amenities and a residual (equation (63) in Appendix F.2). The residual can be computed within the model, and cities' amenities $\{B_\ell^y, B_\ell^o\}_{\ell=1}^L$ are recovered as a solution to a second fixed point. Figure F.4b plots the estimated amenities.

I show as an over-identification exercise that, in the model and the data, young skilled workers have a relatively higher propensity to work in productive cities than their old counterparts.

Worker sorting explains the majority of the between-city wage disparities. In the model, the between-city wage variance can be decomposed into three terms:

$$\mathbb{V}\text{ar}[\mathbb{E}(\log W_\ell)] = \mathbb{V}\text{ar}[\log T_\ell] + \mathbb{V}\text{ar}[\mathbb{E}(\log S_\ell)] + 2\mathbb{C}\text{ov}[\log T_\ell, \mathbb{E}(\log S_\ell)].$$

The first term captures how much of spatial wage inequality is explained by TFP gaps. I find the dispersion in cities' TFP to rationalizes only 16% of the between-city wage variance – and 1.7% of the aggregate variance. The second and third terms capture the spatial differences in skill composition that emanate from the sorting of workers into cities. Together, they explain the remaining 84% of the between-city wage variance. These estimates align with those of the literature that quantifies the impact of cities on wage inequality through movers designs (Dauth et al., 2022; Lhuillier, 2022; Card et al., 2023). In Appendix F.5, I follow this literature and, as a second over-identification exercise, re-estimate cities' TFP using a two-way fixed effect model à la Abowd et al. (1999).[58] The newly estimated TFPs can barely be distinguished from the baseline estimates. I conclude that the small dispersion in TFP across space is a robust feature of the French labor market.

Finally, the model explains 64% of the between-city wage growth variance in the data. To what extent does the spatial segmentation of learning opportunities cause these differences in wage growth? And what are the implications for spatial inequality and aggregate productivity? I use the estimated model to answer these questions.

5 The Tradeoff Between Human Capital and Inequality

In this section, I compare the baseline equilibrium with a counterfactual economy in which interactions are not segmented by cities. In the counterfactual world, interactions continue to be random, but the likelihood of an interaction is given by the aggregate skill density rather than the skill density of the city where individuals work.[59] I compare the two economies in steady state. Table 2 summarizes the key results.

5.1 Learning within Cities and Spatial Inequality

I find that learning opportunities are very spatially concentrated. Figure 4a plots the average skill growth by workplace in the baseline equilibrium (blue circles). On average, workers' skills grow by 15 log points throughout their lifetime. Individuals who work in high-wage cities experience faster human capital accumulation. For instance, workers in Paris experience a skill growth of 23 log points.

These differences in learning originate both from the sorting of fast-learning workers in productive

[58]This approach is not used as the baseline as it requires to take a stance on how to empirically measure skills.
[59]The aggregate skill density of the counterfactual economy may in turn differ from the baseline skill distribution to the extent that cities shape human capital accumulation.

Table 2: The consequences of spatially segmented learning opportunities

	No segmentation	Segmentation	Change
Average skill of old workers	1.231	1.246	1.21%
.. due to learning complementarity	·	·	58.5%
Average wage (€)	3,210	3,296	2.68%
Between-city wage variance	0.007	0.024	0.017
Between-city wage growth variance (log points)			
.. by workplace	0.016	13.85	13.85
.. by birthplace	0.000	3.281	3.281

Table contains aggregate statistics on inequality and productivity under two different assumption regarding the segmentation of learning interactions. First column corresponds to the case in which interactions are random in the whole economy. Second column is when interactions are fully segmented by cities (baseline, estimated model). The third column is the change between the two equilibrium.

cities ($g_1 > 1$) and from the concentration of learning opportunities in those locations ($g_2 > 0$ and $g_{12} > 0$). To quantify the relative importance of both channels, the orange hexagons in Figure 4a display the average skill growth by workplace were interactions not segmented by cities – holding the spatial allocation of workers constant. In this counterfactual economy, workers in Paris continue to experience on average faster skill growth due to their higher ability to learn. Yet, this sorting channel is quantitatively small as it explains 2% of the model-predicted between-city skill growth variance. I conclude that the segmentation of learning opportunities is the primary driver of the spatial differences in human capital accumulation.

The concentration of learning opportunities in a few cities has two consequences on geographic disparities. First, it engenders learning inequality. Figure 4b displays the average skill growth by birthplace against the city's average wage.[60] When workers learn from the individuals around them and face migration costs, workers born in or near skill-dense cities have an easier access to the local learning opportunities and enjoy faster human capital accumulation.[61] This birthplace learning premium is substantial. For instance, workers born in Paris have a lifetime skill growth 5.5 log points higher than workers born in the 25% poorest French cities.[62] The impact of individuals' birthplace on their lifetime learning is exacerbated for middle and high-skill workers who benefit relatively more from meeting other skilled workers (Figure G.1).

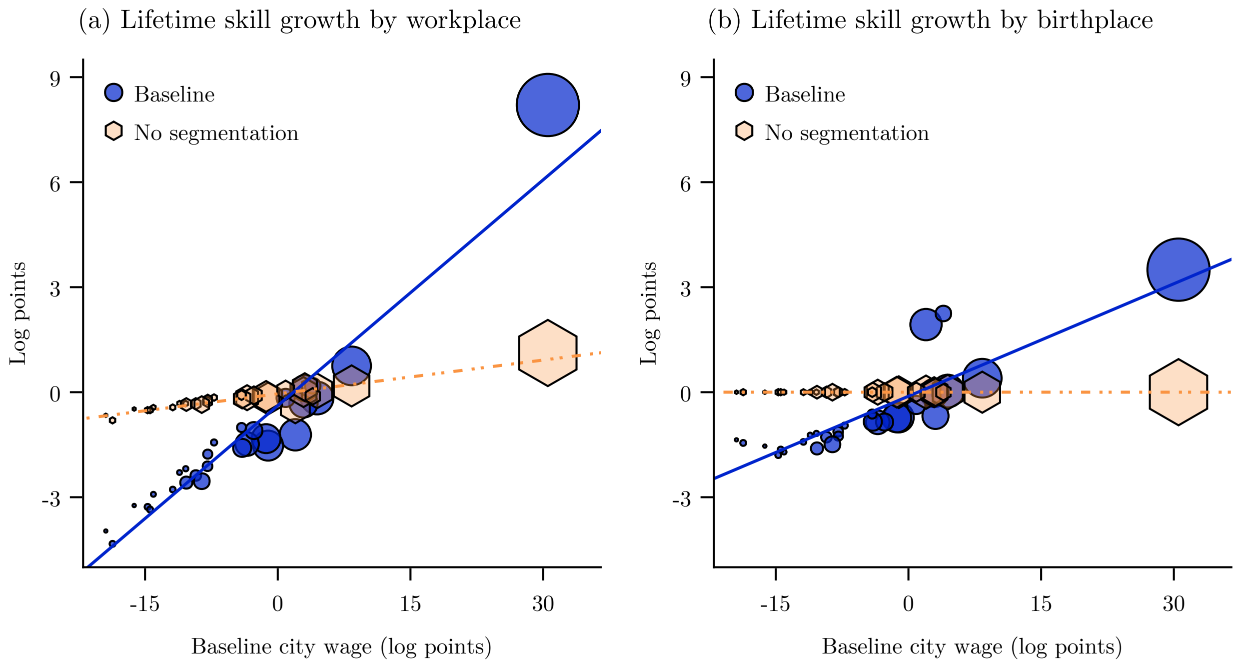
Second, the confinement of learning opportunities to productive cities amplifies spatial wage inequality. Proposition 2 found that, when individuals learn relatively more from skilled workers, productive cities get bigger as young workers go there to learn. In addition, these opportunities are more tailored to skilled workers, and the spatial concentration of high-skill rises. These agglomeration

[60]The model explains 55% of the wage growth variance across birthplaces in the data.

[61]Workers born near Paris indeed experience relatively faster skill growth, as shown in Figure G.2.

[62]The spatial differences in learning feeds into lifetime income inequality. When opportunities are spatially segmented, the lifetime income gaps between workers born in Paris and those born in the bottom 25% of the wage distribution is 12%, compared to 5% when interactions are not segmented.

Figure 4: Between-city differences in human capital accumulation



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.

effects are quantitatively large (Figure G.4): the population of Paris increases from 7% to 14% of the French workforce when workers learn from the individuals around them, and the propensity of workers in the top 10th percentile of the skill distribution to work there triples from 11% to 32%.[63] The greater agglomeration of skilled workers in a few cities amplifies spatial wage inequality, and the between-city wage variance rises from 0.007 to 0.024.[64]

5.2 Local Interactions and Human Capital Accumulation

While the concentration of learning opportunities in space generates geographic disparities, it also substantially boosts aggregate productivity. The average skill of old workers is 1.2% higher when individuals learn within cities, and the aggregate average wage rises by by 2.7%.[65]

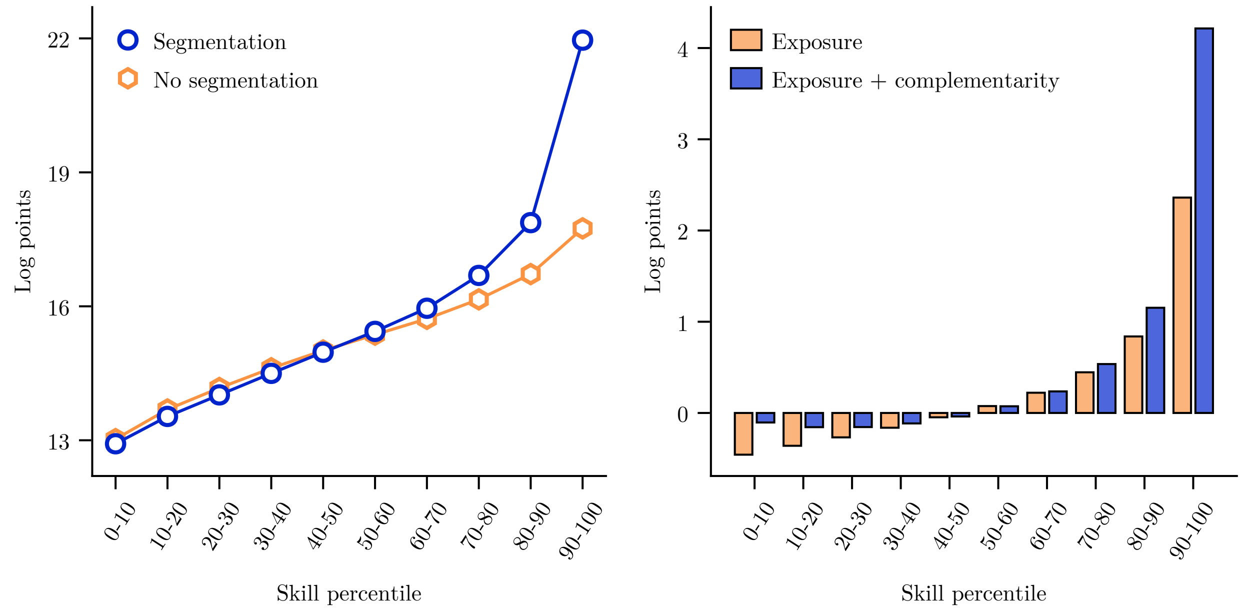
[63] In the presence of costly migration, local interactions agglomerate both young and old workers in productive cities, as the majority of old workers stay in the locations where they lived when young (Figure G.5a).

[64] The change in between-city wage inequality can be decomposed into a TFP effect, a skill disparity effect, and a sorting effect; specifically, $\Delta\mathbb{V}\text{ar}[\mathbb{E}\,(\log W_\ell)] = \Delta\mathbb{V}\text{ar}[\log T_\ell] + \Delta\mathbb{V}\text{ar}[\mathbb{E}\,(\log S_\ell)] + 2\Delta\mathbb{C}\text{ov}[\log T_\ell, \mathbb{E}\,(\log S_\ell)]$. The TFP effect reflects the greater agglomeration of workers to productive cities and explains only 1% of the total change in between-city wage inequality. The remaining two terms capture the heterogeneity in skill composition across cities, and they explain together the remaining 99%.

[65] The average wage increases relatively more than stock of human capital as it also takes into account the greater agglomeration of workers in high-TFP cities. Specifically, the change in the average log wage between the no-segmentation and the segmentation economies can be decomposed into a TFP term and a human capital term, $\mathbb{E}[\log \bar{W}] - \mathbb{E}[\log W] = \sum_\ell \log T_\ell \left(\bar{N}_\ell - N_\ell\right) + \mathbb{E}[\log \bar{S}] - \mathbb{E}[\log S]$, where $\bar{x}$ refers to the value of x in the counterfactual economy. In the aggregate, 16% of the wage increase is explained by the larger stock of human capital, and the remaining 84% are due to the greater agglomeration of workers in productive cities.

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

(a) Lifetime skill growth by initial skill | (b) Change in lifetime skill growth decomposition



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 (24). The blue bars plot the “average exposure” effect, and the orange bars plot the total change.

The spatial concentration of learning opportunities enhances aggregate productivity by boosting the human capital accumulation of skilled workers without dampening that of other individuals. Figure 5a displays the average lifetime skill growth as a function of workers’ initial skill when opportunities are spatially segmented (blue) and when they are not (orange). When individuals learn from the other workers in their city, workers in the top 10% of the skill distribution see their skill grow by 22 log points, 4 log points more than their lifetime human capital accumulation when interactions are not spatially segmented. At the other end of the distribution, the segmentation of learning opportunities reduces the lifetime skill growth of workers in the bottom 10% of the skill distribution by 0.1 log points.[66]

Why do high-skill workers gain from the spatial concentration of learning opportunities? Cities shape skill growth through two channels. First, high-skill workers agglomerate relatively more to skill-dense cities, and as result, are more exposed to productive interactions. I refer to this channel as the exposure effect. Second, high-skill workers are better able to capitalize on skilled interactions in the presence of strong within-skill learning complementarities. I coin this force the complementarity effect. Formally, the change in average lifetime skill growth between the baseline

[66] Learning within cities therefore heightens the difference in human capital accumulation between low- and high-skill. When interactions are spatially segmented, an increase in initial skill by one standard deviation is associated with a lifetime skill growth 1.6 log points larger. In contrast, this skill learning premium flattens to 0.6 log points when individuals learn from every worker.

and the counterfactual economy can be written

$$\mathbb{E}\left[\log\left(\frac{S^o}{s}\right) \mid s\right] - \mathbb{E}\left[\log\left(\frac{\bar{S}^o}{s}\right) \mid s\right] = \underbrace{g_2 \sum_{\ell} \left(\frac{n^y_\ell(s)}{n^y(s)}\right) \Delta_\ell}_{\text{Exposure}} + \underbrace{g_{12} \log s \sum_{\ell} \left(\frac{n^y_\ell(s)}{n^y(s)}\right) \Delta_\ell}_{\text{Complementarity}}, \tag{24}$$

where $\bar{x}$ refers to the value of x when interactions are not spatially segmented, and $\Delta_\ell \equiv \mathbb{E}[\log S_\ell] - \mathbb{E}[\log \bar{S}]$ is the difference between the average skill in city ℓ in the baseline equilibrium and the average aggregate skill in the counterfactual equilibrium.[67]

Learning complementarities are essential in amplifying the effect of cities on human capital accumulation. Figure 5b plots the skill growth decomposition (24). The exposure effect accounts for 56% of the faster learning experienced by workers in the top 10%; the complementarity channel explains the remaining 44% of their learning gains. Meanwhile, the learning complementarities also minimize the learning losses of workers in the bottom half of the skill distribution.[68] By boosting the gains of high-skill and reducing the losses of low-skill, learning complementarities are crucial for the impact of cities on human capital accumulation. Aggregating (24) across workers, the complementarity channel explain 59% of the aggregate increase in human capital.

This section therefore suggests that governments face an equity-efficiency tradeoff. On the one hand, the spatial segmentation of learning opportunities improves aggregate productivity. On the other hand, it amplifies spatial wage and learning inequality. As a result, policies aimed at reducing the spatial concentration of skills may have a detrimental effect on the aggregate stock of human capital. I conclude this chapter by quantifying how steep is the equity-efficiency tradeoff.

6 The Consequences of Spatial Policies

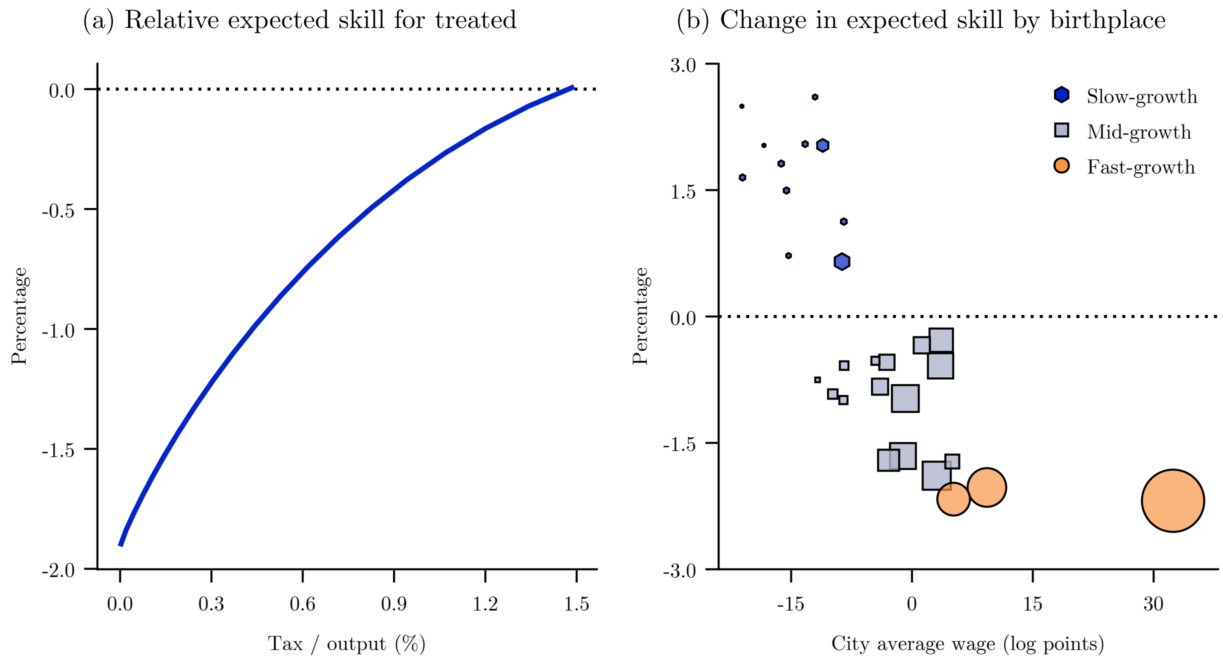
I study the general equilibrium consequences of a moving voucher policy aimed at offering equal learning opportunities to workers born in remote locations (Chetty et al., 2016).[69] Specifically, the policy consists of a subsidy granted to young workers born in the 25% poorest locations conditional on moving to the three largest French cities (Paris, Lyon, and Toulouse). Figure 8a displays the 185 cities treated by the policy. The policy is self-financed within skill and age group to focus on spatial redistribution.

[67] The change in skill growth can further be decomposed into a partial equilibrium effect that holds the average old skill constant, $\mathbb{E}[\log S] - \sum_\ell \left(\frac{n^y_\ell(s)}{n^y(s)}\right) \mathbb{E}[\log S_\ell]$, and a general equilibrium effect, $\mathbb{E}[\log \bar{S}] - \mathbb{E}[\log S]$. The general equilibrium effects alone are however small (see Figure G.3).

[68] In the baseline equilibrium, relatively low-skill workers do not sort much across space (Section 4.4). As a result, the average learning opportunity they have access to is similar to the average nationwide opportunity. This, together with the strong within-skill learning complementarities, explain why these workers do not lose much from the spatial segregation of opportunities.

[69] 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.

Figure 6: 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.

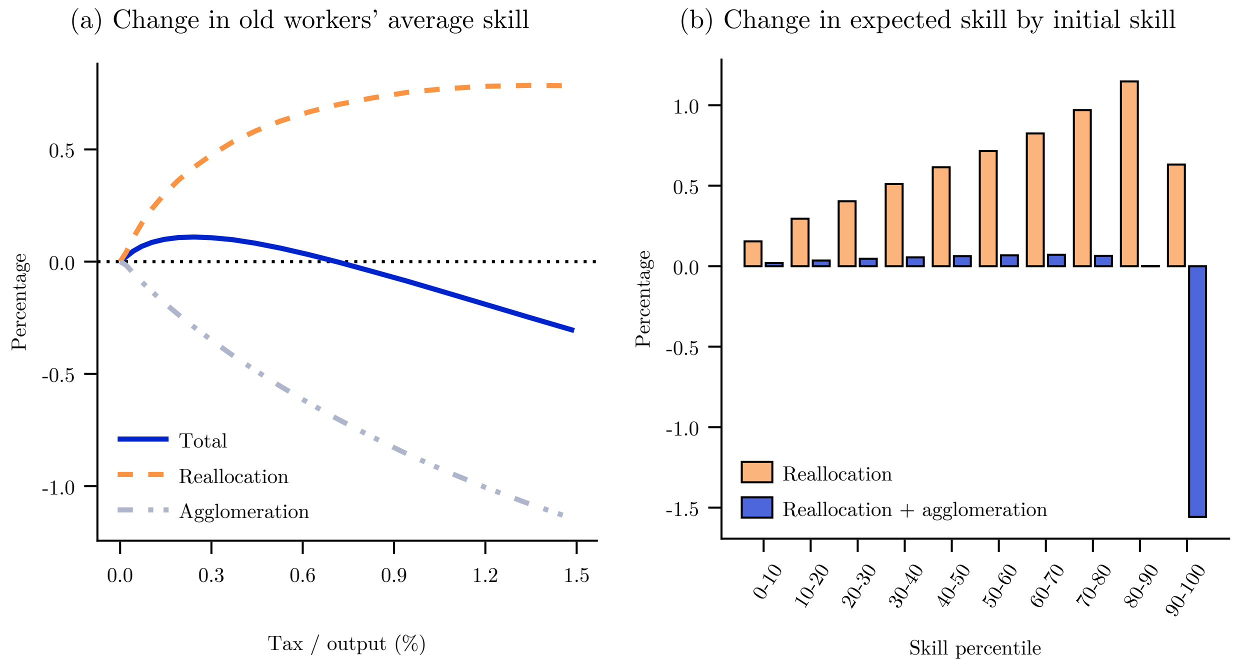
6.1 *Spatial Inequality*

Large moving vouchers have the potential to improve the lifetime human capital of workers born in remote locations. Figure 6a plots the average skill of treated workers at the end of their career compared to the average worker as a function of the policy size. Absent the vouchers, treated workers have a human capital 2% lower than the average French worker at the end of their career. As the size of the vouchers increases, it covers a greater share of the migration cost to productive cities. Treated workers have an easier access to the learning opportunities present in these cities, which increases their lifetime human capital. When the voucher reaches 1,030€, which covers 80% of the average migration cost and represents 1.5% of GDP, the probability that treated workers migrate to the three largest French cities rises from 7% to 45%. As a result, workers born in the 25% poorest locations accumulate human capital as much as the average worker.

The policy has heterogeneous treatment effects across workers. Figure 6b 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 on average an increase in their human capital.[70] The gains are relatively larger for workers born near Lyon, Paris, and Toulouse for whom the take-up rate of the policy is larger (Figure H.1). Similarly, skilled

[70]The average skill obtained by treated workers at the end of their career increases by 1.8%. Combined with an easier access to high-productivity jobs, the moving voucher policy boosts their lifetime labor income by 5.7%.

Figure 7: 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 blue bars represent the partial equilibrium, reallocation effect. The orange bars display the total change.

workers are more likely to opt-in the policy (Figure H.3a). Together with their better capacity to capitalize on the learning opportunities present in productive cities, skilled treated workers benefit relatively more from the policy (Figure H.3b).

The learning gains experienced by the treated workers come at the expense of slower human capital accumulation for the non-treated. The policy reallocates to productive cities workers that are marginally less skilled, and in doing so, worsens the quality of the local opportunities (Figure H.4). As a result, the policy reduces the learning advantage of workers born in Lyon, Paris and Toulouse (orange circles in Figure 6b). The policy also negatively affects the human capital accumulation of non-treated individuals born outside the three largest French cities who used to migrate there to work and learn (grey rectangles).[71]

Combining the effects on the treated and non-treated, the policy reduces spatial learning disparities, and the variance of wage growth across birthplaces decreases from 3.6 to 1.9.[72]

6.2 *Human Capital Accumulation*

The policy also negatively affects aggregate efficiency. Figure 7a displays the change in the average skill of old workers as a function of the size of the policy. When the policy reaches 1.5% of GDP, the stock of human capital is 0.3% lower than in the baseline equilibrium.

The policy negatively impacts aggregate efficiency by reducing the spatial segregation of learning opportunities. The impact of the policy on human capital accumulation can be decomposed into two effects. First, holding the quality of the local interactions constant, the vouchers reallocate workers to skill-dense places and increase the pool of individuals who have access to productive learning opportunities. I refer to this as the reallocation effect. Second, in general equilibrium, the policy affects the spatial distribution of skills, and with that the quality of the local opportunities. I name this channel the composition effect. Formally, the percentage change in the average skill of old workers between the policy and the baseline equilibrium can be decomposed into

$$\frac{\mathbb{E}[\bar{S}^o]-\mathbb{E}[S^o]}{\mathbb{E}[S^o]} = \underbrace{\sum_{\ell}\int\int\left(\frac{\gamma(s,s_p)}{\mathbb{E}[S^o]}\right)\Delta n^y_{\ell}(s)\pi_{\ell}(s_p)ds_p ds +}_{\text{Reallocation}} \underbrace{\sum_{\ell}\int\int\left(\frac{\gamma(s,s_p)}{\mathbb{E}[S^o]}\right)\bar{n}^y_{\ell}(s)\Delta\pi_{\ell}(s_p)ds_p ds,}_{\text{Composition}}$$

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 7a display the reallocation and composition effects. Abstracting from the composition effect, the moving voucher policy would increase aggregate human capital by 1%. However, the composition effect more than offsets the partial equilibrium impact of the policy, and in net, the policy triggers efficiency losses.

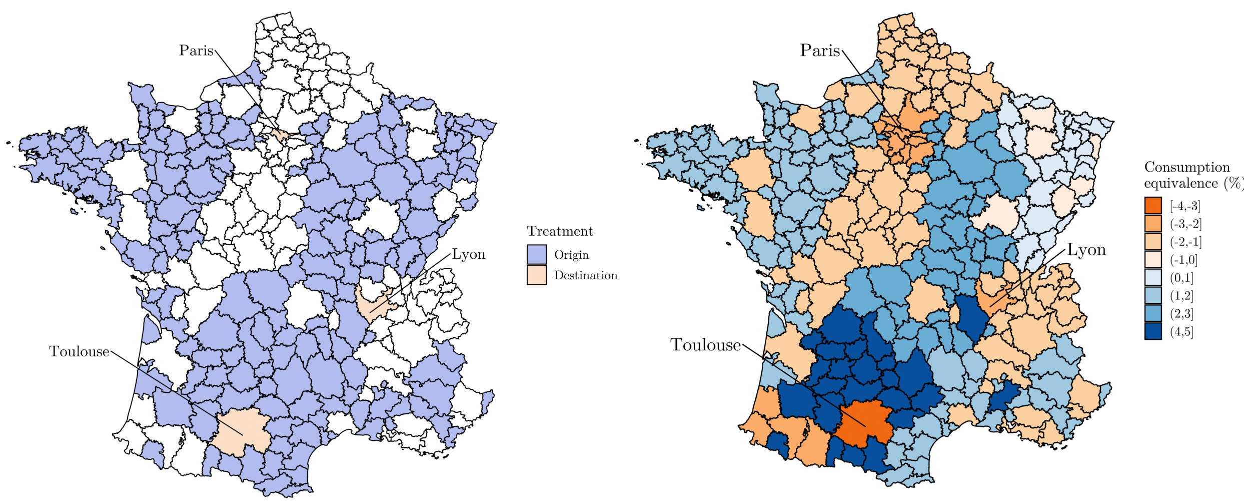
Skilled workers fully bear the learning losses generated by the policy. Figure 7b plots the change in workers' expected skill at the end of their career between the policy and the baseline equilibrium across the skill distribution. Holding constant the quality of the local learning opportunities, all workers would on average experience faster human capital accumulation (orange bars). The gains would be the largest for workers between the 50th and 90th percentile of the skill distribution for whom the enhanced access to interactions with skilled workers is most beneficial. However, the general equilibrium effect neutralizes those partial equilibrium gains, and the policy neither fosters nor dampens the learning of the bottom 90% of the skill distribution (blue bars). In contrast, the largest fraction of skilled workers already agglomerates in Lyon, Toulouse and Paris without the moving vouchers. For those workers, the policy only negatively affects the quality of the interactions

[71] The learning losses for the non-treated born outside Lyon, Paris and Toulouse are relatively lower for two reasons. First, not everyone born outside these cities migrates there to work. Second, in general equilibrium, some of these mid-productivity cities experience an increase in their skill density as skilled workers have marginally less incentives to agglomerate in productive locations (Figure H.4a). As such, the learning opportunities offered by these places improve.

[72] The policy also reduces spatial wage inequality by decreasing the spatial concentration of skilled workers. The between-city wage variance shrinks from 0.024 to 0.018 when the policy reaches 1.5% of GDP.

Figure 8: The consequences of the moving voucher policy on welfare

(a) Treated cities (b) Change in welfare by birthplace



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).

they experience, thus reducing their lifetime learning.[73]

Taking stock, the moving voucher policy reduces geographic learning disparities at the cost of lowered aggregate efficiency, thus facing a substantial equity-efficiency tradeoff. I conclude this section by quantifying the welfare consequences of the policy.

6.3 Welfare

I measure welfare through the expected lifetime utility of a cohort. I express welfare changes in terms of consumption equivalence (Lucas, 1987; Alvarez and Jermann, 2004). Appendix H.1 provides more details on how the consumption equivalence is calculated, as well as a welfare decomposition into different channels.

The policy is spatially redistributive: it generates welfare gains for the treated workers, and welfare losses for the non-treated. Figure 8b displays the average welfare effect of the policy by birthplaces.[74] On the one hand, workers born in treated cities enjoy average welfare gains of 2%. Those gains are relatively larger for workers born near Lyon, Paris, or Toulouse, and they are mainly explained by the faster human capital accumulation these workers experience (Figure H.5). On the other hand, workers born in non-treated cities suffer average welfare losses of 2.3%. For workers born outside of Lyon, Paris and Toulouse, these losses are entirely caused by the financial cost of the policy. For the non-treated born in the three largest French cities, the welfare losses are larger,

[73] Learning complementarities are crucial in shaping the efficiency cost of the policy. Were those absent, the learning losses experienced by workers in the top 10% of the skill distribution would be halved (Figure H.2).

[74] The average welfare change by birthplaces are computed under a utilitarian social welfare function that places equal weights on every worker within the city (see Appendix H.1 for more details). To complement this analysis, Figure H.6 reports the full distribution of welfare changes.

going as high as 3.5% for Toulouse, as the policy negatively affects their lifetime human capital.

In the aggregate, the welfare losses outweigh the gains, and 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.

Wrapping up, moving vouchers constitute an effective redistributive policy that substantially reduces spatial learning disparities. However, these vouchers are effective only if they cover a substantial fraction of the migration costs; taking into account the general equilibrium of the policy, moving vouchers reduce aggregate inefficiency and trigger welfare losses for non-treated workers.

7 Conclusion

This chapter 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 faster wage growth when working in skill-dense cities. I have argued that this pattern reveals the presence of strong within-skill learning complementarities. Using the estimated model, I concluded that the spatial segmentation of learning opportunities increases the aggregate stock of human capital at the cost of amplified geographic disparities.

This equity-efficiency tradeoff matters for spatial policies. On the one hand, local interactions engender learning spillovers and spatial misallocation. Place-based policies incentivizing a higher spatial concentration of skilled workers may correct the local externalities. On the other hand, the spatial concentration of skills 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 the cost of aggregate efficiency losses.

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 other. All these dimensions are sources of local spillovers. A natural direction along which to expand this research agenda is therefore to study spatial policies in the presence of static and dynamic complementarities. A second promising avenue for this agenda is to study how firms and space interact in shaping human capital accumulation; 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.

References

Abowd, J. M., Kramarz, F., and Margolis, D. N. (1999). High wage workers and high wage firms. *Econometrica*, 67(2):251–333.

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.

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.

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.

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.

Dauth, W., Findeisen, S., Moretti, E., and Suedekum, J. (2022). Matching in cities. *Journal of the European Economic Association*, 20(4):1478–1521.

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.

Dingel, J. I. and Tintelnot, F. (2020). Spatial economics for granular settings. Technical report, National Bureau of Economic Research.

Donald, E., FUKU, M., and Miyauchi, Y. (2023). Unpacking aggregate welfare in a spatial economy.

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.

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.

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.

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.

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

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.

Mongey, S. and Waugh, M. E. (2024). Discrete choice, complete markets, and equilibrium. Technical report, National Bureau of Economic Research.

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.

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.

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

Existence 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 shocks.

Assumption 2 (Idiosyncratic learning shocks).
Suppose that (i) the support of F is $\mathbb{R}_+$, and (ii) for any function $h : \mathbb{R}_+ \mapsto \mathbb{R}_+$, $\int h(ex)dF(e) < \infty$ for all $x < \infty$.

I start from the young spatial allocation. Existence of a solution to (5) 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 $\mathcal{M} : C_b(X, Y) \to Q$ by

$$\mathcal{M}[\boldsymbol{\eta}](\boldsymbol{s}) = \left\{ n(s_\ell) \left(\frac{\mathrm{e}^{\vartheta\left(s_\ell T_\ell + B^y_\ell \beta + O(s_\ell, \eta_\ell)\right)}}{\sum_{\ell'} \mathrm{e}^{\vartheta\left(s_\ell T_{\ell'} + B^y_{\ell'} + \beta O(s_\ell, \eta_{\ell'})\right)}} \right) \right\}_{\ell=1}^L$$

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

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

The second part of Assumption 2 guarantees that $O(s, \eta)$ exists. For any $\boldsymbol{\eta} \in C_b(X, Y)$, we know that $\mathcal{M}[\boldsymbol{\eta}](\boldsymbol{s})$ is bounded. Furthermore, Assumption 1 implies that $s \to O(s, n)$ is continuous, and therefore so is $\boldsymbol{s} \to \mathcal{M}[\boldsymbol{\eta}](\boldsymbol{s})$. Hence, $\mathcal{M} : C_b(X, Y) \to C_b(X, Y)$. Finally, $O(s, \eta)$ is continuous in η since $\eta_\ell(s) > 0$ and $\eta_\ell(s) < \infty$ for all $s \in [\underline{s}, \bar{s}]$, all $\ell \in \{1, 2, \ldots, L\}$, and any $\boldsymbol{\eta} \in C_b(X, Y)$. It automatically follows that $\mathcal{M}$ is a continuous operator, and Schauder fixed point theorem implies that there exists at least one $\eta^\star \in C_b(X, Y)$ such that $\mathcal{M}[\boldsymbol{\eta}^\star] = \eta^\star$.

Turning to the old allocation, (4) has a solution for all s as long as $n^o(s)$ has a solution. Using the first part of Assumption 2, the old skill distribution (6) can be rewritten

$$N^o(s^o) = \sum_\ell \int \int n^y_\ell(s_y)\pi^y_\ell(s_p) F\left(\frac{s_o}{\gamma(s_y, s_p)} \right) ds_p ds_y. \tag{25}$$

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. □

Uniqueness Equation (25) implies that, if the spatial allocation of young workers is unique, then so is N^o and therefore the equilibrium. Uniqueness of the young spatial allocation is proven by Banach fixed point theorem. I have already shown that $\mathcal{M}$ maps $C_b(X, Y)$ onto itself. It therefore only remain to prove that $\mathcal{M}$ is a contraction. For any $\boldsymbol{\eta}, \boldsymbol{\eta}' \in C_b(X, Y)$, we want to show

$$\begin{aligned}\|\mathcal{M}_\ell[\boldsymbol{\eta}](\boldsymbol{s}) - \mathcal{M}_\ell[\boldsymbol{\eta}'](\boldsymbol{s})\| &= \left\| n(s_\ell) \left[\left(\frac{\mathrm{e}^{\vartheta\left(s_\ell T_\ell + B^y_\ell + \beta O(s_\ell, \eta_\ell)\right)}}{\sum_{\ell'} \mathrm{e}^{\vartheta\left(s_\ell T_{\ell'} + B^y_{\ell'} + \beta O(s_\ell, \eta_{\ell'})\right)}} \right) - \left(\frac{\mathrm{e}^{\vartheta\left(s_\ell T_\ell + B^y_\ell + \beta O(s_\ell, \eta'_\ell)\right)}}{\sum_{\ell'} \mathrm{e}^{\vartheta\left(s_\ell T_{\ell'} + B^y_{\ell'} + \beta O(s_\ell, \eta'_{\ell'})\right)}} \right) \right] \right\| \\ &\leq K\|\eta_\ell(s_\ell) - \eta'_\ell(s_\ell)\|,\end{aligned}$$

for $K \in [0, 1)$. The term $\|\eta_\ell(s_\ell) - \eta'_\ell(s_\ell)\|$ is independent of $\vartheta\beta$. Meanwhile, the difference in the second equality is bounded, decreasing in $\vartheta\beta$, and converging to zero as $\vartheta\beta = 0$. Hence, for any $\boldsymbol{\eta}$ and $\boldsymbol{\eta}'$ with $\|\eta_\ell(s_\ell) - \eta'_\ell(s_\ell)\| > 0$, there always exist a $\vartheta\beta$ small enough such that the inequality holds. It follows that, for $\vartheta\beta$ small enough, $\mathcal{M}$ is a contraction, and the spatial allocation of young workers is unique. □

A.2 Uniqueness and Stability of the Symmetric Equilibrium

Proposition A.1 (Stability and uniqueness).
Suppose that $T_\ell = \bar{T}$ for all ℓ. Then, a symmetric equilibrium exists. If $\vartheta\beta$ is small, this equilibrium is unique.

Proof. Existence is shown by guessing and verifying that $\pi_\ell^y = n^y$ is an equilibrium. Uniqueness follows from Proposition 1. $\square$

The corollary below derives a bound on $\vartheta\beta$ that ensure the local stability and uniqueness of the symmetric equilibrium under a functional form restriction on the learning technology. For that, define the relative amenity of city ℓ

$$\mathcal{B}_\ell^a \equiv \frac{\mathrm{e}^{\vartheta B_\ell^a}}{\sum_{\ell'} \mathrm{e}^{\vartheta B_{\ell'}^a}}, \quad a \in \{y, o\}. \tag{26}$$

Corollary A.2.
Suppose that $T_\ell = \bar{T}$ for all ℓ, and $\gamma(s, s') = \mathcal{F}(s) + \mathcal{G}(s)\mathcal{H}(s')$ for $\mathcal{F}$, $\mathcal{G}$ and $\mathcal{H}$ some functions. The symmetric equilibrium is stable if

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

and it is unique if L is large and

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

Proof. For simplicity, I do the proof assuming away the idiosyncratic learning shocks. Under $\boldsymbol{T} = \bar{T}$, the expected utility of old workers is $\mathcal{V}^o(s) = s\bar{T} + c$ 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 expected future utility of young workers employed in city ℓ is $O_\ell(s) = \bar{T}\mathcal{F}(s) + \bar{T}\mathcal{G}(s)\mathbb{E}[\mathcal{H}(S_\ell)]$. Let $\mu_\ell \equiv \mathbb{E}[\mathcal{H}(S_\ell)]$ and $\mu \equiv \mathbb{E}[\mathcal{H}(S)]$. The spatial allocation of young workers is given by

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

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

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

Hence, we have reduced the problem to a L-dimensional fixed point. Let $\mathcal{M}$ be the operator that maps $\boldsymbol{\mu}$ onto itself as defined by the above equation. 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 $\mathcal{M} : [\underline{h}, \bar{h}]^L \mapsto [\underline{h}, \bar{h}]^L$. Clearly, $\mu_\ell = \mu$ is a solution to this fixed point. A sufficient condition for the symmetric equilibrium to be stable is $|\partial_{\mu_{\ell'}} \mathcal{M}_\ell(\boldsymbol{\mu})| < 1$ for all pair (ℓ, ℓ') and $\boldsymbol{\mu} = \{\mu\}_{\ell=1}^L$. A sufficient condition for the symmetric equilibrium to be unique is $|\partial_{\mu_{\ell'}} \mathcal{M}_\ell(\boldsymbol{\mu})| < 1$ for all pair (ℓ, ℓ') and all $\boldsymbol{\mu}$, in which case $\mathcal{M}$ is a contraction.[75] For any $\boldsymbol{\mu}$, the partial derivative of $\mathcal{M}_\ell$ w.r.t. μ_ℓ is

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

[75] To see this, note that the condition implies $\|D\mathcal{M}(\boldsymbol{\mu})\| \leq K < 1$. 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

$$\mathcal{M}(\boldsymbol{\mu}) - \mathcal{M}(\boldsymbol{\mu}') = \mathcal{M}[\Delta(1)] - \mathcal{M}[\Delta(0)] = \int_0^1 \frac{d\mathcal{M}[\Delta(\lambda)]}{d\lambda} d\lambda = \int_0^1 d\mathcal{M}[\Delta(\lambda)] \frac{\partial \Delta(\lambda)}{\partial \lambda} d\lambda,$$

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,

$$\|\mathcal{M}(\boldsymbol{\mu}) - \mathcal{M}(\boldsymbol{\mu}')\| = \|\int_0^1 d\mathcal{M}[\Delta(\lambda)] \frac{\partial \Delta(\lambda)}{\partial \lambda} d\lambda\| \leq \int_0^1 \|d\mathcal{M}[\Delta(\lambda)]\| \|\frac{\partial \Delta(\lambda)}{\partial \lambda}\| d\lambda \leq \int_0^1 K\|\boldsymbol{\mu} - \boldsymbol{\mu}'\| d\lambda = K\|\boldsymbol{\mu} - \boldsymbol{\mu}'\|,$$

where the first inequality is from the triangle inequality. Since $K < 1$, $\mathcal{M}$ is a contraction.

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

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

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

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

and

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

for $\ell' \neq \ell$. Imposing $|\partial_{\mu_{\ell'}} F_\ell(\boldsymbol{\mu})| < 1$ 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, if amenities are independent of the number of cities, then $\lim_{L\to\infty} \mathcal{B}_\ell \to 0^+$ for all ℓ. Hence, for L large, the partial derivatives rewrite

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

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

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

The local distributions are equilibrium objects. 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. □

A.3 The Sorting of Old Workers

Proposition A.3 (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':

$$\pi_\ell(s) - \pi_{\ell'}(s) = \left(\frac{N_{\ell'}}{N_\ell}\right) \pi_{\ell'}(s) \mathrm{e}^{\vartheta\left(B_\ell - B_{\ell'}\right)} \left(\mathrm{e}^{\vartheta s (T_\ell - T_{\ell'})} - \mathbb{E}\left[\mathrm{e}^{\vartheta S_{\ell'}(T_\ell - T_{\ell'})}\right]\right). \tag{27}$$

This follows from the definition of $\pi_\ell(s)$, N_ℓ and (4). Let $f : \mathcal{S}^o \mapsto \mathbb{R}$ be any measurable function. Using (27), the difference in the g-transformed mean skill between city ℓ and ℓ' is

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

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

A.4 The First Order Approximation – general technology

Old workers Linearizing the old allocation is straightforward. From (4), we know that

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

for $I^o_{\ell\ell'} \equiv \mathbb{1}\{\ell' = \ell\} - \mathcal{B}^o_{\ell'}$ and $\mathcal{B}^o_\ell$ defined in (26). The first term is the GE effect of increasing the TFP of city ℓ' on the density of skill s. We can guess that this term will be nil. The second term is the reallocation of workers over space. Hence, to a first order, the spatial allocation of old workers is

$$n^o_\ell(s) \approx \bar{n}^o(s)\mathcal{B}^o_\ell + \theta\bar{n}^o(s)\mathcal{B}^o_\ell \log\left(\frac{T_\ell}{\bar{T}}\right) s, \tag{28}$$

where

$$\mathcal{T}^o_\ell \equiv \log\left(\frac{T_\ell}{\bar{T}}\right) - \sum_l \mathcal{B}^o_l \log\left(\frac{T_l}{\bar{T}}\right). \tag{29}$$

When $B^o_\ell = B^o$ for all ℓ and $\mathbb{E}[\log T_\ell] = \log \bar{T}$, this boils down to (7).

Young workers For the young, given (5), we have

$$\left.\frac{\partial n^y_\ell(s)}{\partial \log T_{\ell'}}\right|_{\boldsymbol{T}=\bar{T}} = \vartheta \mathcal{B}^y_\ell n^y(s)\left[\bar{T} s I^y_{\ell\ell'} + \beta\left(\frac{\partial O_\ell}{\partial \log T_{\ell'}} - \sum_l \mathcal{B}^y_l \frac{\partial O_l(s)}{\partial \log T_{\ell'}}\right)\right].$$

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

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

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

$$\left.\frac{\partial O_\ell(s)}{\partial \log T_{\ell'}}\right|_{\boldsymbol{T}=\bar{T}} = \bar{T}\mathcal{B}^o_{\ell'}\int \gamma(s,s_p)n^y(s_p)ds_p + \bar{T}\int\gamma(s,s_p)\frac{\partial \pi^y_\ell(s_p)}{\partial \log T_{\ell'}}ds_p + c,$$

where I have redefined $\gamma = \gamma\int edF(e)$ wlog.[76] Using (2), it follows that

$$\begin{aligned}\left.\frac{\partial O_\ell}{\partial \log T_{\ell'}}\right|_{\boldsymbol{T}=\bar{T}} - \sum_l \mathcal{B}^y_l \left.\frac{\partial O_l}{\partial \log T_{\ell'}}\right|_{\boldsymbol{T}=\bar{T}} &= \bar{T}\int \left.\frac{\partial \pi^y_\ell(s_p)}{\partial \log T_{\ell'}}\right|_{\boldsymbol{T}=\bar{T}} \gamma(s,s_p)ds_p \\ &= \frac{\bar{T}}{\mathcal{B}^y_\ell}\int \left.\frac{\partial n^y_\ell(s_p)}{\partial \log T_{\ell'}}\right|_{\boldsymbol{T}=\bar{T}} \left(\gamma(s,s_p) - \mathbb{E}[\gamma(s,\mathcal{S}^y)]\right)ds_p,\end{aligned}$$

where the second equality follows from the expression for $\partial_{T_{\ell'}}\pi^y_\ell(s_p)$. Plugged in the initial expression, we obtain

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

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

$$\left.\frac{\partial n^y_\ell(s)}{\partial \log T_{\ell'}}\right|_{\boldsymbol{T}=\bar{T}} = \theta n^y(s)\mathcal{B}^y_\ell I^y_{\ell\ell'}\eta(s), \tag{30}$$

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

$$\left.\frac{\partial \pi^y_\ell(s)}{\partial \log T_{\ell'}}\right|_{\boldsymbol{T}=\bar{T}} = \theta n^y(s) I^y_{\ell\ell'}\left(\eta(s) - \mathbb{E}[\eta(S^y)]\right). \tag{31}$$

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

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

where $\Theta(s)$ is defined in (8). 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 $1/\theta$, are large enough.

[76] 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.

Lemma A.1 (Existence and uniqueness).
Suppose that

$$\theta\beta\left(\int\int|\gamma(s,s_p)-\mathbb{E}[\gamma(s,S)]|^2dN(s_p)dN(s)\right)^{1/2}<1.$$

Then, a solution to (8) *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 (8) as

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

This constitutes a Fredholm equation of the second kind. I restrict my attention to solutions to (8) that belong to $L_2(dN,[\underline{s},\bar{s}])$. Under Assumption 1 and $\bar{s}<\infty$, we have[77]

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

for C some constant. In addition, K is globally continuous with respect to s. It follows that any solution to (8) is bounded and continuous in $[\underline{s},\bar{s}]$ (Zabreĭko, 1975, Chapter 2, Theorems 1.7 and 1.8).

Uniqueness of a solution to (8) implies existence (Zabreĭko, 1975, Theorem 1.2). Equation (8) has a unique solution when $\theta\beta<|\bar{\lambda}|$, 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.5). Without this restriction, solving for $|\bar{\lambda}|$ is non-trivial. However, if the Neumann series

$$\eta^f(s)=f(s)+\sum_{j=1}^{\infty}(\theta\beta)^j\int K_j(x,t)tdN(t) \tag{32}$$

converges for any function f satisfying $\int|f(s)|^2dN(s)<\infty$, where K_j is the j-th iterated kernel of K, then $\theta\beta<|\bar{\lambda}|$ (e.g. Zabreĭko, 1975, Chapter 2). The inequality $\theta\beta\|K\|<1$ provides a sufficient condition for (32) to converge. □

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

$$n_\ell^y(s)\approx n^y(s)\mathcal{B}_\ell^y+\theta n^y(s)\mathcal{B}_\ell^y\mathcal{T}_\ell^y s+\theta^2\beta n^y(s)\mathcal{B}_\ell^y\mathcal{T}_\ell^y\Theta(s), \tag{33}$$

where $\mathcal{T}_\ell^y$ is defined analogously to (29). The within-city young skill distribution is

$$\pi_\ell^y(s)\approx n^y(s)+\theta n^y(s)\mathcal{T}_\ell^y\left(s-\mathbb{E}[S^y]\right)+\theta^2\beta n^y(s)\mathcal{T}_\ell^y\left(\Theta(s)-\mathbb{E}[\Theta(S^y)]\right). \tag{34}$$

Accordingly, the young average skill in city ℓ is

$$\mathbb{E}[S_\ell^y]\approx\mathbb{E}[S^y]+\theta\mathcal{T}_\ell^y\mathbb{V}\text{ar}[S^y]+\theta^2\beta\mathcal{T}_\ell^y\mathbb{C}\text{ov}[S^y,\Theta(S^y)],$$

as claimed in (9).

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 (6) w.r.t. $T_{\ell'}$ yields

$$\begin{aligned}\left.\frac{\partial N^o(s_o)}{\partial\log T_{\ell'}}\right|_{\boldsymbol{T}=\bar{T}}=\mathcal{B}_\ell^y\int\int\sum_\ell\left.\frac{\partial\pi_\ell^y(s_p)}{\partial\log T_{\ell'}}\right|_{\boldsymbol{T}=\bar{T}}F\left(\frac{s_o}{\gamma(s_y,s_p)}\right)ds_pdN^y(s_y)+\\ \int\int\sum_\ell\left.\frac{\partial n_\ell^y(s_y)}{\partial\log T_{\ell'}}\right|_{\boldsymbol{T}=\bar{T}}F\left(\frac{s_o}{\gamma(s_y,s_p)}\right)dN^y(s_p)ds_y=0,\end{aligned}$$

where the second equality follows from (2).

A.5 The First Order Approximation – example

Suppose that the learning technology reads:

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

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

for $\mathcal{F}$, $\mathcal{G}$ and $\mathcal{H}$ functions that satisfy Assumption 1. This class of learning technology contains (1). The integral equation simplifies to

$$\eta(s) = s + \theta\beta\mathcal{G}(s)\int\left(\mathcal{H}(s_p) - \mathbb{E}[\mathcal{H}(S^p)]\right)\eta(s_p)dN^y(s_p). \tag{35}$$

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

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

From here, it is easy to show that the characteristic root of (35) is $|\bar{\lambda}| = |1/\mathbb{C}\mathrm{ov}[\mathcal{G}(S^y), \mathcal{H}(S^y)]|$. Hence, a solution to (35) exists and is unique when $\theta\beta|\mathbb{C}\mathrm{ov}[\mathcal{G}(S^y), \mathcal{H}(S^y)]| \leq 1$. 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 (35) to obtain a closed-form expression for Θ:

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

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

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

as claimed in the main text.

A.6 The Second Order Approximation

The second order approximations are implemented assuming identical amenities for simplicity.[78]

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

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

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 in γ wlog. The partial derivative of the expected skill for workers with skill s w.r.t. the TFP of city ℓ is

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

[78] The aggregate effects on productivity are due to the spatial variations in local skill density, see (34). The local skill densities are independent of city amenity. Adding city amenity would therefore not change the conclusion.

By a similar argument as for N^o, there is no first order effect around $\boldsymbol{T} = \bar{T}$. Hence, moving to the second order, differentiate the above expression w.r.t. $T_{l'}$ to get

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

where I am using the notation $\partial_l f(x) = |\partial f(x)/\partial \log T_l|_{\boldsymbol{T}=\bar{T}}$, and the equality on the first line follows from (2). Meanwhile, the second-order derivative of $\pi^y_\ell(s)$ evaluated at $\boldsymbol{T} = \bar{T}$ and summed across cities read

$$\sum_\ell \left.\frac{\partial^2 \pi^y_\ell(s)}{\partial \log T_l \partial \log T_{l'}}\right|_{\boldsymbol{T}=\bar{T}} = -L \sum_\ell \left(\left.\frac{\partial \pi^y_\ell(s)}{\partial \log T_l}\right|_{\boldsymbol{T}=\bar{T}} \left.\frac{\partial N_\ell}{\partial \log T_{l'}}\right|_{\boldsymbol{T}=\bar{T}} + \left.\frac{\partial \pi^y_\ell(s)}{\partial \log T_{l'}}\right|_{\boldsymbol{T}=\bar{T}} \left.\frac{\partial N_\ell}{\partial \log T_l}\right|_{\boldsymbol{T}=\bar{T}} \right).$$

Plugged in the expression for the aggregate change in old skill, we therefore obtain

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

This expression relies entirely on first order terms which we have derived in Section A.4. Using (31), we get

$$\sum_\ell \left.\frac{\partial \pi^y_\ell(s)}{\partial \log T_l}\right|_{\boldsymbol{T}=\bar{T}} \left.\frac{\partial \pi^y_\ell(s')}{\partial \log T_{l'}}\right|_{\boldsymbol{T}=\bar{T}} = \theta^2 I_{ll'} \left(\eta(s) - \mathbb{E}[\eta(S^y)]\right) \left(\eta(s') - \mathbb{E}[\eta(S^y)]\right) n^y(s') n^y(s),$$

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

$$\left.\frac{\partial^2 \mathbb{E}[S^o \mid s]}{\partial \log T_l \partial \log T_{l'}}\right|_{\boldsymbol{T}=1} = \frac{2}{L} \theta^2 I_{ll'} \Omega(s)$$

for

$$\Omega(s) \equiv \left(\eta(s) - \mathbb{E}[\eta(S^y)]\right) \int \gamma(s,s') \left(\eta(s') - \mathbb{E}[\eta(S^y)]\right) dN^y(s'). \tag{36}$$

We can therefore conclude that, to a second-order around $\boldsymbol{T} = \bar{T}$, the expected old skill workers with young skill s is

$$\mathbb{E}\left[S^o \mid s\right] \approx \mathbb{E}\left[\mathrm{S}^o \mid s\right] + \vartheta^2 \mathbb{V}\mathrm{ar}[\log T_\ell] \Omega(s).$$

Aggregating across the skill distribution yields (12).

Between-city wage inequality Let $\bar{\omega}^y_\ell \equiv \mathbb{E}[\log W^y_\ell]$ be the average log wage in city ℓ. Between-city wage inequality is then defined as

$$\mathbb{V}\mathrm{ar}[\bar{\omega}^y_\ell] = \sum_\ell \left(\bar{\omega}^y_\ell - \bar{\omega}^y\right)^2 N^y_\ell,$$

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

$$\frac{\partial \mathbb{V}\mathrm{ar}[\bar{\omega}^y_\ell]}{\partial \log T_l} = 2 \sum_\ell \left(\bar{\omega}^y_\ell - \bar{\omega}^y\right) \left(\frac{\partial \bar{\omega}^y_\ell}{\partial \log T_l} - \frac{\partial \bar{\omega}^y}{\partial \log T_l}\right) N^y_\ell + \sum_\ell \left(\bar{\omega}^y_\ell - \bar{\omega}^y\right)^2 \frac{\partial N^y_\ell}{\partial \log T_l},$$

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

$$\partial^2_{ll'} \mathbb{V}\mathrm{ar}[\bar{\omega}_\ell^y] = \frac{2}{L} \sum_\ell \left(\partial_{l'} \bar{\omega}_\ell^y - \partial_{l'} \bar{\omega}^y \right) \left(\partial_l \bar{\omega}_\ell^y - \partial_l \bar{\omega}^y \right). \tag{37}$$

All that is required is therefore the first order effect of cities' TFP on wages. Using (31), the first marginal effect reads

$$\partial_l \bar{\omega}_\ell^y = \mathbb{1}\{\ell = l\} + \theta I_{l\ell} \mathbb{C}\mathrm{ov}[\log S^y, \eta(S^y)].$$

Using (2), the second marginal effect is

$$\partial_l \bar{\omega}^y = \frac{1}{L}.$$

Combined, (37) rewrites

$$\partial^2_{ll'} \mathbb{V}\mathrm{ar}[\bar{\omega}_\ell^y] = \frac{2}{L} I_{ll'} \left(1 + \theta \mathbb{C}\mathrm{ov}[\log S^y, \eta(S^y)]\right) \left(1 + \theta \mathbb{C}\mathrm{ov}[\log S^y, \eta(S^y)]\right)^2.$$

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

$$\mathbb{V}\mathrm{ar}[\bar{\omega}_\ell^y] \approx \mathbb{V}\mathrm{ar}[\log T_\ell] \left(1 + \theta \mathbb{C}\mathrm{ov}[\log S^y, S^y] + \theta^2 \beta \mathbb{C}\mathrm{ov}[\log S^y, \Theta(S^y)]\right)^2. \tag{38}$$

A.7 *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 (8) 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

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

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, $\eta'(s) > 1 + \theta\beta \mathbb{C}\mathrm{ov}[\gamma_1(s, S), S] \geq 0$, where the second inequality is from the assumption on γ. This verifies our guess.

A.8 *Proof of Proposition 2*

Proposition 2.1 To a first order, (28) and (33) imply that the size of city ℓ is

$$N_\ell \approx \mathcal{B}_\ell^y + \mathcal{B}_\ell^o + \theta \left(\mathcal{B}_\ell^y \mathcal{T}_\ell^y \mathbb{E}[S^y] + \mathcal{B}_\ell^o \mathcal{T}_\ell^o \mathbb{E}[S^o] \right) + \theta^2 \beta \mathcal{B}_\ell^y \mathcal{T}_\ell^y \mathbb{E}[\Theta(S^y)].$$

Meanwhile, if interactions are not spatially segmented, then

$$\tilde{N}_\ell \approx \mathcal{B}_\ell^y + \mathcal{B}_\ell^o + \theta \left(\mathcal{B}_\ell^y \mathcal{T}_\ell^y \mathbb{E}[S^y] + \mathcal{B}_\ell^o \mathcal{T}_\ell^o \mathbb{E}[S^o] \right).$$

When $\boldsymbol{T} \approx \bar{T}$, we thus have that for any ℓ such that $\mathcal{T}_\ell^y > 0$, then $N_\ell > \tilde{N}_\ell \iff \mathbb{E}[\Theta(S^y)] > 0$. Under Lemma 1, η is an increasing function. Hence, $\gamma(s, \cdot)$ increasing implies $\Theta(s) = \mathbb{C}\mathrm{ov}[\gamma(s, S^y), \eta(S^y)] > 0$. Conversely, $\gamma(s, \cdot)$ decreasing implies $\Theta(s) < 0$. When amenities are the same everywhere, $\mathcal{T}_\ell^y > 0 \iff T_\ell > \bar{T}$, which yields Proposition 2.1. $\square$

Proposition 2.2 From (9), the difference between the average skill in city ℓ when interactions are spatially segmented vs. when they are not is given by

$$\mathbb{E}[S_\ell^y] - \mathbb{E}[\tilde{S}_\ell^y] \approx \theta^2 \beta \mathcal{T}_\ell^y \mathbb{C}\mathrm{ov}[S^y, \Theta(S^y)].$$

Similarly, from (38), the difference between the between-city wage inequality when interactions are segmented vs. when they are not is

$$\mathbb{S}\mathrm{d}[\bar{\omega}_\ell^y] - \mathbb{S}\mathrm{d}[\tilde{\bar{\omega}}_\ell^y] \approx \mathbb{S}\mathrm{d}[\log T_\ell]\theta^2\beta\mathbb{C}\mathrm{ov}[\log S^y, \Theta(S^y)].$$

Hence, when $\boldsymbol{T} \approx \bar{T}$, we have: (i) for any ℓ such that $\mathcal{T}_\ell^y > 0$, then $\mathbb{E}[S_\ell^y] > \mathbb{E}[\tilde{S}_\ell^y] \iff \mathbb{C}\mathrm{ov}[S^y, \Theta(S^y)] > 0$, and (ii) $\mathbb{S}\mathrm{d}[\bar{\omega}_\ell^y] > \mathbb{S}\mathrm{d}[\tilde{\bar{\omega}}_\ell^y] \iff \mathbb{C}\mathrm{ov}[\log S^y, \Theta(S^y)] > 0$. If γ is supermodular, then $\gamma_1(s,\cdot)$ is increasing. Hence $\Theta'(s) = \mathbb{C}\mathrm{ov}[\gamma_1(s, S^y), \eta(S^y)] > 0$ (Lemma 1), and $\mathbb{C}\mathrm{ov}[S^y, \Theta(S^y)] > 0$. Conversely, $\gamma_1(s,\cdot)$ is decreasing under γ, and $\mathbb{C}\mathrm{ov}[S^y, \Theta(S^y)] < 0$. □

A.9 Proof of Proposition 3

The second order approximation of $\mathbb{E}[S^o]$ is derived in Section A.6. Define $\chi(s) \equiv \eta(s) - \mathbb{E}[\eta(S^y)]$. From the fundamental theorem of calculus,

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

From the definition of χ, $\int \gamma(s, \underline{s})\chi(s)ds = \int \gamma(\underline{s}, s')\chi(s')ds' = \int \gamma(\underline{s}, \underline{s})\chi(s)ds = 0$. Therefore, Ω rewrites

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

where $\Psi(x) \equiv \int_x \chi(s)dN^y(s)$ and the second equality follows from Fubini's. Note that Ψ rewrites $\Psi(x) = (1 - N(x))\,(\mathbb{E}[\eta(S^y) \mid S^y \geq x] - \mathbb{E}[\eta(S^y)])$. 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.10 The Learning Consequences of Local Interactions across Skills

Proposition A.4 (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 $\mathbb{E}[S^o \mid s] > \mathbb{E}[\bar{S}^o \mid s]$ if and only if $s > s^\star$;*
2. *If $\gamma(s,\cdot)$ is decreasing for all s, then $\mathbb{E}[S^o \mid s] > \mathbb{E}[\bar{S}^o \mid s]$ if and only if $s < s^\star$.*

Proof. The second order approximation for $\mathbb{E}[S^o \mid s]$ is derived in Section A.6. 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 3 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 $s < s^\star$. □

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:

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

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 restrict my attention to allocation without between-skill transfers, wlog. I also suppose that between-age transfers are not allowed to focus on the learning externalities.

The planner thus faces the following constraints, where the variables in the parenthesis refer to the associated Karush–Kuhn–Tucker multiplier:

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

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

2. Young labor supply ($\lambda_{t\ell}^y(s)$):

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

3. Old labor supply $\lambda_{t\ell}^o(s)$:

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

4. The law of motion for the skill distribution of old workers ($\nu_t(s)$):

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

The Lagrangian associated to the planner's problem is

$$\begin{aligned}\mathcal{L} = \frac{1}{\vartheta} \int \log\left(\sum_{\ell} \mathrm{e}^{\vartheta c^{o\star}_{0\ell}(s)}\right) \sum_{\ell} n^{o\star}_{0\ell}(s) ds + \frac{1}{\vartheta} \sum_{t>0} \beta^t_S \int \log\left(\sum_{\ell} \mathrm{e}^{\vartheta\left(c^{y\star}_{t\ell}(s) + \beta O^{\star}_{t\ell}(s)\right)}\right) \sum_{\ell} n^{y\star}_{t\ell}(s) ds - \\ \sum_t \sum_{\ell} \int \lambda^y_{t\ell}(s) \left(n^{y\star}_{t\ell}(s) - n^y(s) \left(\frac{\mathrm{e}^{\vartheta\left(c^{y\star}_{t\ell}(s) + \beta O^{\star}_{t\ell}(s)\right)}}{\sum_{\ell'} \mathrm{e}^{\vartheta\left(c^{y\star}_{t\ell'}(s) + \beta O^{\star}_{t\ell'}(s)\right)}}\right)\right) ds - \\ \sum_t \sum_{\ell} \int \lambda^o_{t\ell}(s) \left(n^{o\star}_{t\ell}(s) - n^{o\star}_t(s) \left(\frac{\mathrm{e}^{\vartheta c^{o\star}_{t\ell}(s)}}{\sum_{\ell'} \mathrm{e}^{\vartheta c^{o\star}_{t\ell'}(s)}}\right)\right) ds - \\ \sum_t \int \nu_t(s_o) \left(n^{o\star}_{t+1}(s_o) - \sum_{\ell} \int\int \frac{1}{\gamma(s_y, s_p)} f\left(\frac{s_o}{\gamma(s_y, s_p)}\right) n^{y\star}_{t\ell}(s_y) \pi^{y\star}_{t\ell}(s_p) ds_p ds_y\right) ds_o - \\ \sum_t \int \kappa^y_t(s) \sum_{\ell} \left(c^{y\star}_{t\ell}(s) - T_\ell s\right) n^{y\star}_{t\ell}(s) ds + \sum_t \int \kappa^o_t(s) \sum_{\ell} \left(c^{o\star}_{t\ell}(s) - T_\ell s\right) n^{o\star}_{t\ell}(s) ds,\end{aligned}$$

where I have kept implicit that

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

B.2 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 $\ell \in \{1, 2, \ldots, L\}$ and $t \in \{0, 1, 2, \ldots\}$. 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 $(\boldsymbol{n}_t^o, \boldsymbol{n}_{\ell t}^o, \boldsymbol{c}_{\ell t}^o)$ for $t \geq 1$. Regarding $\boldsymbol{c}_{lt}^o$, we have

$$\int \kappa_t^o(s) n_{tl}^o(s) \eta(s) ds = \beta_S^{t-1} \beta \int \sum_\ell n_{t-1\ell}^y(s) \delta_{c_{tl}^o} O_{t-1\ell}(s) ds +$$
$$\vartheta \beta \sum_\ell \int \lambda_{t-1\ell}^y(s) n_{t\ell}^y(s) \left(\delta_{c_{tl}^o} O_{t-1\ell}(s) - \sum_{\ell'} \frac{n_{t-1\ell'}^y(s)}{n^y(s)} \delta_{c_{tl}^o} O_{t-1\ell'}(s) \right) ds +$$
$$\vartheta \int n_{tl}^o(s) \eta(s) \left(\lambda_{tl}^o(s) - \sum_\ell \lambda_{t\ell}^o(s) \frac{n_{t\ell}^o(s)}{n_t^o(s)} \right) ds,$$

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 $\lambda_{t-1\ell}^y(s) \equiv \lambda_{t-1}^y(s)$ for all ℓ and t. Hence, the second line is equal to zero. From the expression for $O_{t\ell}$, we further know that

$$\delta_{c_{tl}^o} O_{t-1\ell}(s) = \int \int \eta[e\gamma(s, s_p)] \frac{n_{tl}^o[e\gamma(s, s_p)]}{n_t^o[e\gamma(s, s_p)]} dF(e) \pi_{t-1\ell}^y(s_p) ds_p.$$

Hence, the optimality condition turns into

$$\int \kappa_t^o(s) n_{tl}^o(s) \eta(s) ds = \beta_S^{t-1} \beta \int \eta(q) \frac{n_{tl}^o(q)}{n_t^o(q)} \sum_\ell \int \int \frac{1}{\gamma(s, s_p)} f\left(\frac{q}{\gamma(s, s_p)} \right) n_{t-1\ell}^y(s) \pi_{t-1\ell}^y(s_p) ds_p ds dq +$$
$$\vartheta \int n_{tl}^o(s) \eta(s) \left(\lambda_{tl}^o(s) - \sum_\ell \lambda_{t\ell}^o(s) \frac{n_{t\ell}^o(s)}{n_t^o(s)} \right) ds,$$

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 $n_t^o(q)$. 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 $c_{tl}^o(x) = T_l s$.[79]

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

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

[79]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.

Hence, here as well, this implies $\lambda^y_{lt}(x) \equiv \lambda^y_t(x)$ and $\kappa^y_t(x) = \beta^t_S \equiv \kappa^y_t$. When $\beta_S = \beta$, then $\kappa^y_t = \kappa^o_t$.

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

$$0 = \beta^t_S \beta \int \sum_\ell n^y_{t\ell}(s) \delta_{n^y_{tl}} O_{t\ell}(s) ds + \beta^t_S \int \mathcal{V}^y_t(s)\eta(s)ds -$$
$$\int \left(\lambda^y_{tl}(s)\eta(s) - \vartheta\beta \sum_\ell \lambda^y_{t\ell}(s) n^y_{t\ell}(s) \left[\delta_{n^y_{tl}} O_{t\ell}(s) - \sum_{\ell'} \frac{n^y_{t\ell'}(s)}{n^y(s)} \delta_{n^y_{tl}} O_{t\ell'}(s) \right] \right) ds +$$
$$\int \nu_t(s_o) \int\int \frac{1}{\gamma(s_y, s_p)} f\left(\frac{s_o}{\gamma(s_y,s_p)}\right) \left(\eta(s_y)\pi^y_{tl}(s_p) + \pi^y_{tl}(s_y)\left[\eta(s_p) - \pi^y_{tl}(s_p)\int \eta(\sigma)d\sigma\right]\right) ds_p ds_y ds_o -$$
$$\int \kappa^y_t(s)\left[c^y_{tl}(s) - sT_l\right]\eta(s)ds.$$

Using the spatial equalization of λ^y_{lt}, the second line boils down to $\int \lambda^y_t(s)\eta(s)ds$. Using $\nu^o_t(s) = 0$, the third line vanishes. Finally, from the expression for $O_{tl}(s)$, we have

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

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

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

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 $\kappa^y_t = \beta^t_S$, we obtain

$$c^y_{tl}(x) = xT_l + \beta \int \left(\int \mathcal{V}^o[e\gamma(s,x)]dF(e) - O_{tl}(s) \right) \pi^y_{tl}(s)ds + \mathcal{V}^y_t(x) - \lambda^y_t(x)/\beta^t_S.$$

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

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

for

$$\tau_l(x) \equiv \beta \int\int \left(\mathcal{V}^o[e\gamma(s,x)] - \mathbb{E}[\mathcal{V}^o[e\gamma(s,S^y_\ell)]]\right) dF(e)\pi^y_l(s)ds \tag{39}$$

after substituting the expression for $O_l(s)$. This expression corresponds to (15) in the main text. Using the feasibility condition, it must be that $\mathcal{V}^y(x) - \lambda^y(x) = -\sum_l n^y_l(x)/n^y(x)\tau_l(x)$, and we conclude that[80]

$$c^y_l(x) = xT_l + \tau_l(x) - \sum_\ell \left(\frac{n^y_\ell(x)}{n^y(x)}\right)\tau_\ell(x). \tag{40}$$

In the competitive equilibrium, workers' consumption equate their wage, i.e. $c^y_l(x) = xT_l$. (40) thus yields (14) in the main text. Alternatively, (40) can be rewritten

$$c^y_l(x) = w_l(x) + t_l(x),$$

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

The optimal allocation is given by (39), (40), and the local labor supply,

$$n^{y\star}_\ell(s) = n^y(s)\left(\frac{e^{\vartheta\left(c^{y\star}_\ell(s) + \beta O^\star_\ell(s)\right)}}{\sum_{\ell'} e^{\vartheta\left(c^{y\star}_{\ell'}(s) + \beta O^\star_{\ell'}(s)\right)}}\right), \tag{41}$$

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

which together form a fixed point. This fixed point can be recasted as a function of the local labor supplies only. Hence, existence of the optimal allocation can be shown using similar techniques as in the proof of the first part of Proposition 1. Regarding uniqueness, note that if the local labor supplies are unique, then so are the transfers and consumption. Hence, using similar techniques as in the second part of the proof of Proposition 1, it is possible to show that, if ϑ is small enough, there exists a unique optimal allocation. Finally, consider the case where $T_\ell = \bar{T}$ for all ℓ. It is straighforward to guess and verify that $\tau_l(x) = \tau(x)$ and $c_l^y(x) = x$ is a solution. Hence, when $\boldsymbol{T} = \bar{T}$ and ϑ is small, the unique optimal allocation is the symmetric allocation – as in the laissez-faire equilibrium.

B.3 Linearizing the social optimum

I now approximate the optimal allocation around $\boldsymbol{T} \approx \bar{T}$ under the assumption of ϑ small so that the symmetric allocation is unique when $\boldsymbol{T} = \bar{T}$. Solving for the perturbed equilibrium requires to solve jointly for (39), (40) and (41). I drop the star superscript for notational simplicity.

From (41), we have

$$\left.\frac{\partial n_\ell^y(s)}{\partial \log T_l}\right|_{\boldsymbol{T}=\bar{T}} = \vartheta\left(\frac{n^y(s)}{L}\right)\left.\left\{\frac{\partial c_\ell^y(s)}{\partial \log T_l} + \beta\frac{\partial O_\ell(s)}{\partial \log T_l} - \frac{1}{L}\sum_{\ell'}\left(\frac{\partial c_{\ell'}^y(s)}{\partial \log T_l} + \beta\frac{\partial O_{\ell'}(s)}{\partial \log T_l}\right)\right\}\right|_{\boldsymbol{T}=\bar{T}}. \tag{42}$$

The net change in the learning opportunities of city ℓ is derived as in Section A.4. To continue, we need to derive the net change in consumption. From (40),

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

where $\partial_l x = \partial x/\partial \log T_l|_{\boldsymbol{T}=\bar{T}}$, and I have used $\tau_\ell(s) = 0$ under $\boldsymbol{T} = \bar{T}$ and $\sum_\ell \partial_l\pi_\ell(s) = 0$ from (2). It follows that the perturbed consumption allocation is

$$\partial_l c_\ell^y(x) = x\mathbb{1}\{l=\ell\} + \beta\int \partial_l\pi_\ell^y(s)\left(\gamma(s,x) - \mathbb{E}[\gamma(s,S^y)]\right)ds - \beta\int \mathbb{E}[\gamma(S^y,s_p)]\partial_l\pi_\ell^y(s_p)ds_p. \tag{43}$$

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

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

Combining the expression for $O_\ell(s)$ and $c_\ell(s)$ in (42) yields

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

Guessing and verifying the form of the solution,

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

we end up with the optimal integral equation

$$\begin{aligned}\eta^\star(x) = x + \theta\beta\int\left(\{\gamma(x,s) - \mathbb{E}[\gamma(x,S^y)]\} - \{\mathbb{E}[\gamma(S^y,s)] - \mathbb{E}[\gamma(S^y,S^y)]\}\right)\eta^\star(s)dN^y(s) + \\ \theta\beta\int\left(\{\gamma(s,x) - \mathbb{E}[\gamma(S^y,x)]\} - \{\mathbb{E}[\gamma(s,S^y)] - \mathbb{E}[\gamma(S^y,S^y)]\}\right)\eta^\star(s)dN^y(s).\end{aligned} \tag{44}$$

Finally, using (43) 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,

$$t_l^y(x) \approx T_l + d_l^y(x), \tag{45}$$

where

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

and

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

B.4 Proof of Proposition 5

Using covariances, (44) rewrites

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

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

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

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

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

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

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

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

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

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

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,

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

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

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

Proposition 5.1 To a first order, the optimal spatial allocation is given by

$$n_\ell^{y\star}(s) \approx \frac{n^y(s)}{L} + \theta\left(\frac{n^y(s)}{L}\right)\log\left(\frac{T_\ell}{\bar{T}}\right)\eta^\star(s).$$

Note that $\mathbb{E}[\eta^\star(S^y)] = \int \eta^\star(x) n^y(x)dx = \mathbb{E}[S^y]$. Hence, the optimal number of young workers in city ℓ is

$$N_\ell^{y\star} \approx \frac{1}{L} + \frac{1}{L}\theta \log\left(\frac{T_\ell}{\bar{T}}\right)\mathbb{E}[S^y].$$

In the laissez-faire equilibrium, the number of young workers in city ℓ is

$$N_\ell \approx \frac{1}{L} + \theta\frac{1}{L}\log\left(\frac{T_\ell}{\bar{T}}\right)\mathbb{E}[S^y] + \theta^2\beta\log\left(\frac{T_\ell}{\bar{T}}\right)\mathbb{E}[\Theta(S^y)].$$

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 To a first order, the optimal skill density in city ℓ is

$$\pi_\ell^{y\star}(s) \approx n^y(s) + \theta n^y(s)\log\left(\frac{T_\ell}{\bar{T}}\right)(\eta^\star(s) - \mathbb{E}[S^y]).$$

In particular, the optimal average skill in city ℓ is

$$\mathbb{E}_\ell[S^{y\star}] \approx \mathbb{E}[S^y] + \theta\log\left(\frac{T_\ell}{\bar{T}}\right)\mathbb{C}\text{ov}[\eta^\star(S^y), S^y]$$

In the laissez-faire equilibrium, we had

$$\mathbb{E}_\ell[S^y] \approx \mathbb{E}[S^y] + \theta\log\left(\frac{T_\ell}{\bar{T}}\right)\mathbb{C}\text{ov}[\eta(S^y), S^y].$$

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.[81] 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,

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

Taking a similar derivative in the decentralized equilbrium returns

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

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,

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

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

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

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.

Conclusion of Proposition 5 I now conclude by showing that the optimal allocation implies a larger aggregate stock of human capital. Following exactly the same derivation as in Section A.6, the optimal average skill of

[81] 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 $g' > h'$.

old workers read

$$\mathbb{E}[S^{o\star}] \approx \mathbb{E}[S^o] + \theta \mathbb{V}\text{ar}[\log T_\ell]\Omega^\star \quad \text{where} \quad \Omega^\star = \int\int \gamma_{12}(s,s_p)\Psi^\star(s)\Psi^\star(s_p)dsds_p, \tag{46}$$

for $\Psi^\star(x) = (1 - N(x))\left(\mathbb{E}[\eta^\star(S^y) \mid S^y \geq x] - \mathbb{E}[\eta^\star(S^y)]\right)$ defined analogously to Section A.9. Hence, comparing (46) with (12), we have that

$$\mathbb{E}[S^{o\star}] - \mathbb{E}[S^o] \approx \theta \mathbb{V}\text{ar}[\log T_\ell] \int\int \gamma_{12}(s,s_p)\left[\Psi^\star(s)\Psi^\star(s_p) - \Psi(s)\Psi(s_p)\right] dsds_p.$$

Suppose for an instant that γ supermodular implies $\Psi^\star(s) > \Psi(s)$, and inversely, γ submodular implies $\Psi^\star(s) < \Psi(s)$. Then, it immediately follows that $\mathbb{E}[S^{o\star}] > \mathbb{E}[S^o]$.

From the definition of Ψ, we have

$$\begin{aligned}\Psi^\star(x) - \Psi(x) &\propto \int_x [\chi^\star(s) - \chi(s)]dN^y(s)\\ &= \int_x [\eta^\star(s) - \eta(s)]dN^y(s) - \bar{N}^y(x)\mathbb{E}[\eta^\star(S) - \eta(S)],\end{aligned}$$

and we know that γ supermodular (submodular) implies $\eta'^\star > \eta'$ ($\eta'^\star < \eta'$).

(i) we know that $\eta'^\star$ is higher than η' (ii) we do not know anything about $\eta^\star - \eta$, in fact it is likely that $\eta^\star < \eta$ for some; this should, however, be fine, because Ψ seems to be more about η' than η

Rewrite instead

$$\begin{aligned}\Psi(x) = \int_x \chi(s)dN^y(s) &= \int_x \{\eta(s) - \mathbb{E}[\eta(S)]\}dN^y(s)\\ &= \int_x\int \{\eta(s) - \eta(\sigma)\}dN^y(\sigma)dN^y(s)\\ &= \int_x\int\int_\sigma^s \eta'(x)dxdN^y(\sigma)dN^y(s)\\ &= \int_x\int \left\{\mathbb{1}\{s \geq \sigma\}\int_\sigma^s \eta'(x)dx - \mathbb{1}\{s < \sigma\}\int_s^\sigma \eta'(x)\right\} dxdN^y(\sigma)dN^y(s)\\ &= \int_x \left\{\int^s\int_\sigma^s \eta'(x)dxdN^y(\sigma) - \int_s\int_s^\sigma \eta'(x)dxdN^y(\sigma)\right\} dN^y(s)\end{aligned}$$

C Extensions

C.1 Production complementarities

In this Appendix, I show that Proposition 2 and 3 hold allowing for production complementarities.

C.1.1 Model

The framework is similar to that of Section 2.1 with one modification. Young and old workers differ in their employment opportunities. To earn an income, young workers need to be employed by the competitive local firms. These firms produce the freely-trade good with a CES production function,

$$R(T_\ell, \boldsymbol{n}) = T_\ell \left(\int a(s)n(s)^\sigma ds\right)^{\frac{1}{\sigma}},$$

where $\sigma \leq 1$ is the elasticity of substitution. Let $R(\boldsymbol{n})$ denote the CES aggregator, $R(\boldsymbol{n}) = R(T_\ell, \boldsymbol{n})/T_\ell$. In contrast to young workers, old workers have access to a backyard technology. Specifically, an old worker living in city ℓ with skill s produces sT_ℓ of the freely traded good.[82] This framework nests the model of Section 2.1 under $\sigma \to 1^-$.

[82] I differentiate young and old work opportunities to simplify the analysis by decoupling the yound and old problem.

C.1.2 Equilibrium

The problem of young and old workers is unchanged up to the local wages. Local wages for young workers clear the local labor markets by equating the local labor supply and demand. They are given by

$$w_\ell(s) = T_\ell a(s) \left(\frac{R[\boldsymbol{n}_\ell^y(s)]}{n_\ell^y(s)} \right)^{1-\sigma}. \tag{47}$$

The equilibrium is given by (3), (4), (5), (6), and (47). As in the original problem, old workers are passive, and a unique equilibrium exists when $\vartheta\beta$ is small.

Symmetry Suppose that $T_\ell = \bar{T}$ for all ℓ. Then, there always exists a symmetric equilibrium characterized by

$$n_\ell^y(s) = \frac{n^y(s)}{L} \quad \text{and} \quad w_\ell(s) = \bar{T} a(s) \left(\frac{R(\bar{\boldsymbol{n}})}{\bar{n}(s)} \right)^{1-\sigma}.$$

Define $w(s) = w_\ell(s)/\bar{T}$ in the symmetric equilibrium. This symmetric equilibrium is guaranteed to be unique when $\vartheta\beta$ is small enough. For the rest of the analysis, I suppose that wages are increasing in skills in the symmetric equilibrium. Since I do not impose any conditions on γ, this assumption is wlog.

Assumption 3.
The functions a and n^y are such that wages are increasing in s in the symmetric equilibrium, $w'(s) > 0$.[83]

C.1.3 Proposition 2 with production complementarities

From (5), the partial derivative of the young allocation w.r.t. T_l is

$$\begin{aligned}
\Delta_l n_\ell^y(s) &\equiv \left.\frac{\partial \log n_\ell^y(s)}{\partial \log T_l}\right|_{\boldsymbol{T}=\bar{T}} \\
&= \vartheta \left(\left.\frac{\partial w_\ell(s)}{\partial \log T_l}\right|_{\boldsymbol{T}=\bar{T}} - \sum_{\ell'} \frac{n_{\ell'}^y(s)}{n^y(s)} \left.\frac{\partial w_{\ell'}(s)}{\partial \log T_l}\right|_{\boldsymbol{T}=\bar{T}} \right) + \vartheta\beta \left(\left.\frac{\partial O_\ell(s)}{\partial \log T_l}\right|_{\boldsymbol{T}=\bar{T}} - \sum_{\ell'} \frac{n_{\ell'}^y(s)}{n^y(s)} \left.\frac{\partial O_{\ell'}(s)}{\partial \log T_l}\right|_{\boldsymbol{T}=\bar{T}} \right).
\end{aligned}$$

The second term is similar to the baseline economy of Section 2.1. From (47), the partial of wages w.r.t. $\log T_l$ is

$$\left.\frac{\partial w_\ell(s)}{\partial \log T_l}\right|_{\boldsymbol{T}=\bar{T}} = \mathbb{1}\{\ell = l\}\bar{T} w(s) + (1-\sigma)\bar{T} w(s) \left(\int \frac{w(s')n^y(s')}{R(\boldsymbol{n}^y)} \Delta_l n_\ell^y(s') ds' - \Delta_l n_\ell^y(s) \right).$$

Combining the static and dynamic response, and solving for $\Delta_l n_\ell^y(s)$, we obtain

$$\begin{aligned}
\Delta_l n_\ell^y(s) = {}& \frac{\theta w(s)}{1+\theta(1-\sigma)w(s)} \left(I_{l\ell} + (1-\sigma) \int \frac{w(s')n^y(s')}{R(\boldsymbol{n}^y)} \Delta_l n_\ell^y(s') ds' \right) + \\
& \frac{\theta\beta}{1+\theta(1-\sigma)w(s)} \int \left(\gamma(s, s_p) - \mathbb{E}[\gamma(s, S^y)]\right) \Delta_l n_\ell^y(s_p) dN^y(s_p).
\end{aligned}$$

We can now guess a particular form for $\Delta_l n_\ell^y(s)$. Specifically, $\Delta_l n_\ell^y(s) = \theta I_{l\ell}\eta(s)$, for some function $\eta(s)$. Plugging in the original expression verifies the guess and yields the implicit expression for $\eta(s)$:

$$\eta(s) = \frac{w(s)\Sigma}{1+\theta(1-\sigma)w(s)} + \frac{\theta\beta\Theta(s)}{1+\theta(1-\sigma)w(s)}, \tag{48}$$

for

$$\Sigma = 1 + \theta(1-\sigma) \int \left(\frac{w(s')}{R(\boldsymbol{n}^y)} \right) \eta(s') dN^y(s'), \tag{49}$$

an implicit constant, and

$$\Theta(s) = \int \left(\gamma(s, s') - \mathbb{E}[\gamma(s, S^y)]\right) \eta(s') dN^y(s') = \mathbb{C}\mathrm{ov}[\gamma(s, S^y), \eta(S^y)]. \tag{50}$$

[83] This is clearly satisfied if a is increasing and n^y is (weakly) decreasing. As such, Assumption 3 necessarily holds for a an iso-elastic function of s and n^y a Pareto distribution.

an implicit function. The function $\Theta(s)$ is the same as in the main text, up to the value of the implicit function η. However, there are two novelties in (48). First, the presence of Σ, which captures the static production complementarities. If $\Sigma > 1$, production complementarities generate further agglomeration. Second, the three effects are normalized by a term which reflects the effect of other workers with similar skills on workers' wages. When $\sigma \to 1^-$, $\Sigma \to 1$, and $\eta(s)$ solves (8).

Despite these differences, I now show that Lemma 1 and Proposition 2 extends to $\sigma < 1$ when θ is not too large.

Lemma C.1.
Suppose that θ is small such that $\theta^2 \approx 0$. Then, if

$$\theta\beta\mathbb{C}ov\left[\gamma_1(s, S^y), \frac{w(S^y)}{1+\theta(1-\sigma)w(S^y)}\right] > -w'(s),$$

for all s, SPAM prevails: $\eta'(s) > 0$ for all s.

Proof. Differentiate (48) w.r.t. s to get

$$\eta'(s) \propto w'(s)\Sigma + \theta\beta\mathrm{Cov}[\gamma_1(s, S^y), \eta(S^y)] + \theta^2\beta(1-\sigma)\Theta'(s)w(s) - \theta^2\beta(1-\sigma)\Theta(s)w'(s).$$

When $\theta^2 \approx 0$, this expression simplifies to

$$\eta'(s) \propto w'(s)\Sigma + \theta\beta\Sigma\mathrm{Cov}\left[\gamma_1(s, S^y), \frac{w(S^y)}{1+\theta(1-\sigma)w(S^y)}\right].$$

after plugging in the expression for $\eta(S^y)$. Furthermore, plugging the expression for η in Σ and using $\theta^2 \approx 0$ yields[84]

$$\Sigma = \frac{1}{1-e} \quad \text{for } e \equiv \int \frac{w(s)n^y(s)}{R(\boldsymbol{n}^y)}\frac{\theta(1-\sigma)w(s)}{1+\theta(1-\sigma)w(s)}ds \in (0,1). \tag{51}$$

It is then automatic to see that the condition stated in Lemma C.1 implies $\eta'(s) > 0$. □

The condition stated in Lemma C.1 nests that of Lemma 1 under $a(s) = s$ and $\sigma \to 1^-$, such that $w'(s) = 1$. When $\sigma < 1$, it necessarily holds if the learning technology is supermodular. When the technology is submodular, Lemma C.1 imposes a bound on how submodular the function can be.

With Lemma C.1, it is now possible to generalize Proposition 2 to the case with $\sigma < 1$.

Proposition C.1.
Suppose that θ is small such that $\theta^2 \approx 0$ and Lemma C.1 holds. Then,

1. *If $\gamma(s, \cdot)$ is increasing for all s, (un)productive cities are (smaller) larger when interactions are spatially segmented; the converse holds when $\gamma(s, \cdot)$ is decreasing;*
2. *If γ is supermodular, the average skill of young workers in (un)productive cities is (smaller) larger when interactions are spatially segmented, which leads to greater between-city wage variance amongst young workers; the converse holds when γ is submodular.*

Proof. Let $\tilde{x}$ denote the variable x when interaction are not spatially segmented. Then, the equilibrium solves

$$\tilde{\eta}(s) = \frac{w(s)\tilde{\Sigma}}{1+\theta(1-\sigma)w(s)},$$

where $\tilde{\Sigma}$ is defined analogously to (49) w.r.t. $\tilde{\eta}(s)$. Solving for $\tilde{\Sigma}$ implies $\tilde{\Sigma} = 1/(1-e) \approx \Sigma$, where the approximation is w.r.t. $\theta^2 \approx 0$. Accordingly, the difference between the equilibrium with and without spatially segmented interaction is

$$\eta(s) - \tilde{\eta}(s) = \frac{w(s)\left(\Sigma - \tilde{\Sigma}\right)}{1+\theta(1-\sigma)w(s)} + \frac{\theta\beta\Theta(s)}{1+\theta(1-\sigma)w(s)} \approx \frac{\theta\beta\Theta(s)}{1+\theta(1-\sigma)w(s)}.$$

Proposition C.1.1 immediately follows: $\gamma(s, \cdot)$ increasing implies $\Theta(s) > 0$ (Lemma C.1), and therefore $\eta(s) > \tilde{\eta}(s)$. Meanwhile, the segmentation of interactions amplifies spatial inequality to the extent that $\eta(s) - \tilde{\eta}(s)$ is increasing. Differentiating w.r.t. s, we get

$$\eta'(s) - \tilde{\eta}'(s) \propto w'(s)\left(\Sigma - \tilde{\Sigma}\right) + \beta\Theta'(s)\left(\theta + \theta^2(1-\sigma)w(s)\right) - \theta^2\beta(1-\sigma)\Theta(s)w'(s) \approx \theta\beta\Theta'(s).$$

[84] The constant $e \in (0,1)$ since constant returns to scale imply $\int w(s)n(s)ds = R(\boldsymbol{n})$.

Proposition C.1.2 then follows: γ supermodular (submodular) implies $\Theta' > 0$ ($\Theta' < 0$), and therefore $\eta'(s) > \tilde{\eta}'(s)$ for all s. □

C.1.4 Proposition 3 with production complementarities

Proposition 3 automatically generalizes to the presence of static complementarities – conditional on Lemma C.1 holding. To see that, three observations are useful. First, the expression for the average skill is the same regardless of σ. Second, the derivation of the second-order approximation is independent to the definition of η. Third, the proof of Proposition 3 only relies on η increasing.

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,

$$V^o_l(s, \boldsymbol{\varepsilon}) = \max_\ell \; \frac{y_{l\ell}(s)}{P_\ell} + \varepsilon_\ell - \kappa^o_{l\ell}.$$

Letting $m^o_l(s)$ denote the spatial distribution of old workers before their new location choice, the spatial allocation of workers is given by

$$n^o_{l\ell}(s) = m^o_l(s) \left(\frac{\mathrm{e}^{\vartheta\left(y^o_{l\ell}(s)/P_\ell + B^o_\ell - \kappa^o_{l\ell}\right)}}{\sum_{\ell'} \mathrm{e}^{\vartheta\left(y^o_{l\ell'}(s)/P_{\ell'} + B^o_{\ell'} - \kappa^o_{l\ell'}\right)}} \right), \tag{52}$$

where $n^o_{l\ell}(s)$ 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

$$n^o_\ell(s) = \sum_l n^o_{l\ell}(s),$$

and the mass of old workers in ℓ is $N^o_\ell = \int n^o_\ell(s) ds$. Finally, let $\mathcal{V}^o_l(s) \equiv \mathbb{E}[V^o_l(s, \boldsymbol{\varepsilon}) \mid s, l]$ be the expected utility before the realization of the idiosyncratic preferences $\boldsymbol{\varepsilon}$, given by

$$\mathcal{V}^o_l(s) = \frac{1}{\vartheta} \log \left(\sum_\ell \mathrm{e}^{\vartheta\left(y^o_{l\ell}(s)/P_\ell + B^o_\ell - \kappa^o_{l\ell}\right)} \right).$$

Young workers Young workers solve a dynamic location choice problem which reads

$$V^y_l(s, \boldsymbol{\varepsilon}) = \max_\ell \; \frac{y^y_{l\ell}(s)}{P_\ell} + \varepsilon_\ell - \kappa^y_{l\ell} + \beta \int \int \mathcal{V}^o_\ell \left[e \gamma(s, s_p)\right] dF(e) \pi_\ell(s_p) ds_p,$$

where $y^y_{l\ell}(s) = w_\ell(s) + \tau^y_{l\ell}(s) - t$. In a similar fashion as for the old workers, the spatial allocation of young workers is given by

$$n^y_{l\ell}(s) = m^y_l(s) \left(\frac{\mathrm{e}^{\vartheta\left(y^y_{l\ell}(s)/P_\ell + B^y_\ell - \kappa^y_{l\ell} + \beta O_\ell(s)\right)}}{\sum_{\ell'} \mathrm{e}^{\vartheta\left(y^y_{l\ell'}(s)/P_{\ell'} + B^y_{\ell'} - \kappa^y_{l\ell'} + \beta O_{\ell'}(s)\right)}} \right). \tag{53}$$

I define $n^y_\ell(s)$ and N^y_ℓ in the same manner as $n^o_\ell(s)$ and N^o_ℓ.

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[85]

$$p_\ell = \left(\frac{(1-\alpha)\, Y_\ell}{\mathcal{H}} \right)^{\frac{1}{1+\delta}}, \tag{54}$$

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 (21).

Skill density The within-city skill density of location ℓ must be consistent with workers' location choices, given by

$$\pi_\ell(s) = \frac{\sum_a n_\ell^a(s)}{\sum_a N_\ell^a}. \tag{55}$$

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

$$M_\ell^o(s) = \int \int n_\ell^y(x) \pi_\ell(y) F\left(\frac{s}{\gamma(x,y)} \right) dy dx. \tag{56}$$

From here, one can define the location-specific measure as $m_\ell^o(s) = \partial M_\ell^o(s)/\partial s$ and the aggregate old skill density as $n^o(s) = \sum_\ell m_\ell^o(s)$. Finally, given the birth process, the location-specific distribution of young workers is $m_\ell^y(s) = N_\ell^y n^y(s)$.

Definition D.1 (Steady state equilibrium)**.**
A steady state equilibrium is a collection of young and old spatial allocation, $s \to n_{l\ell}^y(s)$ and $s \to n_{l\ell}^o(s)$ for each $l, \ell \in \{1, 2, \dots, L\}$, city-specific skill distribution, $s \to m_\ell^y(s)$ and $s \to m_\ell^o(s)$ 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* (53) *and* (52) *respectively;*
2. *The local skill densities are consistent with workers' location decisions, given by* (55)*;*
3. *Given young workers' location decisions, the city-specific young and old skill distribution are given by $m_\ell^y(s) = N_\ell^y n^y(s)$ and* (56) *respectively;*
4. *Given workers' location decisions, housing prices solve* (54)*.*

D.3 Algorithm

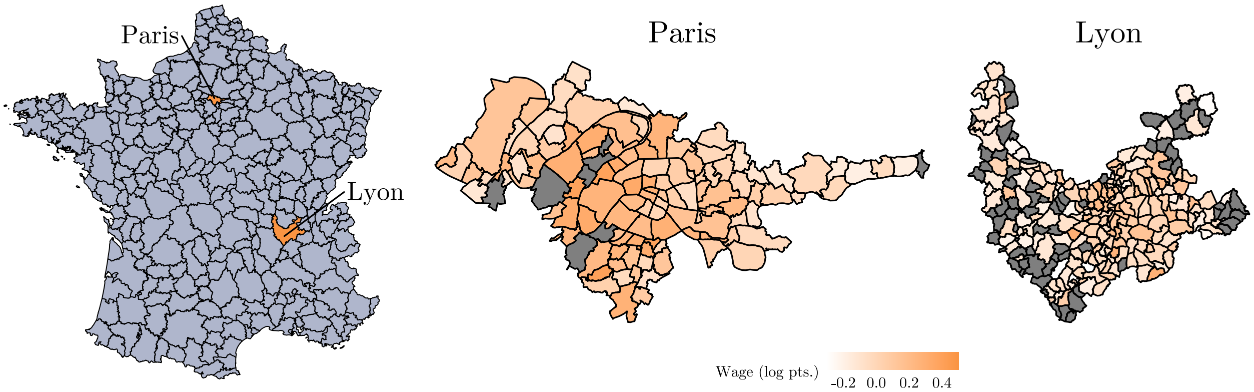
I use two different algorithms to solve for the steady state equilibrium depending on whether I am estimating the model or computing counterfactuals. Both algorithms are constituted of an inner loop and an outer loop. The inner loops are identical. Given city fundamentals, $\boldsymbol{T}$ and $\boldsymbol{B}$, housing prices $\boldsymbol{p}$, and skill distributions $\boldsymbol{m}^y$ and $\boldsymbol{m}^o$, they iterate on (53) and (52) 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 (54) 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 (63).

Counterfactual In the counterfactuals, city TFP and amenities are taken as given. The outer loop therefore iterates on (54) to solve for the housing prices, as well as on $m_\ell^y(s) = N_\ell^y n^y(s)$ and (55) to solve for $\boldsymbol{m}^y$ and $\boldsymbol{m}^o$.

[85] 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.

Figure E.1: The geography of commuting zones and municipalities



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

As described in the main text, the French matched employer-employee dataset comes in two format. The first format is a long-panel that tracks the entire labor market history of 4% of the French workforce. The second format is a repeated short-panel (two years) that provides information on the job providing the main source of labor income within each year for the universe of the French workforce.[86] I apply a first set of similar restrictions on both datasets:

- Exclude workers younger than 25 and older than 55 year old;
- Exclude workers employed in the private sector;
- Exclude the agriculture, education and health industries;
- Keep only workers employed full-time;[87]
- Truncate the yearly wage distribution at the 5% and 99.9%.

The long-panel comes at the job-spell level, where a job-spell is defined as a year-establishment-occupation. Workers are observed multiple times within the same year if they switch jobs. Since the focus of this paper is on middle- and long-term wage growth, I aggregate the long-panel at the yearly level. To do so, I keep for each worker the job that provides the highest earnings within the year. I exclude from the sample jobs that last less than 30 days, and I also exclude workers that have been unemployed for more than three consecutive years.

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.

[86]In the short-panel, the worker ID are shuffled every year so that it is not possible to link workers for more than two periods.

[87]Full-time jobs are identified first by their contracts, and second by the fact that the individuals work at the job more than three hours per worked day.

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 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 (19) identifies the learning technology. I do under the assumption of $g_{12} = 0$ to reduce the length of the derivation. With a slight abuse of notation, let s_{it} denote $\log s_{it}$. Then, the old skill of a worker i who starts with skill s_i^y and meets with a partner with skill s_i^p is

$$s_i^o = g_0 + g_1 s_i^y + g_2 s_i^p + e_i. \tag{57}$$

Equation (57) always rewrite

$$s_i^o = g_0 + g_1 s_i^y + g_2 \mathbb{E}\left[s_j \mid \ell_i^y\right] + e_i + \nu_i, \tag{58}$$

where ℓ_i^y is the city where i works when young, and $\nu_i \equiv g_2\left(s_i^p - \mathbb{E}[s_j \mid \ell_i^y]\right)$. Equation (58) has two error terms. First, e_i captures the learning sources other than interactions. Second, ν_i measures the difference between the particular interaction experienced by i and the average interaction in city ℓ_i^y. If the two error terms are uncorrelated with s_i^y and $\mathbb{E}[s_j \mid \ell_i^y]$, then this justifies estimating the local projection (58) by OLS.

Assumption 6 guarantees the two error terms are orthogonal to the regressors. First, Assumption 6.2 implies that $\mathbb{E}[s_i e_i] = \mathbb{E}[\mathbb{E}[s_j \mid \ell_i^y] e_i] = 0$. Second, Assumption 6.1 implies that ν_i is uncorrelated with the regressors. To see why, note that

$$\mathbb{E}[\nu_i] = \mathbb{E}[\mathbb{E}[\nu_i \mid s_i^y, \ell_i^y]] = \mathbb{E}[g_2 \mathbb{E}[s_i^y - \mathbb{E}[s_j \mid \ell_i^y] \mid s_i^y, \ell_i^y]] = 0,$$

where the first equality follows from the L.I.E., the second from the definition of ν_i, and the third from $\mathbb{E}[s_j^p \mid s_i^y, \ell_i 6y] = \mathbb{E}[s_j \mid \ell_i^y]$ when interactions are random within cities. A similar argument implies $\mathbb{E}[\nu_i s_i^y] = \mathbb{E}[\nu_i \mathbb{E}[s_j \mid \ell_i^y]] = 0$.

These arguments justify estimating the learning technology through the local projection

$$s_i^o = \alpha + \beta s_i^y + \gamma \mathbb{E}[s_j \mid \ell_i^y] + u_i. \tag{59}$$

Is (59) exposed to the standard reflection problem (Manski, 1993)? In particular, when estimating the static framework $s_i = \alpha + \beta \mathbb{E}[s_j \mid \ell_i^y] + u_i$ by OLS, then $\hat{\beta} = 1$ regardless of the presence of peer effects. I now show that (59) does not suffer from this problem, and in particular, that $\hat{\beta} = g_2$.

Proposition E.1.
If s_i and $\mathbb{E}_{\ell_i^y}[s_j]$ are not colinear, (59) *yields $\hat{\gamma} = g_2$.*

Proof. By Frisch-Waugh–Lovell, we know that estimating

$$\epsilon_i = a + b\nu_i + \zeta_i \tag{60}$$

by OLS generates $\hat{b} = \hat{\gamma}$, where ϵ_i and ν_i are defined by

$$s_i^o = \mathtt{a} + \mathtt{b} s_i^y + \epsilon_i, \tag{61}$$

$$\mathbb{E}_{\ell_i^y}[s_j] = \mathrm{a} + \mathrm{b} s_i^y + \nu_i. \tag{62}$$

Using standard OLS algebra, the coefficient estimated by (61) are

$$\hat{\mathtt{a}} = \mathbb{E}[s_i^o] - \hat{\mathtt{b}}\mathbb{E}[s_i^y] \quad ; \quad \hat{\mathtt{b}} = \frac{\mathbb{C}\mathrm{ov}\left(s_i^y, s_i^o\right)}{\mathbb{V}\mathrm{ar}[s_i^y]}.$$

Using the learning process (57) yields

$$\hat{\mathsf{b}} = g_1 + \tilde{b} g_2 \qquad ; \qquad \epsilon_i = \tilde{a} + g_2 \left(s_i^p - \tilde{b} s_i^y \right) + e_i$$

for $\tilde{a}$ some constant and

$$\tilde{b} \equiv \frac{\mathbb{C}\mathrm{ov}\left(s_i^y, s_i^p\right)}{\mathbb{V}\mathrm{ar}[s_i^y]} = \frac{\mathbb{E}\left[\mathbb{C}\mathrm{ov}\left(s_i^y, s_i^p \mid \ell_i^y\right)\right]}{\mathbb{V}\mathrm{ar}[s_i^y]} + \frac{\mathbb{C}\mathrm{ov}\left(\mathbb{E}[s_j^y \mid \ell_i^y], \mathbb{E}[s_j \mid \ell_i^y]\right)}{\mathbb{V}\mathrm{ar}[s_i^y]} = \frac{\mathbb{C}\mathrm{ov}\left(\mathbb{E}[s_j^y \mid \ell_i^y], \mathbb{E}[s_j \mid \ell_i^y]\right)}{\mathbb{V}\mathrm{ar}[s_i^y]},$$

where the second equality follows from the law of total covariance and the last from the random matching assumption within cities. Likewise, the slope estimated by (62) is

$$\hat{\mathrm{b}} = \frac{\mathbb{C}\mathrm{ov}\left(s_i^y, \mathbb{E}\left[s_j \mid \ell_i^y\right]\right)}{\mathbb{V}\mathrm{ar}\left[s_i^y\right]} = \frac{\mathbb{C}\mathrm{ov}\left(\mathbb{E}\left[s_j^y \mid \ell_i^y\right], \mathbb{E}\left[s_j \mid \ell_i^y\right]\right)}{\mathbb{V}\mathrm{ar}\left[s_i^y\right]} = \tilde{b} \qquad ; \qquad \nu_i = \mathbb{E}[s_j \mid \ell_i^y] - \hat{\mathrm{a}} - \tilde{b} s_i^y$$

Hence, altogether, the estimated slope of (60) is

$$\hat{b} = \frac{\mathbb{C}\mathrm{ov}[\epsilon_i, \nu_i]}{\mathbb{V}\mathrm{ar}[\nu_i]} = g_2 \left(\frac{\mathbb{C}\mathrm{ov}\left[s_i^p - \tilde{b} s_i^y, \nu_i\right]}{\mathbb{V}\mathrm{ar}[\nu_i]} \right).$$

Expanding the numerator, using the definition of $\tilde{b}$ and ν_i, and the law of total covariance, it is easy to conclude that $\mathbb{C}\mathrm{ov}(s_i^p - \tilde{b} s_i^y, \nu_i) = \mathbb{V}\mathrm{ar}(\nu_i)$. Hence, $\hat{b} = g_2$ if $\mathbb{V}\mathrm{ar}[\nu_i] \neq 0$, which necessarily holds if s_i^y and $\mathbb{E}[s_j \mid \ell_i^y]$ are not colinear. $\square$

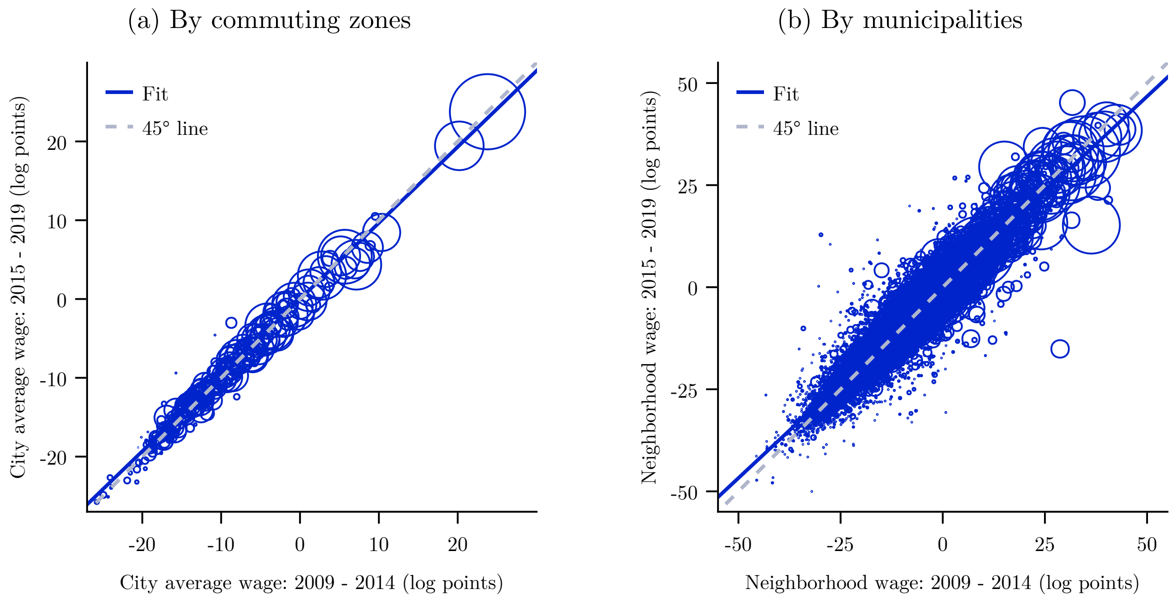
E.4 Additional Tables and Figures

Table E.1: The returns to local interactions (robustness)

Dep. variable	5-year wage				3-year wage	Δ 3-year wage	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	OLS	2SLS	OLS	OLS	OLS	OLS	2SLS
Skill (g_1)	1.003 (0.002)	1.023 (0.003)	0.940 (0.004)	0.975 (0.002)	0.978 (0.002)		
City skill (g_2)	0.119 (0.010)	0.110 (0.005)			0.053 (0.003)		
Skill × city skill (g_{12})	0.065 (0.019)	0.091 (0.007)			0.053 (0.006)		
Neighborhood skill			0.063 (0.005)	0.037 (0.008)			
Wage × neighborhood skill			0.083 (0.019)				
Establishment wage				0.072 (0.006)			
Firm wage				0.011 (0.006)			
Δ Skill (g_1)						-0.080 (0.009)	0.040 (0.098)
Δ Local skill (g_2)						0.042 (0.015)	0.040 (0.023)
Δ Skill × local skill (g_{12})						0.280 (0.017)	0.161 (0.078)
Controls							
Worker-level	·	✓	✓	✓	·	·	·
Distance work-residence	·	·	✓	✓	·	·	·
Instrument							
Skill ∼ past skill	·	✓	·	·	·	·	·
Δ Skill ∼ past Δ skill	·	·	·	·	·	·	✓
Fixed effects							
Year	✓	✓	·	·	✓	✓	✓
Year × residence × city	·	·	✓	✓	·	·	·
Year × occupation × industry × city	·	·	✓	·	·	·	·
Obs.	971,401	971,401	971,401	508,236	520,938	520,938	520,938
Weights	City size^{-1}	Unit	Unit	Unit	Unit	Unit	Unit
Period	2009-2019	2009-2019	2009-2019	2009-2019	2002-2019	2002-2019	2002-2019
F-stat.	·	103,412	·	·	·	·	20,308

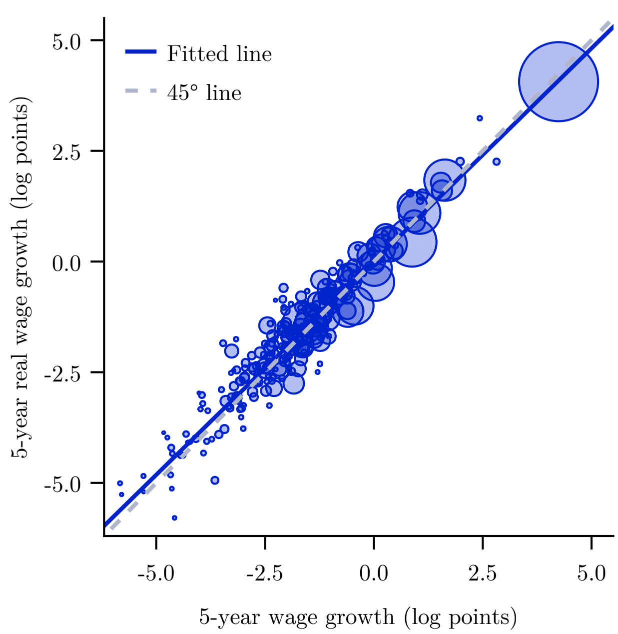
Standard errors clustered at the city level for column 1 and 2, at the neighborhood of work for 3 and 4, and at the worker level for 5 to 7. Sample includes all workers between 25 and 40 year old. Worker-level controls include age, job, occupation, and industry tenure fixed effects, and a dummy for employer switching. Occupations and industry measured at the one-digit level. First column weights observation by the inverse of the size of the city. Second column instrument present skills by the skills in the previous year. To introduce worker F.E., the last three columns extend the sample from 2002 to 2019 and looks at 3-year skill growth. Column 6 estimates the local projection in first difference to remove the worker F.E., and column 7 instruments skill change by past skill change to take into account the autocorrelation in the error term.

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.1). The marker size is proportional to the size of the commuting zone.

Figure E.4: The returns to local interactions by time horizon

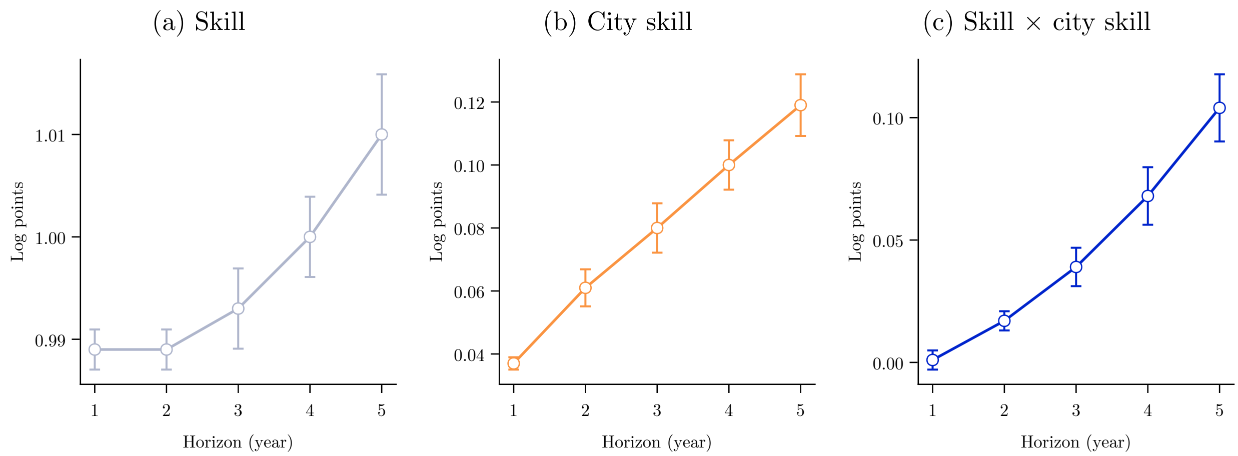


Figure E.5: The returns to local interactions by age

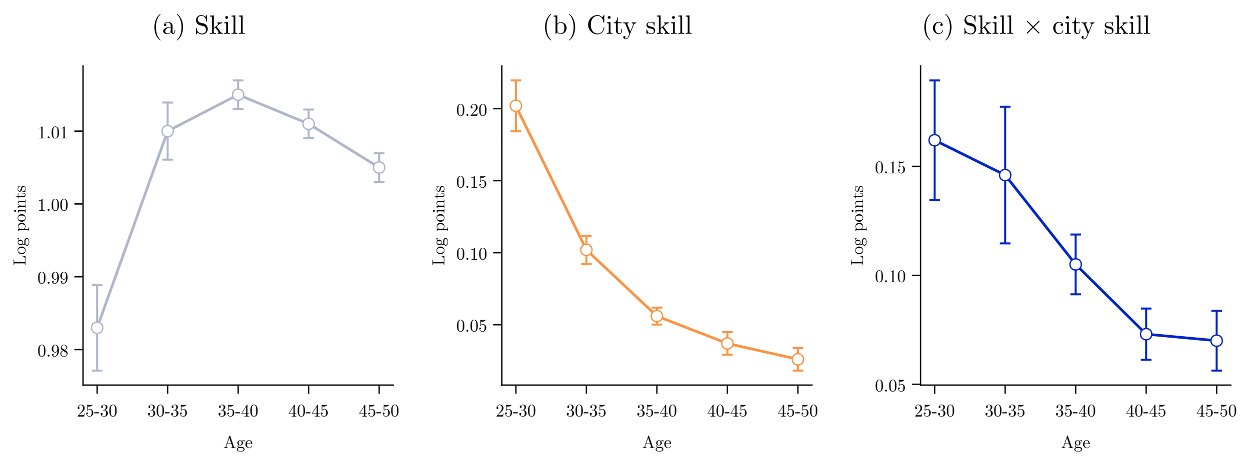


Figure E.6: The returns to local interactions by city size

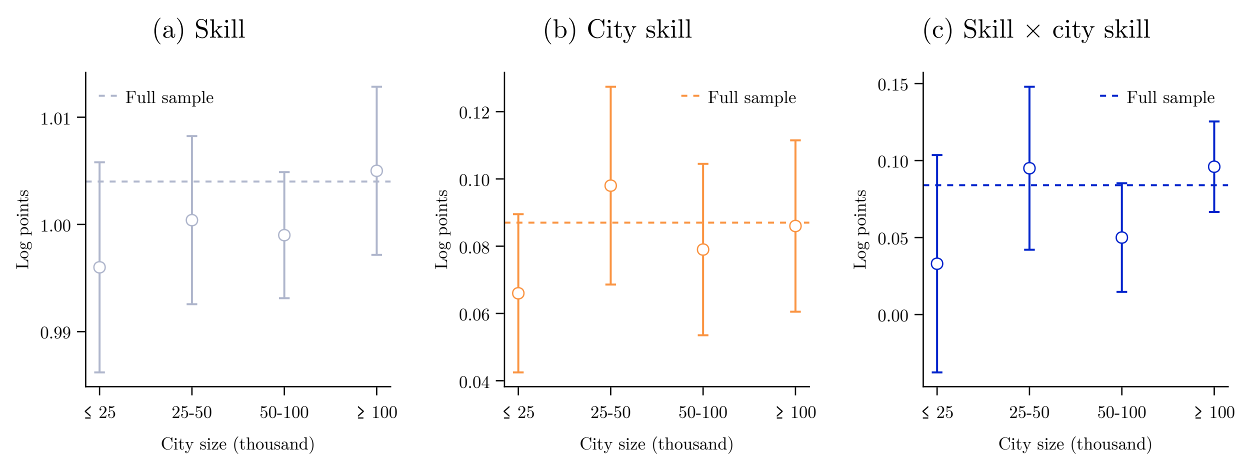


Figure E.7: The returns to local interactions by city inequality

(a) Skill (b) City skill (c) Skill $\times$ city skill

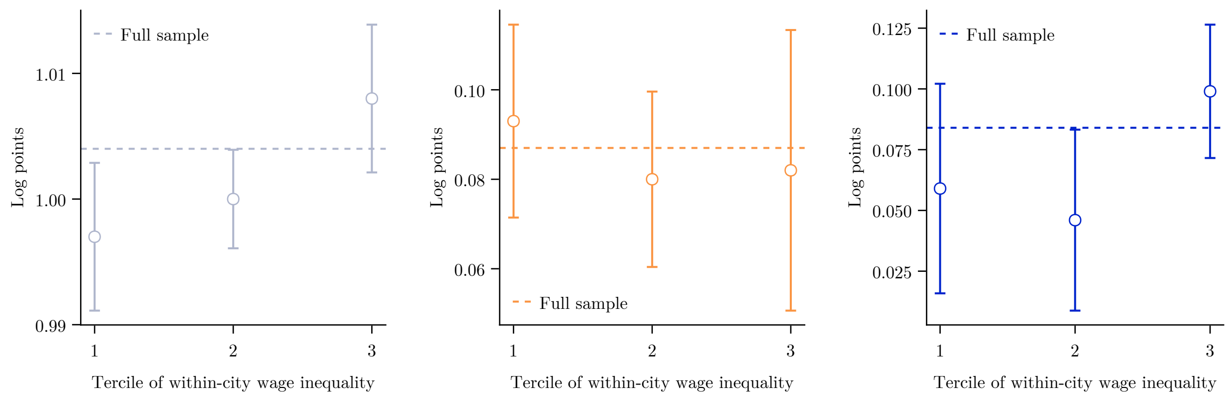
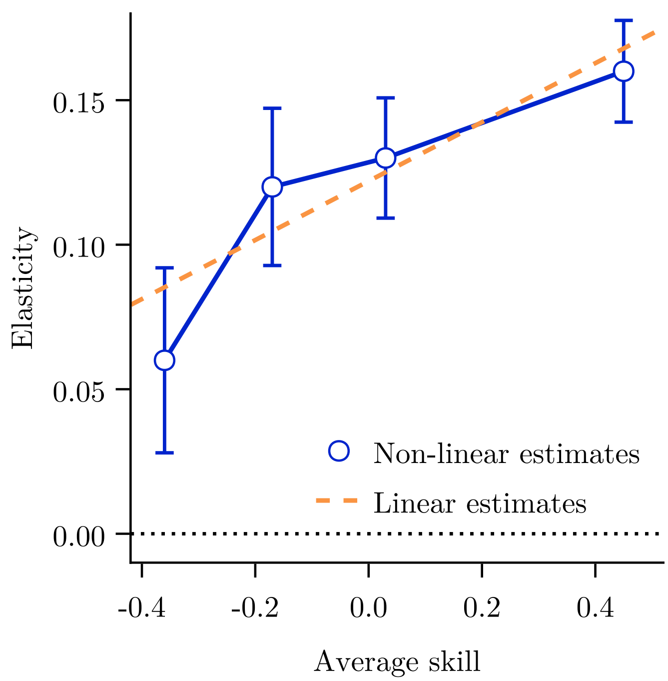
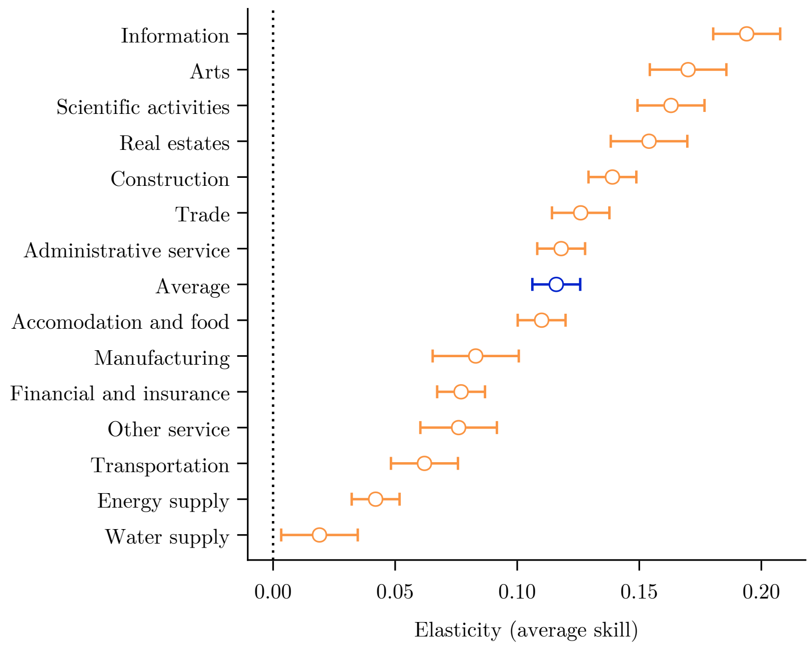


Figure E.8: Non-parametric estimation of the returns to local interactions



Note: estimates produced from a non-parametric version of (19), $\log s_{it+1} = \alpha_t + \beta \log s_{it} + \gamma_q \overline{\log s}_{\ell_{it}} + u_{it}$, for q the quartile of workers' current skills. The figure displays $\{\hat{\gamma}_q\}_q$ (blue circles) along with 10% confidence intervals, and the parametric fit produced by (19) (orange line).

Figure E.9: The returns to local interactions by industry



Note: estimates produced from an industry-specific version of (19), $\log s_{it+1} = \alpha_{t\iota} + \beta_\iota \log s_{it} + \beta_\iota \overline{\log s}_{\ell_{it}} + u_{it}$, for ι the industry of worker i. The figure displays $\{\hat{\gamma}_\iota\}_\iota$ (orange circles) and the average effect (blue circle) along with 5% confidence intervals.

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). I include in the clustering algorithm variables that proxy for the geographic primitives of the model: the coordinates of the city center to account for the migration costs, its average wage to account for city's TFP, and its population to account for its amenity.

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.2 lists the number of cities by city type, the largest city in each type, the average size, as well as the average wage.

How good is the spatial approximation generate by the k-mean algorithm? 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.

F.2 Estimating equations

I derive here the equations that I use for the estimation of the model. First, in each city, the housing market must clears. The housing supply is by assumption $H_\ell = \mathcal{H} p_\ell^\delta$, where the housing demand is $H_\ell^D = \alpha Y_\ell / p_\ell$. Equating the two, solving for p_ℓ and taking logs yield (21).

I then turn to the migration cost equation (22). I derive the expression for young workers, but a similar expression exists for the old. From (53), we know that

$$\frac{n_{l\ell}^y(s)}{n_{ll}^y(s)} = \frac{\mathrm{e}^{\vartheta\left(sT_\ell/P_\ell + B_\ell^y - \kappa_{l\ell}^y + \beta O_\ell(s)\right)}}{\mathrm{e}^{\vartheta\left(sT_l/P_l + B_l^y + \beta O_l(s)\right)}}.$$

after using $\kappa^y_{ll} = 0$. Similarly,

$$\frac{n^y_{\ell\ell}(s)}{n^y_{\ell l}(s)} = \frac{\mathrm{e}^{\vartheta\left(sT_\ell/P_\ell + B^y_\ell + \beta O_\ell(s)\right)}}{\mathrm{e}^{\vartheta\left(sT_l/P_l + B^y_l - \kappa^y_{\ell l} + \beta O_l(s)\right)}}.$$

Dividing the first expression by the second, using the symmetry $\kappa^y_{l\ell} = \kappa^y_{\ell l}$, and integrating across young workers w.r.t. $dN^y(s)$ yields (22).

Third and last, I derive the expression that I iterate on to obtain cities' amenities. Here as well, I derive the expression for young workers, but a similar expression exists for the old. From (53), we know that

$$n^y_{l\ell}(s) = \mathrm{e}^{\vartheta\left(B^y_\ell - B^y_P\right)} \mathrm{e}^{\vartheta\left(\frac{sT_\ell}{P_\ell} - \frac{sT_P}{P_P} + \beta O_\ell(s) - \beta O_P(s)\right)} \mathrm{e}^{-\vartheta(\kappa^y_{l\ell} - \kappa^y_{lP})} n^y_{lP}(s),$$

for some reference city P. Using $n^y_\ell(s) = \sum_l n^y_{l\ell}(s)$, and then integrating across workers, we obtain

$$N^y_\ell = \mathrm{e}^{\vartheta\left(B^y_\ell - B^y_P\right)} \int \mathrm{e}^{\vartheta\left(\frac{sT_\ell}{P_\ell} - \frac{sT_P}{P_P} + \beta[O_\ell(s) - O_P(s)]\right)} \sum_l \mathrm{e}^{-\vartheta(\kappa^y_{l\ell} - \kappa^y_{lP})} n^a_{lP}(s) ds. \tag{63}$$

The left-hand side is the city ℓ employment share of young workers. The first-term is city ℓ's relative amenities. Finally, the integral on the right is a term that can be computed within the model. Normalizing $B^y_P = B^o_P$, cities' amenities are recovered by iterating on this expression.

F.3 Migration heterogeneity

The model predicts substantial heterogeneity in migration probabilities across young workers. On average, an 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 the reduced-form equation in the model

$$\mathrm{mig}_\ell(s) = \alpha_\ell + \beta_\ell \log s + u_\ell(s), \tag{64}$$

where $\mathrm{mig}_\ell(s)$ 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.7a and Figure F.7b 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.

F.4 Sensitivity

Figure F.8 and F.9 reports information to help assess the sources of identification in the indirect inference procedure. Specifically, Figure F.8 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.8 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.8 also shows that some parameters affect mainly one moment whereas others drive several moments.

Figure F.9 complements the analysis of Figure F.8 by reporting the sensitivity measure of Andrews et al. (2017). Intuitively, this metric is the inverse of the moment elasticities reported in Figure F.8. 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.9 displays the absolute value of $\boldsymbol{\Lambda}$, and the sign of $\boldsymbol{\Lambda}$ is reported in parenthesis. As in Figure F.8, the blue bar highlights which moment is supposed to identify the specific parameter.

I draw two takeaways from Figure F.9. 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 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.5 Over-identification exercises: sorting in the data and in the model

F.5.1 TFP estimates

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,

$$\underbrace{\mathbb{E}[\log W_\ell] - \mathbb{E}[\log W]}_{\text{Between-city wage gap}} = \underbrace{\log T_\ell - \mathbb{E}[\log T]}_{\text{TFP gap}} + \underbrace{\mathbb{E}[\log S_\ell]}_{\text{Skill gap}}, \tag{65}$$

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 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 (65) 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.[88]

Figure F.10a and F.10b 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 F.10a), while the between-city skill gaps explain the remaining 57% of the variation (orange dotted line in Figure F.10b). 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).

F.5.2 Spatial sorting across the lifecycle

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

$$\log\left(\frac{n^y_\ell(s)}{n^o_\ell(s)}\right) = \vartheta\left(B^y_\ell - B^o_\ell\right) - \log\left(\frac{\mathcal{N}^y(s)}{\mathcal{N}^o(s)}\right) + \vartheta\beta O_\ell(s), \tag{66}$$

where $\mathcal{N}^a(s) \equiv \sum_\ell \mathrm{e}^{\vartheta[w_\ell(s)+B^a_\ell+\beta O^a_\ell(s)]}$ 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 (66), I measure in the data the between-age difference in willingness to sort through the reduced-form specification

$$\log\left(\frac{N^y_{\ell q}}{N^o_{\ell q}}\right) = \alpha_\ell + \beta_q + \gamma_q \mathbb{E}[\log W_\ell] + e_{\ell q}, \tag{67}$$

where $N^a_{\ell q}$ 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

[88]Suppose $s_{it} = i$ so that the empirical wage model reads $\log w_{it} = \alpha_{\ell_{it}} + \beta_i + u_{it}$. City TFPs are then estimated off within-worker time variation in wages for workers who switch locations: $\log w_{it+1} - \log w_{it} = \alpha_{\ell_{it+1}} - \alpha_{\ell_{it}} + u_{it+1} - u_{it}$. That is, city TFPs are estimated off wage growth variation for movers. Hence, these estimates conflate the proper city productivity with estimates of the learning technology.

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 (67) in the data and in the model by OLS.

Figure F.10c 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.11 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.[89]

F.6 Additional Tables and Figures

Table F.1: 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 (22)	·	·
δ	Housing price elasticity	7.473	Housing price $\sim$ exp.: constant (21)	·	·
$\mathcal{H}$	Supply stock	7.2×10^{-9}	Housing price $\sim$ exp.: slope (21)	·	·
g_1	Learning technology	1.057	Wage growth $\sim$ wage (19)	·	·
g_2	Learning technology	0.327	Wage growth $\sim$ city wage (19)	·	·
g_{12}	Learning technology	0.393	Wage growth $\sim$ wage $\times$ city wage (19)	·	·
		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	Local inequality $\sim$ local wage	0.29	0.28
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.

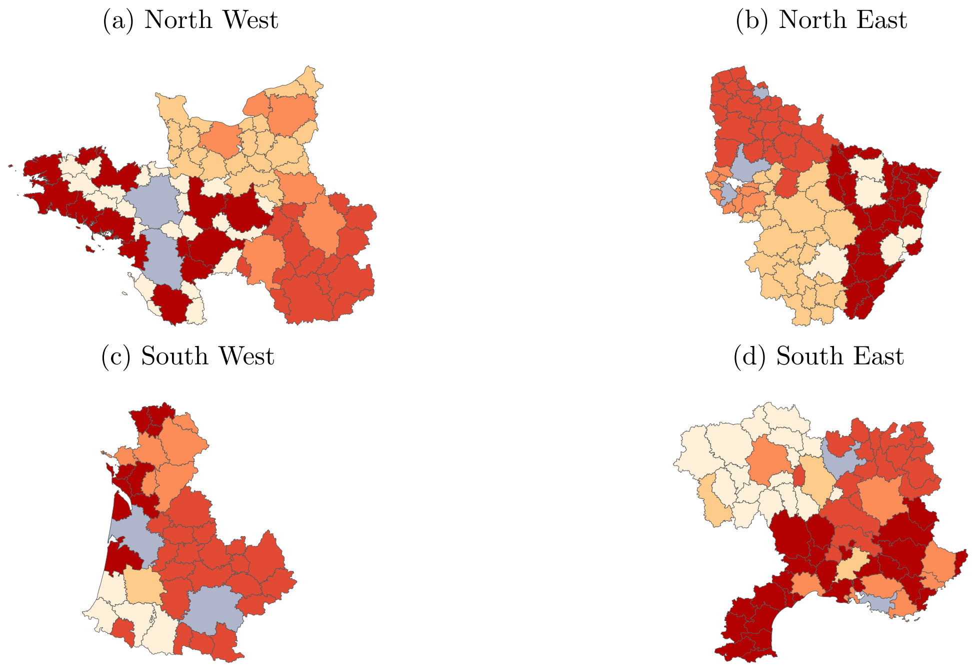
[89] When migration costs are positive, the option value of productive cities reflect both the learning opportunities and the future migration costs. Figure F.11 plots the spatial differences in option values when interactions are not segmented by cities (Section 5). In this case, skilled workers continue 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.

Table F.2: 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

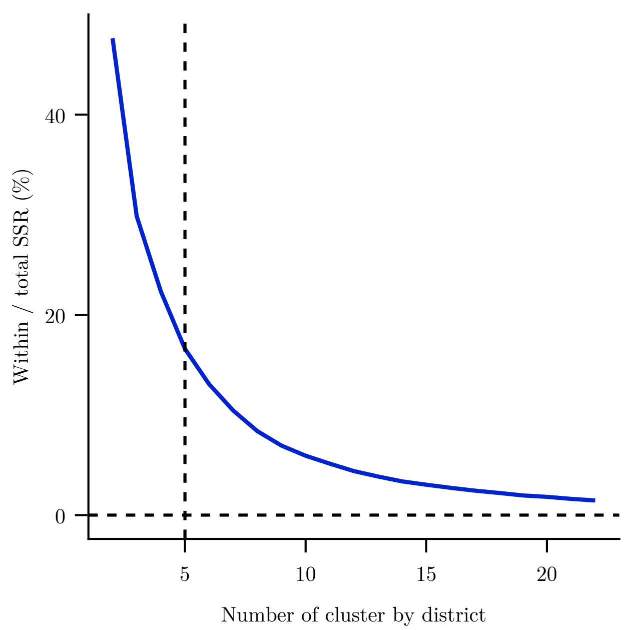
The tables report statistics on the city types used for the estimation of the quantitative model. The second column is the district in which the city type belongs (see Figure F.1). The third column is the number of cities included in the city city type. For the first ten types, there is a single city 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.1: City clustering



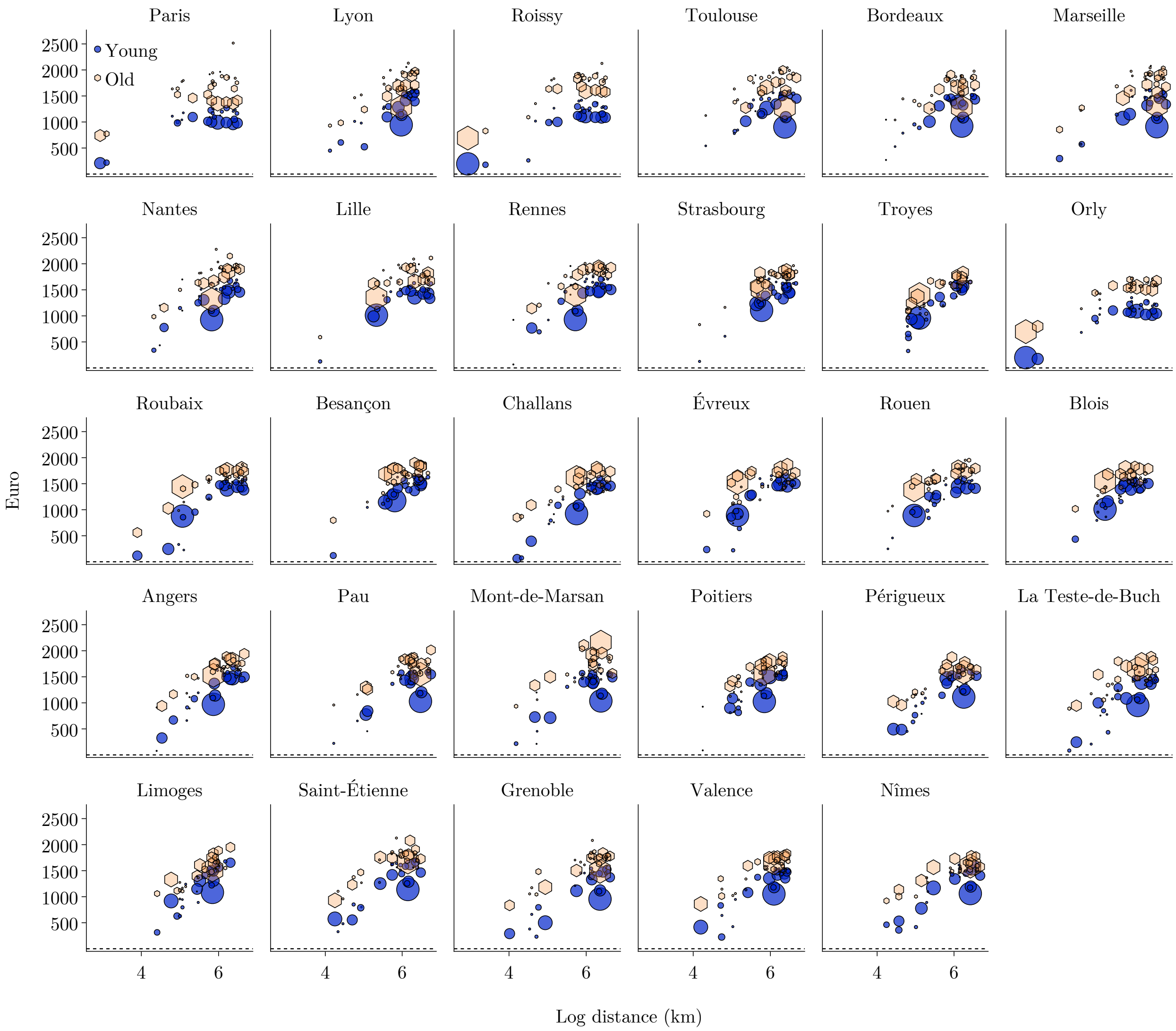
Note: each subplot is a French district. The grey areas are the ten largest French cities. The remaining commuting zones are subdivided into five groups within each district. Each city cluster is represented by a different color on the map. See Table F.2 for a list of the clusters.

Figure F.2: City clustering residual



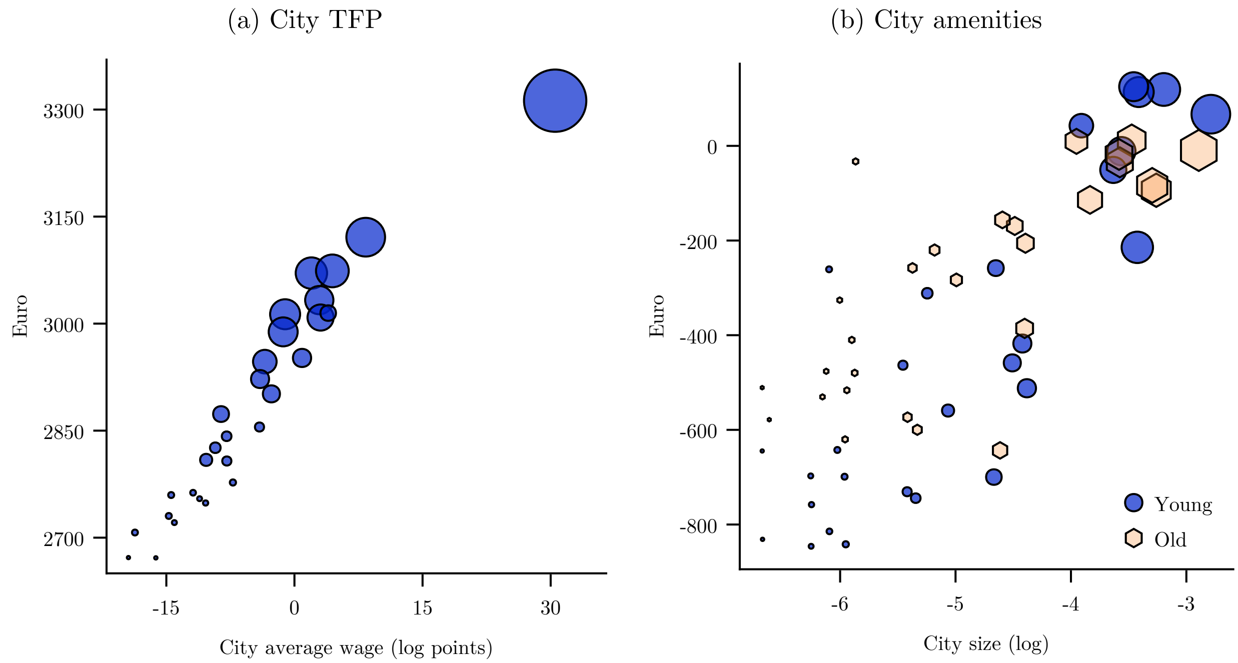
Note: average within cluster sum of square residuals (SSR) relative to the total SSR against the number of cluster (per district) used in the k-mean algorithm. The quantitative model features five city-types per district, as represented by the vertical dashed black line.

Figure F.3: Estimated age-specific migration costs by city of origin



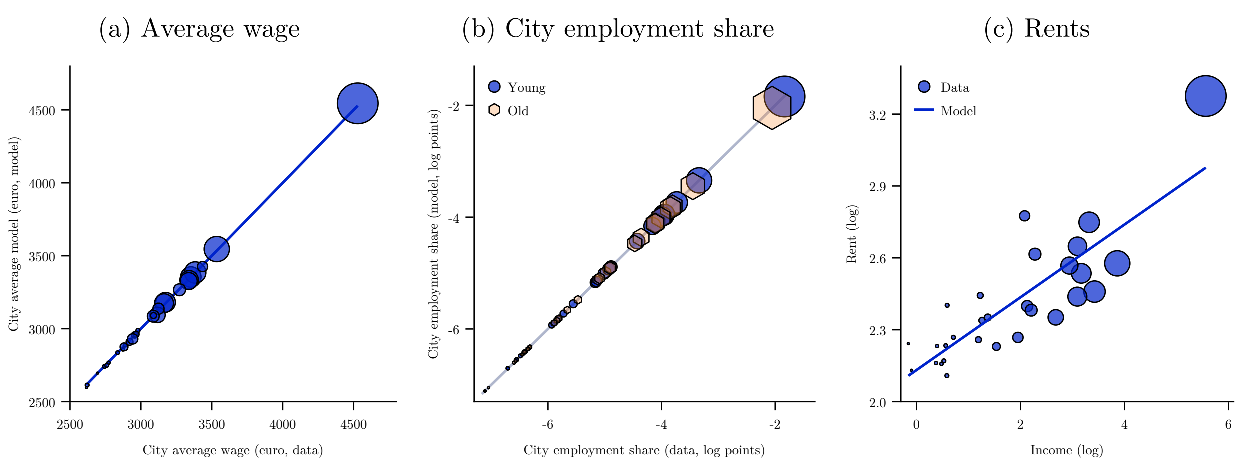
Note: estimated age-specific migration costs by city of origin as a function of the distance between the city of origin and the destination. Each subplot corresponds to a city of origin. The migration costs are estimated according to (22). The estimated costs are in utils, and they are reported here in euros through $\kappa^a_{\ell\ell} P_\ell$. 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 (63). 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.

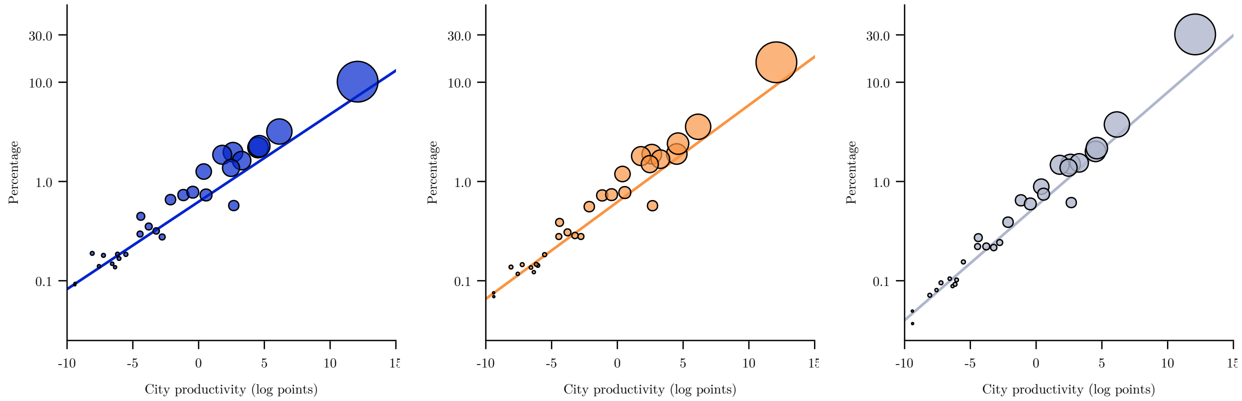
Figure F.5: Model fit



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: Spatial allocation of workers across the skill distribution

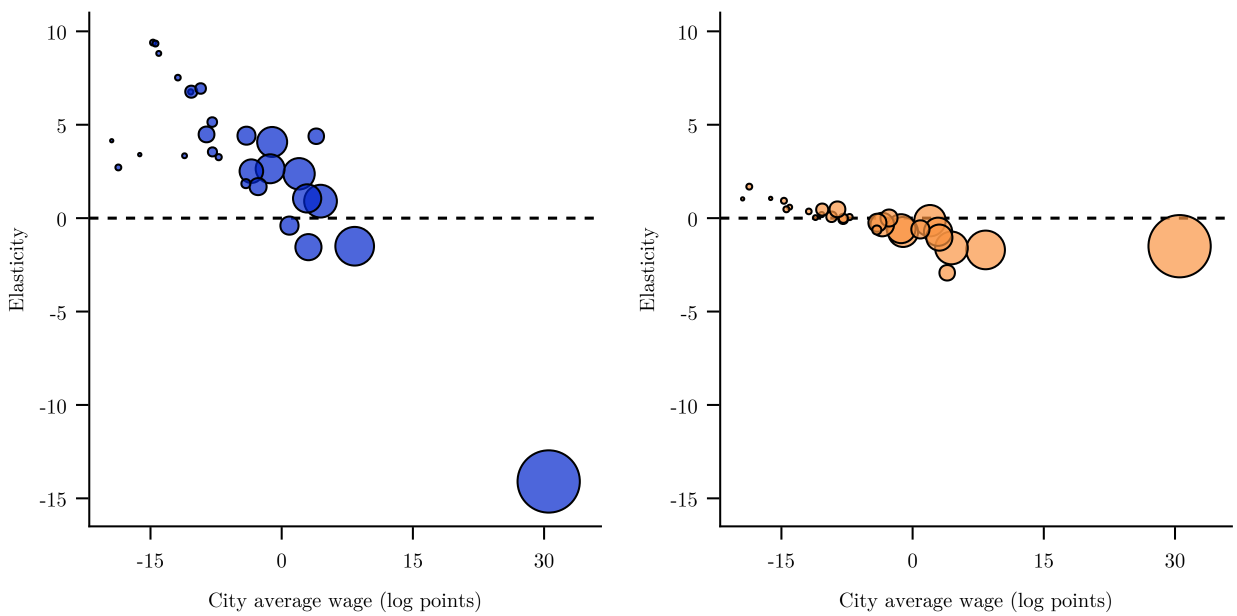
(a) Skill ≤ P50 (b) Skill ∈ P50-P90 (c) Skill ≥ P90



Note: three panels display the probability to work in a particular city if your skill is below the 50th percentile (a), between the 50th and 90th percentile (b), or above the 90th percentile (c). The x-axis is cities' relative log TFP, and the y-axis in log-scale. The size of the circles are proportional to the size of the cities.

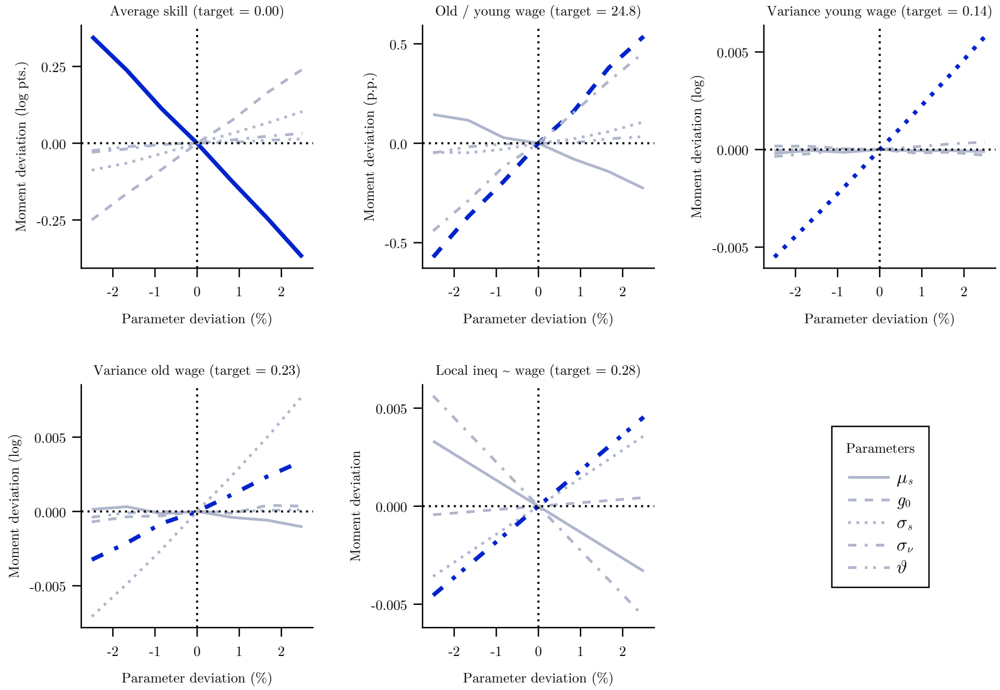
Figure F.7: Migration elasticities across space

(a) Young migration elasticity by birthplace (b) Old migration elasticity by birthplace



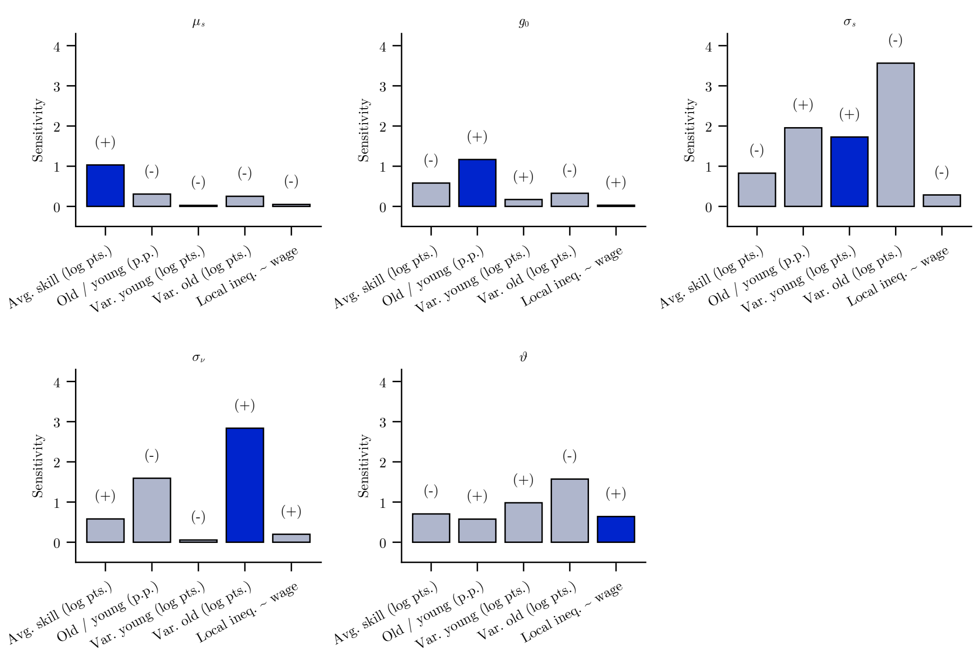
Note: panel (a) and (b) display the migration elasticities for young and old workers as estimated by (64). Skills are normalized in the regression, so that the elasticities correspond to the percentage point change in migration probability in response to an increase in skill by one standard deviation. The size of the circles are proportional to the size of the cities.

Figure F.8: Sensitivity of targeted moments to parameters



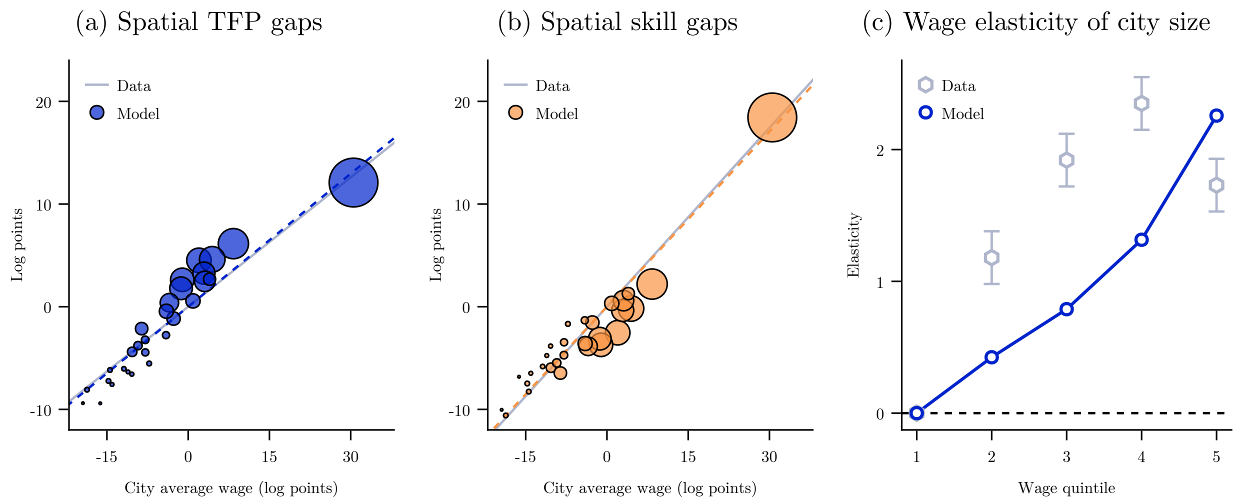
Note: each panel is a targeted moment in the indirect inference procedure. Each line is a parameter calibrated by the indirect inference procedure. The blue lines highlight the parameter that are heuristically identified by the particular moment of interest. I consider local deviations around the estimated parameters of 2.5% in each direction. The deviations are represented in percentages on the x-axis.

Figure F.9: Sensitivity of parameters to targeted moments



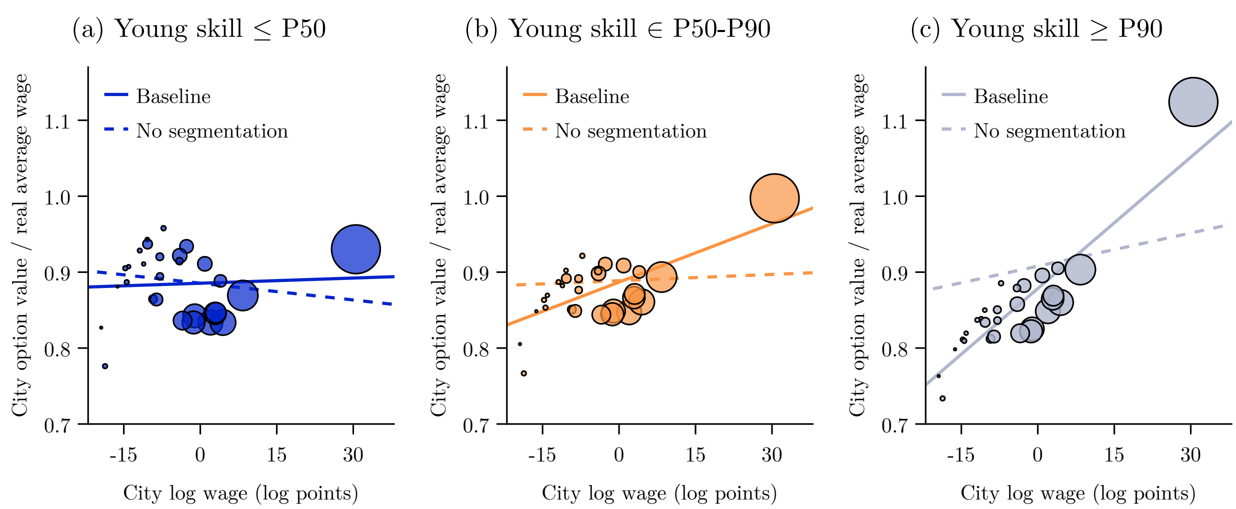
Note: figure reports the sensitivity measure of Andrews et al. (2017) in absolute value. The sign of the sensitivity is reported in parenthesis. Each panel is a parameter calibrated in the indirect inference procedure. Each column is a targeted moment. The blue bar highlights which moment is supposed to identify the parameter of interest.

Figure F.10: Worker sorting in the model and in the data.



Note: panel (a) displays the estimated city TFP. 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 projection of log wages on city and skill fixed effects. Panel (b) displays the average skill in city ℓ, $\mathbb{E}[\log S_\ell]$. 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. Panel (c) plots the point estimates of $\{\gamma_q\}$ in (67) obtained from the model (blue circles) and from the data (grey hexagons).

Figure F.11: The option value of cities across the skill distribution



G The Tradeoff Between Human Capital and Inequality

Figure G.1: Skill growth by birthplace and initial skill

(a) Low skill: $\leq$ P50 (b) Middle skill: P50 - P90 (c) High skill: $\geq$ P90

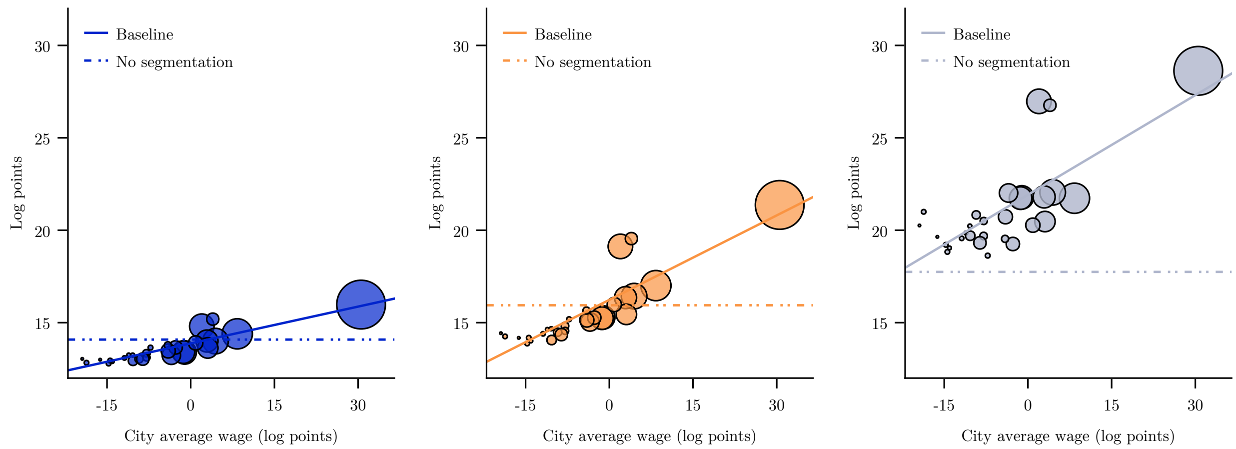


Figure G.2: Relative skil growth by birthplace

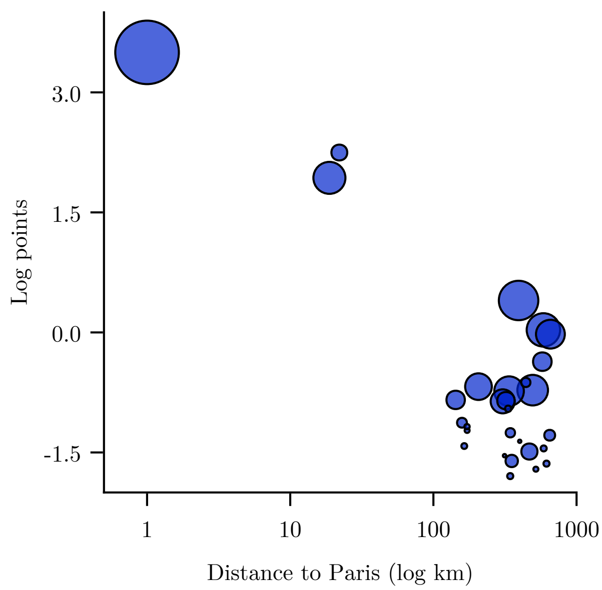
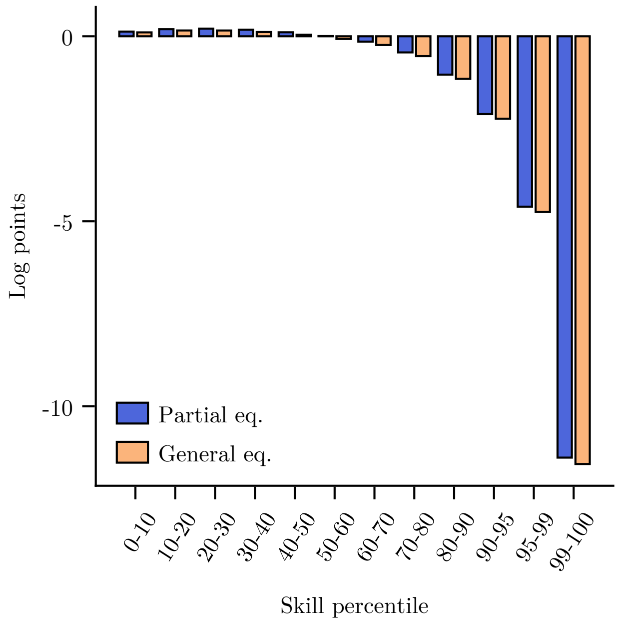
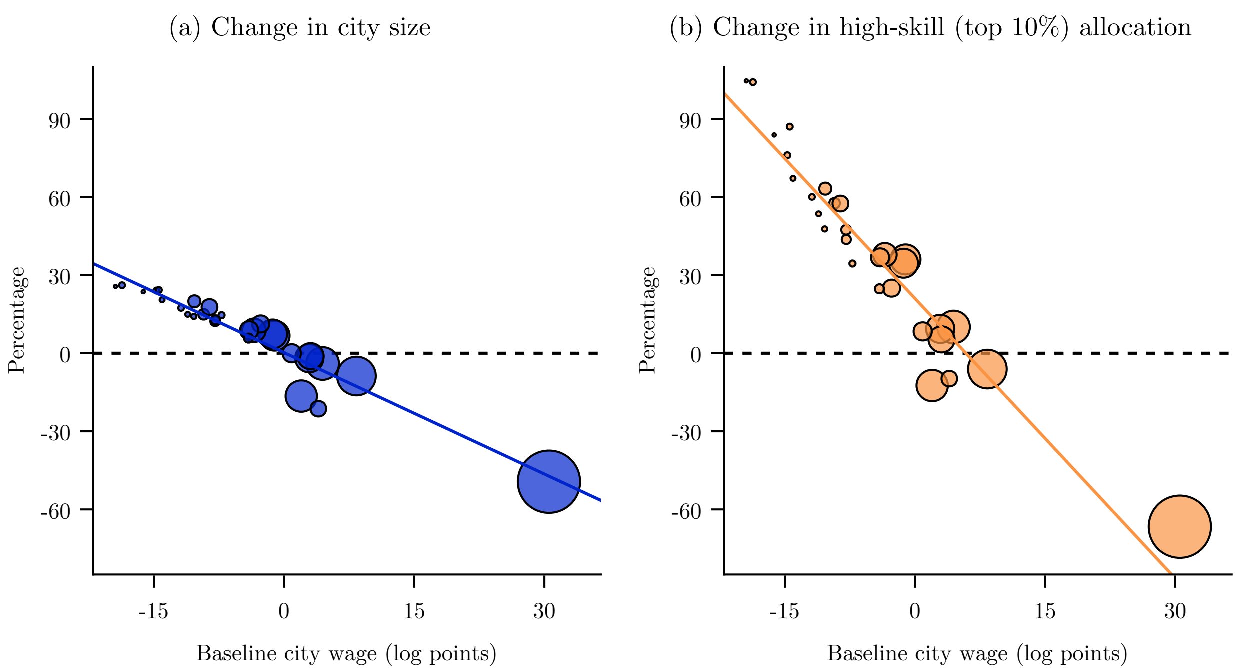


Figure G.3: The partial and general equilibrium effect of local interactions on skill growth



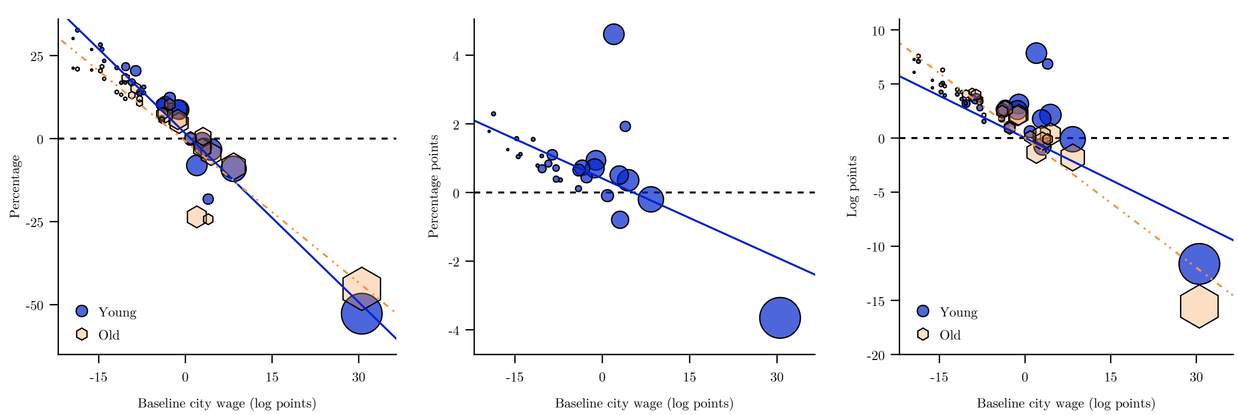
Note: figure plots the change in average skill growth between the baseline equilibrium and the no segmentation equilibrium across the skill distribution of young workers. The blue bars are the change in skill growth when interactions are no longer spatially segmented but the aggregate skill distribution is held constant (partial equilibrium). The orange bars are in general equilibrium letting the aggregate skill distribution adjusts.

Figure G.4: 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 percentage change in the propensity of high-skill workers (top 10% of the skill distribution) to work in a particular city. The orange line is the fitted line. The size of the markers is proportional to the baseline city size in both panels.

Figure G.5: The agglomeration effect of local interactions by age



H The Consequences of Spatial Policies

H.1 Welfare analysis

Consumption equivalence I first present how I derive the consumption-equivalent welfare metric of Section 6.3. Let $\mathcal{V}_l^y(s,\zeta,\boldsymbol{\tau})$ denote the lifetime utility of a young worker born in l with skill s when their lifetime consumption is multiplied by ζ and they face a transfer function $\boldsymbol{\tau}$,

$$\mathcal{V}_l^y(s,\zeta,\boldsymbol{\tau}) = \frac{1}{\vartheta}\log\left(\sum_{\ell} \mathrm{e}^{\vartheta\left(\frac{\left[sT_\ell+\tau_{l\ell}^y(s)\right]\zeta}{P_\ell}+B_\ell^y-\kappa_{l\ell}^y+\beta O_\ell(s,\zeta,\boldsymbol{\tau})\right)}\right),$$

where[90]

$$O_\ell(s,\zeta,\boldsymbol{\tau}) = \frac{1}{\vartheta}\int\int\log\left(\sum_{\ell'} \mathrm{e}^{\vartheta\left(\frac{\left[s^oT_{\ell'}+\tau_{\ell\ell'}^o(s^o)\right]\zeta}{P_{\ell'}}+B_{\ell'}^o-\kappa_{\ell\ell'}^o\right)}\right)dF\left(\frac{s^o}{\gamma(s,s_p)}\right)\pi_\ell(s_p).$$

Let $\zeta_l(s)$ denote the consumption-equivalent change in welfare induced by the policy $\boldsymbol{\tau}$ for workers born in l with skill ℓ. This consumption-equivalent measure is defined implicitly by $\mathcal{V}_l^y[s,\zeta_\ell(s),0] = \mathcal{V}_l^y(s,0,\boldsymbol{\tau})$. Intuitively, $\zeta_l(s)$ measure by how much does the lifetime consumption of workers born in l with skill s needs to be multiplied in the baseline equilbrium to make these workers indifferent between the policy $\boldsymbol{\tau}$ and the baseline economy.

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 "representative" workers. 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 equilibrium,

$$\int \mathcal{V}_{\ell_i}^y(s_i,\zeta_g,0)dF_i^g = \int \mathcal{V}_{\ell_i}^y(s_i,0,\boldsymbol{\tau})d\tilde{F}_i^g,$$

where ζ_g is the welfare measure, and dF_i^g denote the distribution of workers in group g for simplicity.[91] This second welfare metric is consistent with a utilitarian social welfare function that places equal weights on workers within group g.[92]

[90] I use the change of variable $e\gamma(s,s_p) \to s^o$ to derive the option value of city ℓ.

[91] Note that dF_i^g and $d\tilde{F}_i^g$ need not be the same as the policy is reallocating workers across space.

[92] 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.

Welfare decomposition 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 $\chi_l^y(s)$.[93] 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

$$\begin{aligned} O_\ell(s) &= \int\int \mathbb{E}\left[\max_{\ell'}\left\{\frac{e\gamma(s,s_p)T_{\ell'}}{P_{\ell'}} - \kappa_{\ell\ell'}^o + \varepsilon_{\ell'}\right\}\right]\pi_\ell(s_p)dF(e)ds_p \\ &= \int\left(\sum_{\ell'}\frac{n_{\ell\ell'}^o(s^o)}{m_\ell^o(s^o)}\left(\frac{s^oT_{\ell'}}{P_{\ell'}} - \kappa_{\ell\ell'}^o\right) + \chi_\ell^o(s^o)\right)\omega_\ell(s^o,s)ds^o. \end{aligned}$$

where the second line follows from the change of variable $e\gamma(s,s_p) \to s^o$ and the definition

$$\omega_\ell(s^o,s) \equiv \int \frac{\pi_\ell(s_p)}{\gamma(s,s_p)} f\left(\frac{s^o}{\gamma(s,s_p)}\right)ds_p.$$

Combined with the expression for young workers, we can therefore write lifetime utility as

$$\begin{aligned} \mathcal{V}_l^y(s) = &\underbrace{\sum_\ell\left(\frac{\tau_{l\ell} - \kappa_{l\ell}^y P_\ell}{P_\ell}\right)\frac{n_{l\ell}^y(s)}{m_\ell^y(s)} - \beta\sum_\ell \frac{n_{l\ell}^y(s)}{m_\ell^y(s)}\int \omega_\ell(s^o,s)\left(\sum_{\ell'}\frac{n_{\ell\ell'}^o(s^o)}{m_\ell^o(s^o)}\kappa_{\ell\ell'}^o\right)ds^o}_{\text{Transfers net of lifetime migration costs } \equiv V_l^T(s)} \\ &\underbrace{\sum_\ell\left(\frac{sT_\ell}{P_\ell}\right)\frac{n_{l\ell}^y(s)}{m_\ell^y(s)}}_{\text{Present real income} \equiv V_l^I(s)} + \underbrace{\beta\sum_\ell\frac{n_{l\ell}^y(s)}{m_\ell^y(s)}\int\omega_\ell(s^o,s)\left(\sum_{\ell'}\frac{n_{\ell\ell'}^o(s^o)}{m_\ell^o(s^o)}\frac{s^oT_{\ell'}}{P_{\ell'}}\right)ds^o}_{\text{Future real income} \equiv V_l^{EI}(s)} + \\ &\underbrace{\chi_l(s) + \beta\sum_\ell\frac{n_{l\ell}^y(s)}{m_\ell y(s)}\int\chi_\ell(s^o)\omega_\ell(s^o,s)ds^o}_{\text{Preferences} \equiv V_l^A(s)}, \end{aligned}$$

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

$$\begin{aligned} 0 = \mathcal{V}_l^y[s,\zeta_l(s)] - \mathcal{V}_l^{y\star}(s) &= \mathcal{V}_l^y[s,\zeta_l(s)] - \mathcal{V}_l^y(s) + \mathcal{V}_l^y(s) - \mathcal{V}_l^{y\star}(s) \\ &= \mathcal{V}_l^y[s,\zeta_l(s)] - \mathcal{V}_l^y(s) + \left(V_l^A(s) - V_l^{A\star}(s)\right) + \left(V_l^T(s) - V_l^{T\star}(s)\right) + \\ &\quad \left(V_l^I(s) - V_l^{I\star}(s)\right) + \left(V_l^{EI}(s) - V_l^{EI\star}(s)\right), \end{aligned}$$

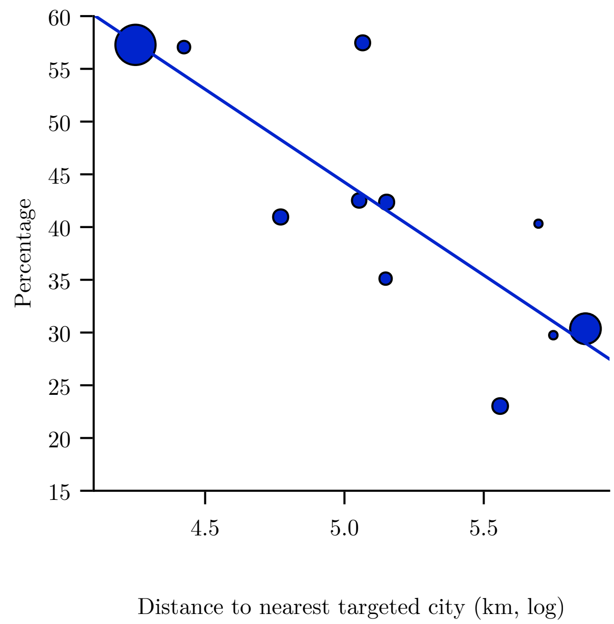
where the first equality is the definition of $\zeta_l(s)$, the second is an identity, and the third follows from the decomposition I just derived. From there, we can define the four consumption-equivalent welfare metric:

$$\begin{aligned} \mathcal{V}_l^y[s,\zeta_l^A(s)] &= \left(V_l^{A\star}(s) - V_l^A(s)\right) + \mathcal{V}_\ell^y(s), \\ \mathcal{V}_l^y[s,\zeta_l^{A+T}(s)] &= \left(V_l^{A\star}(s) - V_l^A(s)\right) + \left(V_l^{T\star}(s) - V_l^T(s)\right) + \mathcal{V}_\ell^y(s), \\ \mathcal{V}_l^y[s,\zeta_l^{A+T+I}(s)] &= \left(V_l^{A\star}(s) - V_l^A(s)\right) + \left(V_l^{T\star}(s) - V_l^T(s)\right) + \left(V_l^{I\star}(s) - V_l^I(s)\right) + \mathcal{V}_\ell^y(s), \\ \mathcal{V}_l^y[s,\zeta_l(s)] &= \mathcal{V}_l^{y\star}(s). \end{aligned}$$

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

[93] No closed-form solution exists for $\chi_l^y(s)$ but it can be obtained as utility residual.

Figure H.1: Average take-up rate of moving vouchers by birthplace



$\zeta_l^{A+T+I}(s)$ 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 H.5 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.[94]

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

H.2 Additional Figures

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

Figure H.2: The role of complementarities in shaping the learning costs of the voucher policy

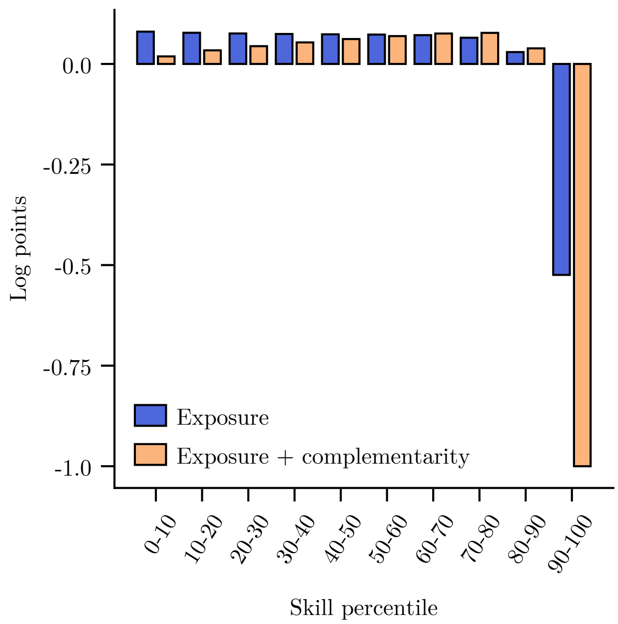


Figure H.3: The consequences of moving vouchers by skill and birthplace

(a) Policy take-up rate

(b) Change in expected skill

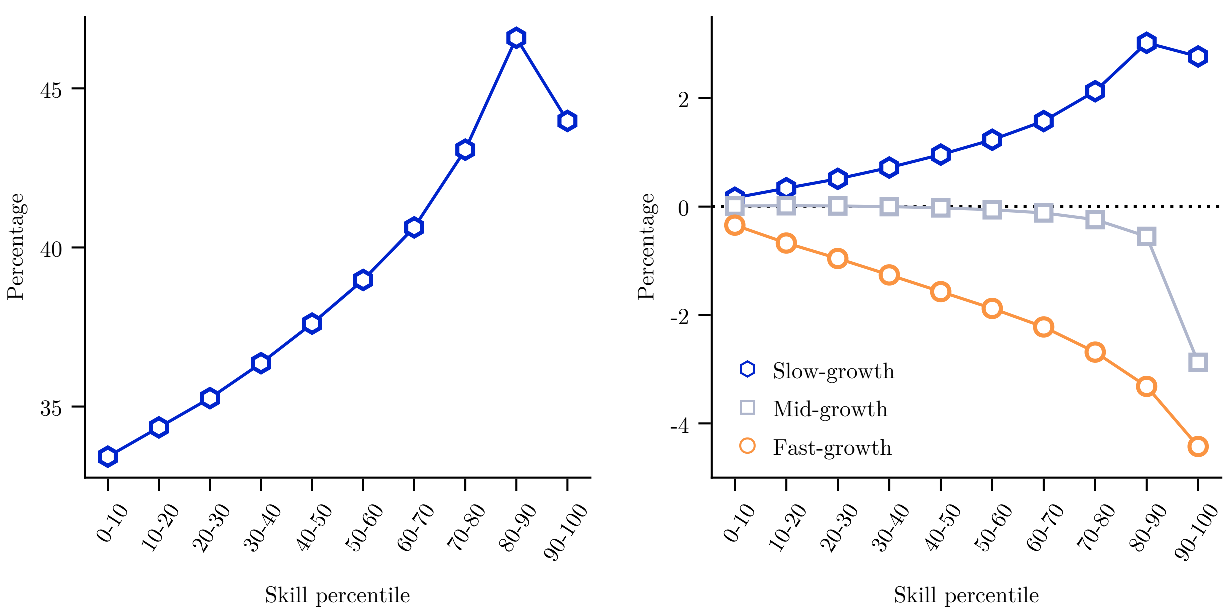


Figure H.4: The consequences of the voucher policy on local interactions

(a) Change in average skill by workplace (b) Change in average skill growth by workplace

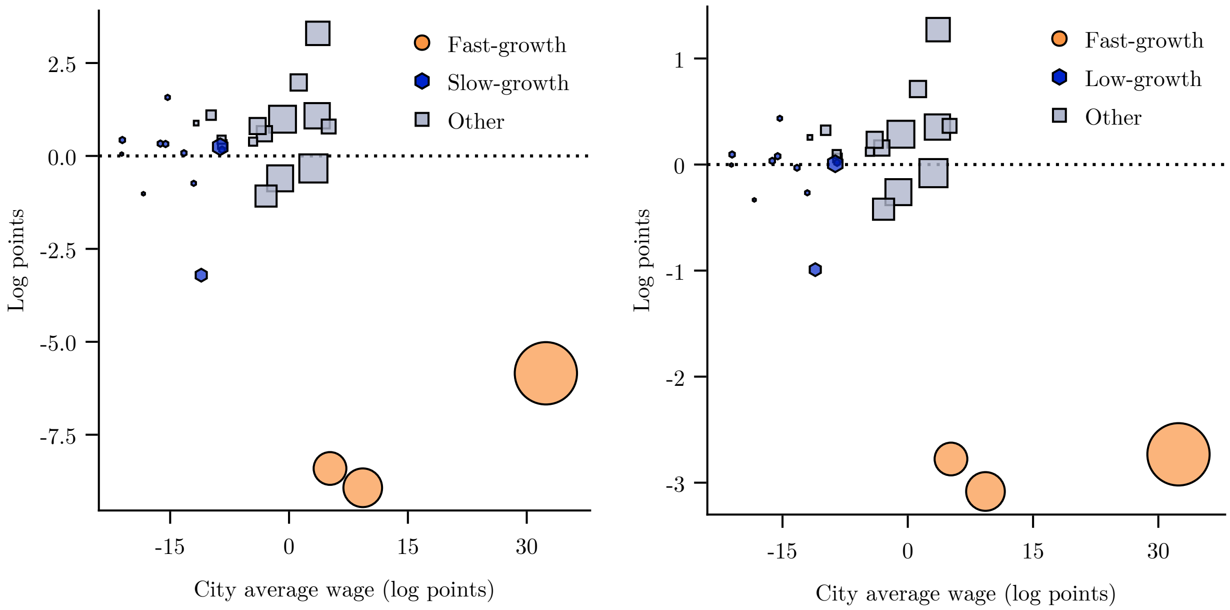


Figure H.5: Welfare decomposition of the moving voucher policy by birthplace

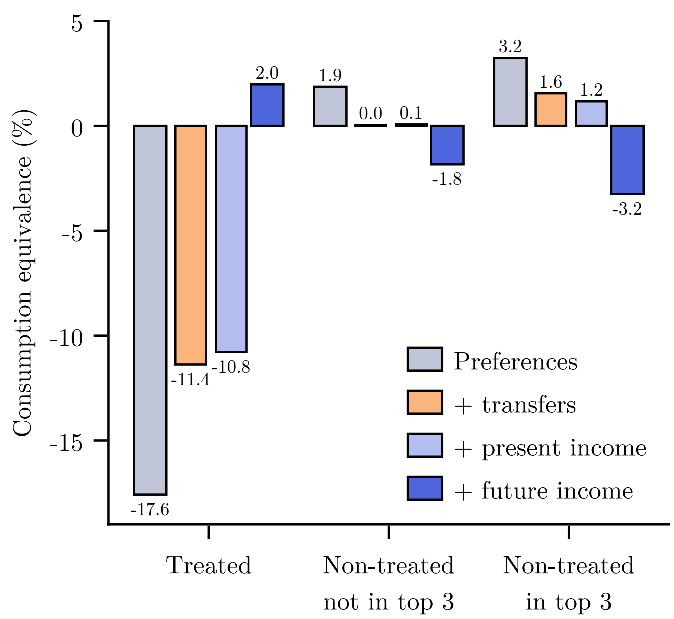


Figure H.6: The consequences of the moving voucher policy across skills and birthplaces