A Dynamic Spatial Knowledge Economy

Levi Crews*
UCLA

13th December, 2024

Please click here for the latest version.

Abstract

Cities have long been thought to drive economic growth. Despite this, analyses of spatial policies have largely ignored the effects of such policies on growth. In this paper, I develop a spatial endogenous growth model in which heterogeneous agents make forward-looking migration decisions and human capital investments over the life cycle. Local externalities in the human capital investment technology drive both agglomeration and growth. I show that, along a balanced growth path, the growth rate depends on the spatial distribution of human capital, making it sensitive to spatial policies. I calibrate the model to data on U.S. metropolitan areas and show that it can rationalize the faster wage growth of workers in big cities, as well as other key patterns in life-cycle wage profiles, migration decisions, and city characteristics. Because workers accumulate human capital at different rates depending on where they live, the model provides an environment in which spatial policy can not just attract skilled workers, but produce them, too. I find that policies that further concentrate skilled workers in large cities are growth-enhancing.

JEL codes: E24, J24, O15, O18, O41, R11, R13, R23

*This paper was the main chapter of my Ph.D. dissertation at the University of Chicago. I am grateful to my advisors Esteban Rossi-Hansberg, Fernando Alvarez, Jonathan Dingel, and Felix Tintelnot for their invaluable advice and support. I thank Laurent Gobillon, Sam Kortum, and Marta Prato for insightful conference discussions; Rodrigo Adão, Adrien Bilal, Olivia Bordeu, Gilles Duranton, Fabian Eckert, Pablo Fajgelbaum, Cecile Gaubert, Mikhail Golosov, Brian Greaney, Agustín Gutiérrez, Greg Kaplan, Paolo Martellini, Simon Mongey, Ishan Nath, Galo Nuño, Florian Oswald, Michael Peters, Steve Redding, Andrés Rodríguez-Clare, Younghun Shim, Robert Shimer, Jonathan Vogel, Conor Walsh, Antonis Zitridis, and numerous conference and seminar participants for helpful comments; and Nancy Stokey and Tom Winberry for guidance on an earlier iteration of this project. I thank the International Economics Section at Princeton University for their hospitality while working on this project. All errors are my own. Contact: lgcrews@econ.ucla.edu.

1 Introduction

That cities drive economic growth is a refrain often repeated since Jacobs (1969) and even Marshall (1890). Ideas flow freely through city streets, the story goes, and pass all the more rapidly between people the denser is the city. It is these ideas—embodied as human capital in the workers who learn, use, and share them—that drive economic growth (Lucas, 1988; Glaeser, 2011).

Recent panel estimates suggest that workers do indeed learn more in big cities and that they anticipate the ensuing wage gains when making migration decisions over the life cycle.[1] But models of idea-driven growth have yet to incorporate this evidence, instead restricting attention to settings with no explicit space whatsoever.[2] Accordingly, it remains unclear how the allocation of workers across space contributes to determining the pace of aggregate growth. At the same time, work evaluating spatial policies has been hamstrung by this gap, unable to weigh the potential growth effects of policies that shift workers around the urban landscape.

In this paper, I address these concerns by developing a quantitative spatial endogenous growth model that can be used to evaluate spatial policies. In so doing, I make three contributions. First, my model demonstrates a mechanism through which the allocation of human capital across space determines the pace of aggregate growth. Second, in quantifying the model, I jointly rationalize a host of empirical facts about U.S. cities in the cross-section and U.S. workers over the life cycle. Third, with the quantified model in hand, I solve for the outcome of a commonly-proposed place-based policy.

In the model, a finite number of cities are populated by heterogeneous workers who make forward-looking migration decisions and human capital investments over the life cycle. At the root of the model is a local externality in the investment process. Being near the human capital of her neighbors quickens a worker's accumulation of her own. As such, she is attracted to large, skilled cities for the permanent productivity boost they afford her.[3] It is this desire, shared to varying degrees by all workers, that draws individuals into larger cities despite their higher land rents; it is this permanent boost, experienced to varying degrees by all workers, that fosters aggregate growth.

The state of the economy at any moment is summarized by the distribution of human capital across locations and cohorts. An individual worker's migration and investment decisions depend on this distribution because the mass and composition of other workers in each city determine her own returns to living there. Workers' migration and investment decisions in turn determine how this distribution evolves. Accordingly, one of the equilibrium conditions is that workers solve a Hamilton-Jacobi-Bellman equation, taking the human capital distribution as given. Another is that the distribution evolves according to a Kolmogorov

[1] On learning in big cities, De la Roca and Puga (2017) show that wage growth is higher in bigger cities, these wage gains persist even after migration, and the wage gains are higher for workers with higher initial ability. Their results are discussed in more detail in Section 3. See also Glaeser and Maré (2001); Baum-Snow and Pavan (2012); Wang (2016); Carlsen, Rattsø, and Stokke (2016). On forward-looking migration decisions, first see Sjaastad (1962), then Gould (2007), Kennan and Walker (2011), and Bilal and Rossi-Hansberg (2021).

[2] See, for example, Lucas (2009), Lucas and Moll (2014), Gabriel and Lucas (2019), and Caicedo, Lucas, and Rossi-Hansberg (2019), all of which consider workers to be the relevant economic agents. A related literature following Luttmer (2012) and Perla and Tonetti (2014) that considers firms to be the relevant agents has incorporated space—but not location choice—through international trade (Alvarez, Buera, and Lucas, 2014; Buera and Oberfield, 2020; Perla, Tonetti, and Waugh, 2021).

[3] As Glaeser (1999) puts it, cities are "forges of human capital" and "intellectual furnaces where new ideas are formed."

forward equation given the decision rules of individual workers.

I focus on a particular class of solutions to these equations, namely balanced growth paths (BGPs), along which aggregate production grows at a constant rate and the spatial distribution of relative human capital levels remains stationary. As a consequence of this stationarity, city sizes are constant along a BGP—Detroit does not deteriorate, New York does not bounce back, the Sunbelt does not boom. Instead, differences across cities persist, with bigger cities typically being more expensive, more productive, and more skilled. Nevertheless, all cities experience shared output growth as workers migrate through the urban system while accumulating human capital over the life cycle.[4] So although the restriction to BGPs precludes studying the growth *of* individual cities, it proves to be a tractable way to study growth *because of* cities.

My main theoretical result is an expression for the aggregate growth rate along a BGP. The growth rate is pinned down endogenously as a function of the entire human capital distribution across space. In particular, I show that it can be expressed as the weighted average of the returns to investment in human capital, with more weight given to investments made in larger, more-skilled locations. The key here is that, in this economy, all productive capacity—both raw labor and knowledge—is embodied in workers.[5] Workers invest in their own knowledge by combining their labor and what they know with the knowledge of those around them. Whether the knowledge they create is wholly new to the economy or simply new to them is immaterial. What matters is that, contrary to many discussions of human capital, their knowledge is only partially rival.[6] It is rival to the extent that a worker cannot clone her raw labor—what one could call the "one brain, one body" constraint—but it is nonrival to the extent that it spills over into the learning processes of her neighbors. Notice the restriction to her *neighbors*: when every idea must be in somebody's brain, it matters where those brains are.

I solve for a BGP numerically by adapting an algorithm from Achdou et al. (2022). In so doing, I address what has been called the "hard problem" of regional economics, which asks how to solve models in which growth and agglomeration are outcomes of a joint process (Breinlich, Ottaviano, and Temple, 2014, p.685). This problem has been hard because it is high-dimensional: to make choices today, (heterogeneous) agents must understand the distribution of economic activity over space *and* time, and those choices in turn determine how the distribution behaves (Desmet and Rossi-Hansberg, 2014, pp.2–3). The key to my solution is that my equilibrium is a mean field game, a type of differential game first formulated by Lasry and Lions (2007). Mean field games are now the subject of a growing mathematical literature discussing their properties and numerical schemes to solve them, of which the algorithm from Achdou et al. (2022) is one example.[7]

[4] Admittedly, the same mechanism that determines the pace of aggregate growth in this model—namely, local knowledge externalities—may play a key role in what Moretti (2013a) calls the Great Divergence, wherein average wages in cities with larger endowments of skilled labor have pulled away from those in cities with smaller endowments since the early 1980s as technical change and globalization have pressed upon the U.S. economy (see also Glaeser and Saiz, 2004; Berry and Glaeser, 2005; Giannone, 2022). Through the lens of this model, though, these secular trends would drive transitions between BGPs, not movements along them. The study of the model's transitional dynamics is left for future work.

[5] As in Caicedo, Lucas, and Rossi-Hansberg (2019): "all knowledge in the economy is held by the individual people who comprise it: there is no abstract technology hovering above them in the ether."

[6] See, for example, Romer (1990, §2), who compares the nonrivalry of a design with the rivalry of the ability to add, and Jones (2005, §1).

[7] See Achdou et al. (2020) for an overview. To date, Achdou et al.'s (2022) algorithm has been used mostly for consumption-savings problems with incomplete markets, where the continuous choice variable is household wealth. Lucas and Moll (2014) used an early version of this algorithm to solve a model with

I quantify the model using recent data on 378 U.S. metropolitan areas in three steps. First, I back out measures of migration frictions from five years of observed flows between cities that are cross-tabulated by age group in the American Community Survey. Second, I calibrate the parameters governing the human capital investment sub-problem—which nests the canonical Ben-Porath (1967) model—to values that have been shown to rationalize key patterns in life-cycle wage profiles and the U.S. earnings distribution. Finally, I estimate the remaining parameters using a minimum distance estimator to best fit Duranton and Puga's (2023) estimates of workers' returns to experience in big cities, in addition to observed city employment and total wage bills.

The quantified model can match its aggregate targets exactly and implies city fundamentals that correlate well with observed features of those places. Despite solving for the spatial distribution of human capital nonparametrically, the quantified model generates reasonable marginal distributions of skill and age within each city. This is so in part because the model also does well in rationalizing observed patterns in human capital investments and migration over the life cycle. In particular, consistent with recent evidence, earnings growth is higher in big cities, and this boost in earnings persists even after migration to a smaller city, because workers accumulate human capital more quickly in bigger places.[8] This dynamic component to the earnings premium is anticipated by young workers, who migrate disproportionately to bigger and more-skilled cities in order to increase the returns to their time spent learning.

As an application of my quantitative framework, I consider a counterfactual in which I relax land use restrictions in the "brain hubs" of New York and San Francisco to the level of the median city, thereby making it cheaper for those two cities—already large and productive—to expand. I find that, along the new balanced growth path, both cities are larger and more skilled, with New York especially so. This is not primarily the result of skill-biased migration, as if the brain hubs simply syphoned off skilled workers from other cities. Instead, the rate of human capital accumulation in both cities increases, meaning that the spatial policy helps *produce* more skilled workers. Ultimately, I find that the aggregate growth rate for the entire economy increases by 13 basis points in response to the policy.

Related literature

The canonical paper from which this paper develops is Lucas (1988), though the model in its mechanics is actually closer to that in Uzawa (1965). In both papers, growth is driven by the costly accumulation of human capital by an infinitely-lived representative individual. Lucas (1988, §6) observes that this mechanism may undergird the power of cities argued for by Jacobs (1969), but he does not model it explicitly.[9] In his later work, Lucas further developed his observation by modeling the accumulation of human capital through meetings between heterogeneous individuals but, crucially, he continued to abstract from the space

balanced growth but with only one location. A few contemporaneous papers have used the algorithm in spatial models. The closest of these to the present paper is Martellini (2022), which investigates potential causes of the urban wage premium in a stationary equilibrium. See also Bilal (2023) and Greaney (2023).

[8] See especially Glaeser and Maré (2001), Baum-Snow and Pavan (2012), and Duranton and Puga (2023) who study U.S. workers.

[9] "It seems to me that the 'force' we need to postulate to account for the central role of cities in economic life is of exactly the same character as the 'external human capital' I have postulated as a force to account for certain features of aggregative development. . . . What can people be paying Manhattan or downtown Chicago rents for, if not for being near other people [in order to learn from them]?" (pp.38–39).

in which those meetings happen and by which they are circumscribed.[10] By contrast, this paper models cities explicitly while incorporating skill heterogeneity and adding life-cycle dynamics that are critical for rationalizing evidence from worker panels.

Others have embedded Lucas's (1988) endogenous growth model in an urban system, most notably Black and Henderson (1999).[11] Their model is highly stylized: there are two types of perfectly-specialized cities, all populated by interchangeable members of a household who are allocated to locations by a patriarch. Such bones are unfit to carry the flesh of facts from worker-level studies of migration and human capital investment.[12] Moreover, in their model agglomeration and growth occur through separate processes: the agglomeration spillovers are wholly static, and the learning technology has no local effects. This leaves aside the "hard problem" and is inconsistent with recent evidence that my model can rationalize (Glaeser and Maré, 2001; Baum-Snow and Pavan, 2012; Wang, 2016; De la Roca and Puga, 2017).

Still others have developed static models in which local knowledge spillovers among heterogeneous workers beget agglomeration.[13] Glaeser (1999) considers one city and its hinterland, wherein city residents are more likely to become skilled.[14] His model predicts that younger, skilled individuals will sort disproportionately into the city because they value the learning that the city provides. Davis and Dingel (2019) embed a similar mechanism into a system-of-cities model, which they use to rationalize the higher nominal wages, housing prices, productivity, and average skill level observed in larger U.S. cities. My model matches the same facts, but it also rationalizes observed patterns of life-cycle migration and human capital investments over time. Moreover, because my model is dynamic, it can trace out the effects of local knowledge spillovers on long-run growth, not just cross-sectional city characteristics.

Work on spatial policy has also suffered from the restriction to static models.[15] To conclude their review of place-based policies, Glaeser and Gottlieb (2008, p.215) advised that "[l]ocal policies that either attract or produce skilled people seem likely to offer the best chance of improving the fortunes of a troubled urban area." With static models, attention has been paid only to how (and who) those policies *attract* (Gaubert, 2018; Hsieh and Moretti, 2019; Fajgelbaum and Gaubert, 2020; Rossi-Hansberg, Sarte, and Schwartzman, 2021; Donald, Fukui, and Miyauchi, 2023b; Gaubert et al., 2024).[16] This paper studies how such policies might affect the *production* of skilled workers, too.[17] In principle, policies that

[10] See Lucas (2009), Gabriel and Lucas (2019) and Caicedo, Lucas, and Rossi-Hansberg (2019), which include cohort structures, and Lucas and Moll (2014), which includes costly on-the-job learning.

[11] See also Eaton and Eckstein (1997), whose framework is closer in spirit to mine but whose specification of endogenous growth is incomplete, and Rossi-Hansberg and Wright (2007), who further develop Black and Henderson's approach in a stochastic environment with many industries. Using an augmented version of Black and Henderson's model, Davis, Fisher, and Whited (2014) estimate that agglomeration, by itself, raises per capita consumption growth by 10.2%.

[12] Compare Klette and Kortum (2004), who use "[f]irm-level studies of research and development, productivity, patenting, and firm growth" to "add flesh to [the] bones" of Schumpeterian growth theory.

[13] Behrens and Robert-Nicoud (2015) reviews agglomeration theory with heterogeneous agents.

[14] Glaeser's agents live two periods, but the equilibrium conditions (1) and (2) are static.

[15] Contemporaneous work by Donald, Fukui, and Miyauchi (2023a) studies optimal spatial policy in a dynamic model with frictional migration à la Caliendo, Dvorkin, and Parro (2019).

[16] For Hsieh and Moretti (2019), who interpolate between two static equilibria of their model to infer TFP growth from the data, an additional concern is that they treat as exogenous a variable (namely, productivity growth) that my model demonstrates is in fact sensitive to spatial policies.

[17] Rossi-Hansberg, Sarte, and Schwartzman (2021, p.43) make this explicit: "Our analysis abstracts from the role that spatial polarization might have on human capital formation." So do Fajgelbaum and Gaubert (2020, pp.1014-15): "Finally, we only considered a static model, where each worker type is fixed regardless

attract human capital to cities where it might be best employed today are not necessarily the same as those that would produce the most human capital in the long run, especially when life-cycle dynamics are considered. The framework I develop allows policymakers to take this into consideration.

Finally, the quantitative model introduced in Section 2 contributes to a burgeoning literature on dynamic spatial models.[18] Because dynamic spatial models can be particularly difficult to solve, spatial economists have compiled a menu of techniques to ease the computational burden. The simplest is to assume that agents still solve static problems, be that because agents are explicitly myopic (Peters, 2022) or because, by assumptions put on technology and market structure, their dynamic objectives reduce to maximizing only current returns (Desmet and Rossi-Hansberg, 2014).[19] The closest such paper to the present one is Duranton and Puga (2023), who develop a model of growth and agglomeration in which agents are homogeneous and solve only one-period problems. By contrast, the method I propose allows me to specify a forward-looking migration and investment problem for heterogeneous workers, which I can then connect explicitly to the data and with which I can evaluate policies that have anticipatory or protracted effects. Another technique, called "dynamic hat algebra," involves specifying a dynamic discrete choice model across labor markets and solving for counterfactuals in changes from some baseline equilibrium under perfect foresight (Caliendo, Dvorkin, and Parro, 2019). In recent work, Kleinman, Liu, and Redding (2023) have extended this method to incorporate forward-looking capital accumulation. Crucially, though, the agents that accumulate capital in their model are *not* the workers who migrate; the two groups interact only through markets. In my model, workers migrate and accumulate human capital *jointly*, with the rate of the latter depending on the destination choice of the former.[20] A final set of recent papers use similar techniques to the ones used here (Martellini, 2022; Bilal, 2023; Greaney, 2023). This paper extends those methods to analyze balanced growth paths.

The rest of the paper is organized as follows. Section 2 presents the model and its predictions. Section 3 describes the quantitative procedure that casts the U.S. as a dynamic spatial knowledge economy. Throughout, I highlight salient features from the data on the U.S. urban system and workers' migration and human capital investments within it. Section 4 uses the quantified model to evaluate the proposed place-based policy. Section 5 concludes.

of location. We leave it to future work to study dynamic and long-run implications of spatial policies when worker productivity or tastes can change over time through skill formation or as a function of the skill mix in the community."

[18]See Desmet and Rossi-Hansberg (2010) and Rossi-Hansberg (2019) for reviews. The latter advocates for the type of work I do in this paper: "On the theoretical side, [existing] frameworks are still missing fully forward-looking dynamics in contexts where agents naturally care about the impacts of their actions in the future. Modeling forward-looking behavior in dynamic-spatial models with growth is essential when aiming to characterize optimal government policy, for example. . . . Progress on this front is urgent."

[19]See Proposition 1 of their paper and the surrounding discussion. To date, Desmet and Rossi-Hansberg's paper and follow-up work (see, e.g., Desmet, Nagy, and Rossi-Hansberg, 2018) have been the best available solution to the hard problem.

[20]Recent work by Dvorkin (2023) incorporates life-cycle human capital and wealth accumulation into a framework that can be solved with dynamic hat algebra. Cai et al. (2023) incorporate productivity growth into the framework of Kleinman, Liu, and Redding (2023), but growth there is still only semi-endogenous, which precludes studying the "hard problem."

2 Model

2.1 Environment

Time is continuous and indexed by $t \in \mathbb{R}_+$. There is a finite number of cities $n = 1, 2, \ldots, N$, with their set denoted by $\mathcal{N}$. There are two goods in the economy: the consumption good, which is homogeneous and freely traded, and land. The consumption good is the numeraire.

The economy is populated by a mass L of risk-neutral workers who discount their future flow utility at rate ρ and live a fixed length of time A, with age indexed by a. Upon death, a worker is replaced by an entrant somewhere in the economy.[21] Each worker born in city n at date t is endowed with a unit stream of raw labor and an initial level of human capital, indexed by z, drawn from a distribution with density $\underline{g}_n(z,t)$, to be specified later. The spatial distribution of human capital at any moment t has density $g_n(a,z,t)$, to be determined in equilibrium. If one integrates this density over its full support of ages and human capital levels and then scales by the total population L, the result is the population of city n at date t, denoted by $L_n(t)$.

Workers derive utility from the consumption good and land. They cannot write debt contracts with each other, hence they split all of their income between the two goods at each date. Land is assumed to be a strict necessity: after renting one unit of it, workers spend all of their remaining income on the consumption good. A resident of city n with human capital z at date t pays the flow cost

$$P_n(z,t) \equiv \left[p_n L_n(t)^{\theta_n}\right] z \tag{1}$$

for her accommodation in that city. This functional form can be microfounded with a monocentric city framework wherein commuting costs are paid in forgone earnings, which will be proportional to a worker's human capital (see Appendix A.1). Without loss of generality, the flow utility from land can be normalized to zero. Thus, the worker's income allocation problem is always static and trivial, and her flow utility is proportional to her flow of the consumption good. An exogenous local amenity, B_n, multiplies this consumption flow.

Workers make forward-looking decisions over *where* and *how* to allocate their labor. We start with *where*. Migration between cities is subject to two distinct frictions: one governs the frequency of migration; the other, how much it costs. First, each worker receives the opportunity to move across cities according to a Poisson process with arrival rate $\lambda > 0$, else she must stay where she is. This proxies for the fixed costs of moving in a manner that greatly simplifies the analysis.[22] Second, if a worker chooses to move to city i from city n, she must pay a permanent multiplicative flow-utility cost $1/\tau_{ni}$, with $\tau_{nn} = 1$ for all $n \in \mathcal{N}$ and $\tau_{ni} > 1$ otherwise. Insofar as these bilateral costs are correlated with distance, they impose a geography on the urban system.

Upon arrival of each migration opportunity, a worker receives a set of preference shocks, one for each city $n \in \mathcal{N}$, that are distributed i.i.d. Fréchet with shape parameter $\epsilon > 1$. Like migration costs, the preference shock for the chosen destination is multiplicative and

[21] That all workers live for the same fixed length of time and are replaced one-for-one upon death implies that the marginal distribution of age must be uniform for it to be stationary.

[22] For one, it simplifies the worker's problem, since she needs only choose a destination *conditional* on being mobile, not the timing of the move itself. Moreover, it simplifies how the human capital distribution evolves, since those who are mobile are just a random subset of the whole population and are therefore distributed according to the same law.

permanent. Accordingly, a worker's flow utility is the product of her flow of the consumption good and the accumulation of her preference realizations and migration costs along her chosen migration history.[23] Since we will soon take expectations over these preference shocks, I relegate the full technical description to Appendix A.2 to save on notation in the main text.

Now we turn to *how* each worker allocates her labor in a given location. A worker's income equals her output. Production uses labor only and is linear in a worker's effective labor supply, which is the product of her raw labor allocated to working, $1-s$, and her human capital, z. Production in each city benefits from static agglomeration economies: labor productivity in city n is log-linear in the local population with a city-specific scale, T_n, but common agglomeration elasticity, α.[24] It follows that a worker in city n supplying $l = (1-s)z$ units of effective labor will produce (hence earn)

$$y_n(l,t) = [T_n L_n(t)^\alpha]\, l \tag{2}$$

units of the consumption good.

Whatever labor she does not use in production is used to invest in her human capital. The key assumption is that there are local externalities in this investment technology. In particular, let

$$\frac{dz}{dt} = \kappa(s) z^\beta Z_n(t)^{1-\beta}, \quad \beta \in (0,1), \tag{3}$$

where $\kappa : [0,1] \mapsto \mathbb{R}_+$ is an increasing, concave function and $Z_n(t)$ is a local human capital index, the *vibrancy* of city n, given by

$$Z_n(t) \equiv \left(L \iint z^\zeta\, g_n(a,z,t)\, dz\, da \right)^{\frac{1}{\zeta}}, \quad \text{for } \zeta > 0. \tag{4}$$

The vibrancy is a weighted measure of the human capital in a city, accounting for both the size of the city and the distribution of human capital within it, where ζ governs the relative weights of those two factors. As $\zeta \to \infty$, a city's vibrancy is determined entirely by the city's maximum level of human capital; as $\zeta \to 0$, it is determined entirely by the city's size. An alternative expression for the vibrancy makes this clear:

$$Z_n(t) = L_n(t)^{\frac{1}{\zeta}} \left(\iint z^\zeta \frac{g_n(a,z,t)}{\ell_n(t)}\, dz\, da \right)^{\frac{1}{\zeta}} \equiv L_n(t)^{\frac{1}{\zeta}}\, \bar{z}_{n,\zeta}(t),$$

where $\ell_n(t) \equiv L_n(t)/L$ is the population share of city n at time t and thus $\bar{z}_{n,\zeta}(t)$ is the power-ζ mean of human capital within city n at time t. When $\zeta = 1$, the vibrancy decouples into the product of the city's size and its arithmetic mean of human capital.

Equations (3) and (4) encode many of the key features of the model. I will highlight six of them.

Dynamic agglomeration—First, holding fixed the marginal distribution of human capital within a given city, the returns to investment are strictly increasing in the city's size. This is a dynamic agglomeration economy—a mechanism through which city size has a prolonged positive effect on local productivity—because it quickens the accumulation of human capital by city residents. Contrast this with the static agglomeration economies introduced in Equation (2), which have only an instantaneous effect.

[23] Desmet, Nagy, and Rossi-Hansberg (2018) and Bilal (2023) use similar specifications for preferences.

[24] Examples of static agglomeration economies include sharing the gains from individual specialization and improving the quality of matches in a frictional labor market. See Duranton and Puga (2004) for a review.

Marshallian knowledge spillovers—Second, unlike in models where workers search randomly for more-productive partners from whom to learn, here workers make continuous use of the entire stock of local human capital, $Z_n(t)$, as they forge their own. This is partly for technical convenience,[25] but it also reflects an observation about local economies first made by Marshall (1890, p.271): "The mysteries of trade become no mysteries; but they are as it were in the air." Of particular note is that a worker need not meet a partner who has *more* human capital than she does in order to learn.[26] Like chefs swapping recipes, software developers swapping code snippets, or a chef and a developer discussing the wisdom of *mise en place*, both can learn regardless of relative human capital.[27] Equation (3) just takes the further step of asserting that meetings are sufficiently unconstrained in cardinality, frequency, and duration that we lose little by simply endowing workers with access to stock of local knowledge, defined in Equation (4).

Moreover, there is no segmentation of the population into inventors/entrepreneurs, for whom ideas matter, and others, for whom they do not. All benefit from access to knowledge in the city. As Glaeser (1999, p.255) puts it, reflecting on the same quote from Alfred Marshall, "As impressive as the role of cities in generating new innovations may be, the primary informational role of cities may not be in creating cutting edge technologies, but rather in creating learning opportunities for everyday people. ... [Marshall's quotation] pertains to young apprentices learning commonly known mechanical skills in industrial centers, not just to cutting edge innovations."

Learning complementarity—Third, the investment technology is supermodular in a worker's own human capital and the vibrancy of her location. So highly productive workers benefit more from living in big cities, especially when those big cities are populated with other highly productive workers. This complementarity pushes toward sorting of highly productive workers into bigger cites (Gaubert, 2018). Sorting will not be perfect, though, because other complementarities and margins of heterogeneity affect migration choices, as we discuss more in Section 2.3.

Nesting of Ben-Porath model—Fourth, ignoring the external effect from vibrancy, this is the classical form of the human capital investment technology dating back to Ben-Porath (1967).[28] Workers who live finite lives make on-the-job investments in their general human capital, using a technology that is concave in both time invested and their current human capital, forgoing part of their wage, which is proportional to their current human capital, to do so. The Ben-Porath model specifies $\kappa(s) = \frac{\bar{\kappa}}{\eta} s^{\eta}$ with $\eta \in (0, 1)$, so I will use that functional form in all that follows. Recent empirical work has focused on dissecting general human capital into firm-, industry-, occupation-, and even task-specific components, but the classical model does a good job of capturing the life cycle dynamics we are focused on here (Heckman, Lochner, and Taber, 1998; Sanders and Taber, 2012). We will discuss this further once the worker's dynamic program has been presented in Section 2.2.1.

[25] Random search for an individual learning partner with Poisson arrival would require the law of motion for a worker's human capital to be a jump process, where her human capital jumps upon meeting a partner, with the size of the jump potentially dependent on her own human capital and that of her partner.

[26] Davis and Dingel (2019, p.159, fn.13) argue that models of idea exchange in which only the lower-productivity agent benefits from a random meeting, like those in Lucas and Moll (2014) and Perla and Tonetti (2014), are poorly-suited to be embedded in a spatial economy.

[27] Berliant, Reed, and Wang (2006) and Berliant and Fujita (2008) consider models in which agents possessing vectors of task-specific knowledge search for partners with whom to exchange ideas in order to improve their productivity. Any agent can learn from any other so long as the other's knowledge vector is not a strict subset of her own.

[28] Heckman (1976) and Rosen (1976) are other seminal contributions.

Costly knowledge spillovers—Fifth, the investment technology bears a purposeful resemblance with the analogous technology in Davis and Dingel (2019). Davis and Dingel were the first to microfound agglomeration as an outcome of costly idea exchange within a system of cities.[29] Both our technologies require costly time investment and feature complementarities between an individual's skill and the size of and average skill within her city. A key difference between our models, however, is that, because their model is static, the return from learning in their environment is instantaneous, whereas in mine the return accrues over time.[30]

Affordance of balanced growth—Finally, the investment technology exhibits constant returns in the pair of human capital measures, (z, Z_n). Because production is linear in human capital, constant returns in the accumulation thereof is a necessary condition for balanced growth (Jones, 1999). Also, because κ is increasing and concave, the technology exhibits diminishing returns in time invested, s. Since production is linear in time, the concavity is needed to ensure that optimal time allocations are interior.

2.2 Equilibrium

2.2.1 Worker's problem

A worker chooses her labor allocation to maximize her lifetime utility, subject to the learning process (3) and the migration frictions outlined above, taking the path of local populations and vibrancies as given. I show in Appendix A.2 that, after taking expectations over her idiosyncratic preference shocks, the worker's problem can be written recursively as the following Hamilton-Jacobi-Bellman equation,

$$\begin{aligned}\rho V_n(a,z,t) = \max_{s\in[0,1]} &\left\{ B_n \left[T_n L_n^{\alpha}(1-s) - p_n L_n^{\theta_n}\right] z + \partial_z V_n(a,z,t)\left[\kappa(s) z^{\beta} Z_n^{1-\beta}\right]\right\} \\ &+ \lambda \sum_i m_{ni}(a,z,t)\left[\xi_{ni}(a,z,t) V_i(a,z,t) - V_n(a,z,t)\right] \\ &+ \partial_a V_n(a,z,t) + \partial_t V_n(a,z,t),\end{aligned} \tag{5}$$

with terminal condition

$$V_n(A,z,t) = 0 \quad \text{for all } (n,z,t), \tag{6}$$

where I have omitted the time arguments from local characteristics, and where

$$m_{ni}(a,z,t) = \frac{\tau_{ni}^{-\epsilon} V_i(a,z,t)^{\epsilon}}{\sum_k \tau_{nk}^{-\epsilon} V_k(a,z,t)^{\epsilon}} \tag{7}$$

is the optimal migration share, whose functional form follows from the preference shocks being distributed Fréchet with shape parameter ϵ, and

$$\xi_{ni}(a,z,t) = \frac{1}{\tau_{ni}} m_{ni}(a,z,t)^{-\frac{1}{\epsilon}} \tag{8}$$

[29] Farrokhi (2021) develops and estimates a static quantitative spatial model with agglomeration forces that stem from costly idea exchange.

[30] Davis and Dingel (2019) have a variable called the "value of the local idea-exchange environment" that is analogous to what I call the vibrancy of the city. Their construction is more complicated insofar as it weighs levels of human capital not simply by their prevalence (my g) but also by the time that workers with each level of human capital invest in learning (my s). To copy their construction would require an additional loop in the algorithm I present in Section 3, so I decline to do so.

accounts for the selection effect that accompanies optimal migration.[31]

The Hamilton-Jacobi-Bellman equation has seven distinct components. The left-hand side is the expected annuity value at date t of a worker of age a with human capital z living in city n. The first term on the right-hand side is that worker's consumption flow, the difference between her income (2) and her urban cost (1), multiplied by the relevant local amenity. The second term accounts for the change in her value function that comes from human capital investment according to (3) and (4). A max operator with respect to raw labor, s, encases the first two terms as the worker trades off consumption today with future human capital. I denote by $s_n(a,z,t)$ the solution to this maximization problem. The third term is the option value of moving, of which the right to exercise, recall, arrives at rate λ. With probability $m_{ni}(a,z,t)$, defined in (7), the worker exchanges the value of her current location n for that of her new location i. The latter is multiplied by $\xi_{ni}(a,z,t)$, defined in (8), to account for the underlying selection of her idiosyncratic optimal destination.[32] Notice that even when a worker chooses to stay put, which occurs with probability $m_{nn}(a,z,t)$, she still receives the selection bonus. Now looking at the last line, the fourth term on the right-hand side captures the aging of the worker. The fifth term accounts for the change in value over time as the economy evolves.[33] Finally, the terminal condition in (6) states that the value at death is zero regardless of location, human capital, or date.

2.2.2 State evolution

Local populations and vibrancies evolve in accordance with the spatial distribution of human capital, summarized by its density g. The density follows a Kolmogorov forward equation

$$\begin{aligned}\partial_t g_n(a,z,t) = -\partial_z[h_n(a,z,t)g_n(a,z,t)] - \lambda[1-m_{nn}(a,z,t)]g_n(a,z,t)\\ + \lambda\sum_{i\neq n} m_{in}(a,z,t)g_i(a,z,t) - \partial_a g_n(a,z,t),\end{aligned} \tag{9}$$

with initial condition $g_n(0,z,t) = \frac{1}{A}\underline{g}_n(z,t)$, where

$$h_n(a,z,t) = \kappa[s_n(a,z,t)]z^\beta Z_n(t)^{1-\beta} \tag{10}$$

is the optimally-controlled drift of an individual worker's human capital and, recall, $\underline{g}$ is the density from which each new worker's human capital is drawn.[34]

The Kolmogorov forward equation has six distinct components. The left-hand side is the instantaneous change in the density at a particular point (n,a,z) of the state space at date t, so it will be natural to describe the right-hand side in terms of flows of workers into or out of that point of the state space. The first term on the right-hand side is the outflow along the productivity dimension that comes from investment, where $h_n(a,z,t)$, defined in (10), measures the speed of the flow and $g_n(a,z,t)$ measures how many workers are "traveling" at that speed. The second term is the gross outflow of (a,z)-workers from

[31] Workers select their optimal migration destination after observing their idiosyncratic preference shocks. The mean value of a worker conditional on this choice is then strictly greater than the unconditional mean.

[32] Because $\xi_{ni}(a,z,t)$ is just a transformation of $m_{ni}(a,z,t)$, the Hamilton-Jacobi-Bellman in (5) is equivalent to that of a representative agent in state (n,a,z,t) who maximizes with the set of *continuous* control variables $\{s_n(a,z,t), \mathbf{m}_{n\cdot}(a,z,t)\}$ subject to the constraint $\sum_i m_{ni}(a,z,t) = 1$. See Appendix A.2 for a more detailed discussion.

[33] Although the pace of aging conforms with calendar time, the fourth and fifth terms are not the same. One can see this by holding t fixed and thinking about what it means to vary a, and vice versa.

[34] See Appendix A.3 for a heuristic derivation of the Kolmogorov forward equation.

city n. The third term is the gross inflow of the same type of workers from other cities. Flows go both ways because of the idiosyncratic preference shocks. In general, the two terms will not cancel because different cities have different marginal densities at (a, z). The fourth term is the outflow along the age dimension. Finally, the initial condition states that the marginal human capital density at entry is given by $\underline{g}$.

2.2.3 Equilibrium

In an equilibrium, workers invest and migrate optimally given the evolution of location characteristics, the underlying distribution evolves as workers' choices dictate, and at every moment the distribution integrates properly to the location characteristics. A formal definition follows.

Definition 1. An *equilibrium* is a tuple of functions $\{V, s, m, g\}$ on $\mathcal{N} \times [0, A] \times \mathbb{R}_{++} \times \mathbb{R}_+$ and a tuple of functions $\{\mathbf{L}, \mathbf{Z}\}$ on $\mathcal{N} \times \mathbb{R}_+$ such that

(i) workers solve the Hamilton-Jacobi-Bellman equation (5), taking paths of populations $\mathbf{L}$ and vibrancies $\mathbf{Z}$ as given;

(ii) the density $g_n(a, z, t)$ evolves according to the Kolmogorov forward equation (9), taking workers' optimal decision rules as given;

(iii) populations and vibrancies satisfy their definitions given $g_n(a, z, t)$:

$$L_n(t) = L \iint g_n(a, z, t)\, dz\, da, \qquad Z_n(t) = \left(L \iint z^{\zeta}\, g_n(a, z, t)\, dz\, da \right)^{\frac{1}{\zeta}};$$

(iv) local population shares sum to one for all t:

$$1 = \sum_{n \in \mathcal{N}} \frac{L_n(t)}{L} = \sum_{n \in \mathcal{N}} \iint g_n(a, z, t)\, dz\, da.$$

The equilibrium of this economy is an example of a *mean field game*. In a mean field game, each agent is one of infinitely many rational players controlling her own state (here, her human capital, current location, and age, though the last one is uncontrolled). Agents are indistinguishable modulo their state, so an agent does not care what any other particular agent does, since the latter's action would be drowned out regardless. Instead, each agent only needs to track the *distribution* across states (here, the density g). The resulting Nash equilibrium is the natural extension of a rational expectations equilibrium to an economy populated by heterogeneous agents.[35]

2.2.4 Balanced growth paths

A complete analysis of this economy would require the ability to calculate solutions for all initial distributions. I limit myself in this paper to the analysis of a set of particular solutions for which the growth rate and the distribution of relative human capital are both constant over time.

[35] With a representative agent, the distribution is degenerate at what one would call the "aggregate state." An equilibrium would require that the state of the representative agent be consistent with this aggregate state, as in the "big K, little k trick" (Ljungqvist and Sargent, 2018, §12.8.1).

Definition 2. A *balanced growth path* (BGP) is a number $\gamma \in \mathbb{R}_+$ and functions (v, σ, μ, ϕ) on $\mathcal{N} \times [0, A] \times \mathbb{R}_{++}$ such that

$$\begin{aligned} V_n(a, z, t) &= e^{\gamma t} \quad v_n\,(a, x) \\ s_n(a, z, t) &= \qquad \sigma_n\,(a, x) \\ m_{ni}(a, z, t) &= \qquad \mu_{ni}\,(a, x) \\ g_n(a, z, t) &= e^{-\gamma t}\ \phi_n\,(a, x) \end{aligned}$$

and (V, s, m, g) is an equilibrium with initial condition $g_n(a, z, 0) = \phi_n(a, z)$, where $x \equiv ze^{-\gamma t}$ is relative human capital.

A BGP is an equilibrium path along which the marginal distribution of age is stationary and the marginal distribution of human capital evolves as a "traveling wave," holding its shape while shifting to the right at a constant rate γ. To hold its shape, the distribution must have its quantiles growing at the same rate: letting capital letters denote cumulative density functions, the BGP requires $G_n(a, z, t) = \Phi_n(a, ze^{-\gamma t})$ for any (n, a), hence the qth quantile $z_n^q(a, t)$, defined by $\Phi_n(a, z_n^q(a, t)e^{-\gamma t})$, must satisfy

$$z_n^q(a, t) = e^{\gamma t}[\Phi_n(a)]^{-1}(q).$$

That the value function and decision rules take the forms given in Definition 2 follows immediately.

If workers lived forever, each could learn ad infinitum, hence so too could the human capital distribution shift to the right. But workers here live only finite lives, so how the population turns over will play a critical role in enabling balanced growth. To that end, I assume that the human capital distribution of entrants, summarized by the density $\underline{g}$, tracks the human capital distribution of existing workers in the following way.

Assumption 1. The c.d.f. of initial human capital, $\underline{G}$, whose density is $\underline{g}$, satisfies two conditions: (i) its human capital quantiles grow at the same rate as those of G; (ii) the lower bound of its support is strictly positive and also grows at this rate.

For example, $\underline{G}$ could be uniform over an interval of fixed relative size around, say, the 25th-percentile of G. For another, $\underline{G}$ could be a two-sided truncation of G with the same bounds. In any case, Assumption 1 has the interpretation that successive cohorts enter the labor market with more human capital because, for instance, they spent an unmodeled preperiod (call it "compulsory schooling") learning from previous cohorts, whose median level of human capital is growing at rate γ. The productivity growth of labor market entrants is therefore still endogenous because it is piggybacking on that of current workers. This growth is also an additional externality because current workers do not account for the effect of their investments on future entrants.

Assumption 1 is necessary because it ensures two things. First, there is a well-defined detrended density $\underline{\phi}_n(x) \equiv e^{\gamma t}\underline{g}_n(z, t)$. Second, the lower bound of the equilibrium support will be strictly positive, so no worker will get stuck at $z = 0$, which is an absorbing state.

Under Assumption 1, the conditions of a balanced growth path can be restated in terms of relative human capital. First, restate the location fundamentals as

$$L_n \equiv L \iint \phi_n(a, x)\, dx\, da, \qquad X_n \equiv \left(L \iint x^{\zeta}\, \phi_n(a, x)\, dx\, da\right)^{\frac{1}{\zeta}}, \tag{11}$$

so $L_n(t) = L_n$ and $Z_n(t) = e^{\gamma t} X_n$ for all t. The former implies that the city size distribution is invariant along a BGP. The Hamilton-Jacobi-Bellman equation (5) becomes

$$\begin{aligned}(\rho-\gamma)v_n(a,x) = B_n &\left(T_n L_n^{\alpha}[1-\sigma_n(a,x)] - p_n L_n^{\theta_n}\right) x \\ &+ \partial_x v_n(a,x)\left(\kappa[\sigma_n(a,x)]x^{\beta}X_n^{1-\beta} - \gamma x\right) \\ &\quad + \lambda \sum\nolimits_i \mu_{ni}(a,x)\left[\xi_{ni}(a,x)v_i(a,x) - v_n(a,x)\right] \\ &\quad\quad + \partial_a v_n(a,x)\end{aligned} \tag{12}$$

with boundary condition $v_n(A,x) = 0$ for all (n,x). Notice that the relevant discount rate is now $\rho-\gamma$, which I assume is positive, and the relevant drift is now the difference between how much human capital the worker accrues, $\kappa[\sigma_n(a,x)]x^{\beta}X_n^{1-\beta}$, and how much the underlying human capital distribution shifts, γx. In the space of relative human capital, if a worker does not accumulate human capital quickly enough, she falls behind. The Kolmogorov forward equation (9) becomes

$$\begin{aligned}0 = -\partial_x &\left\{\left(\kappa[\sigma_n(a,x)]x^{\beta}X_n^{1-\beta} - \gamma x\right)\phi_n(a,x)\right\} \\ &- \lambda[1-\mu_{nn}(a,x)]\phi_n(a,x) + \lambda\sum\nolimits_{i\neq n}\mu_{ni}(a,x)\phi_i(a,x) \\ &- \partial_a\phi_n(a,x)\end{aligned} \tag{13}$$

with boundary condition $\phi_n(0,x) = \frac{1}{A}\underline{\phi}_n(x)$ for all (n,x). Total production on a BGP is given by

$$Y_n(t) = e^{\gamma t} L \iint T_n L_n^{\alpha}[1-\sigma_n(a,x)]x\,\phi_n(a,x)\,dx\,da, \qquad Y(t) \equiv \sum_n Y_n(t),$$

so production, too, grows at rate γ. The definitions of the detrended decision rules will be shown in the next section. All derivations are shown in Appendix A.4.

2.3 Properties of a balanced growth path

A balanced growth path must satisfy Equations (11)–(13) in addition to an equation that determines the growth rate γ. I will show that the growth rate can be written as a weighted average of returns to investment in human capital across workers, with larger weight given to the investment that happens in more vibrant cities. First, though, I establish properties of workers' decision rules and the cross-section of cities for a generic γ.

2.3.1 Characterization of worker decision rules and city aggregates

The following lemma will prove useful throughout this section.

Lemma 1. *The value function $v_n(a,x)$ is decreasing in age, a, and increasing in relative human capital, x. Moreover, the second partial $\partial_{x,x}v_n(a,x)$ is always non-positive, as is the cross partial $\partial_{a,x}v_n(a,x)$.*

Unsurprisingly, for a given relative human capital, a worker would prefer to be younger, and for a given age, she would prefer to be more productive. The signs on the second-order partial derivatives are not much more difficult to see. A formal proof is in Appendix A.4.

Worker decision rules—Consider the optimal investment rule along the BGP,

$$\sigma_n(a,x) = \left(\frac{\bar{\kappa}\,\partial_x v_n(a,x)}{B_n T_n L_n^{\alpha}} \left(\frac{X_n}{x}\right)^{1-\beta} \right)^{\frac{1}{1-\eta}}. \tag{14}$$

where, recall, $\kappa(\sigma) = \frac{\bar{\kappa}}{\eta}\sigma^{\eta}$. We are interested in how investment varies with worker states and city characteristics according to (14). For states, one must consider both the direct effect and the indirect effect that is mediated through the value function. For city characteristics, one must do the same, keeping in mind that the indirect effects will also be mediated through the composition of residents that the city attracts, which determines its vibrancy.

By Lemma 1, the investment rate is weakly decreasing in an individual's relative human capital, x, for any fixed age. Relative to a worker in the same city with less human capital, a more productive worker faces diminished returns (at rate β) to using her human capital in investment but undiminished returns to using it in production, so she optimally shifts more of her labor toward production. This does not imply by itself that her accumulation of human capital is slower—the higher x may more than offset the lower $\kappa(\sigma)$ in the drift component $\kappa(\sigma)x^{\beta}X_n^{1-\beta}$. Conditional on human capital, accumulation *is* slower for older workers, though, because Lemma 1 ensures that the marginal value of relative human capital $\partial_x v_n(a,x)$ is non-increasing in a.[36]

The effects of location are much more difficult to disentangle, at least in part because it is not obvious how $\partial_x v_n(a,x)$ will vary across cities for given arguments.[37] The direct effects on investment of better amenities, B_n, higher fundamental productivity, T_n, and larger city size, L_n, are all negative, as they raise the return to allocating labor to production today. The direct effect on investment of greater vibrancy, X_n, by contrast, is positive, as it raises the return to allocating labor to investment. The total effect of any of these city characteristics on investment, however, is, in general, ambiguous: changes in fundamentals, for example, will induce changes in city size and skill composition in equilibrium that are potentially offsetting on the incentive to invest.[38] These conclusions are summarized in the following proposition.

Proposition 1. *The optimal investment rule along the* BGP, $\sigma_n(a,x)$, *is non-increasing in both age,* a, *and relative human capital,* x, *but varies ambiguously across cities,* n, *for any state.*

Now consider the optimal migration rule along the BGP,

$$\mu_{ni}(a,x) = \frac{\tau_{ni}^{-\epsilon} v_i(a,x)^{\epsilon}}{\sum_k \tau_{nk}^{-\epsilon} v_k(a,x)^{\epsilon}}, \tag{15}$$

which specifies the probability with which a worker of age a and relative human capital x living in city n would choose to migrate to city i upon the arrival of an opportunity

[36] Contrast this with a perpetual youth model, wherein the age of an agent has no effect on her decisions because her hazard rate of death is always the same.

[37] Total land costs, for example, vary with both location and individual human capital, which means they affect σ indirectly through this term.

[38] The total effect on investment of increasing fundamental productivity in city n, for example, is

$$d\ln\sigma_n(a,x) = \left(\frac{1}{1-\eta}\right)\left[\frac{d\ln(\partial_x v_n(a,x))}{d\ln T_n} + \left(\frac{1-\beta}{\zeta} - \alpha\right)\frac{d\ln L_n}{d\ln T_n} + (1-\beta)\frac{d\ln \bar{x}_{n,\zeta}}{d\ln T_n}\right] d\ln T_n$$

where I split vibrancy into its size and composition components.

shock. As Equation (15) makes clear, how migration decisions vary with individual states and location characteristics is mediated almost entirely through the corresponding variation in the expected value function. This makes it difficult to tease apart how each model component affects migration rates. Nevertheless, one can make four observations.

First, workers are generally attracted to nearby or centrally-located vibrant cities with high amenities and high consumption net of urban costs (e.g., high T_n and low p_n), since each of these features either saves on τ today or raises the value v. The value is fully forward-looking, so it accounts for the effect of location on a worker's lifetime consumption profile, not just her current consumption. Workers may choose to move to a city that affords them lower consumption today if it allows them to accumulate human capital more quickly. This is consistent with the "location as an asset" view of Bilal and Rossi-Hansberg (2021).[39]

Second, as previewed in Section 2.1, sorting on individual human capital is imperfect. Because of the extreme-value specification of preference heterogeneity, the share in (15) is strictly positive for any state (n, a, x) and destination i, which implies that all ages and relative human capital levels will be present in any given city i.

That is not to say the proportions in which they are present will be the same, though, which brings one to the third observation: different cities may be disproportionately young or skilled or varied along either dimension. A city's particular mix of fundamentals, agglomeration and congestion forces, and skill composition will attract different types of workers to varying degrees.[40] A young worker, for example, would typically care more about the vibrancy of a city than would an old worker, because the young worker would typically invest more of her time in learning.

The final observation is that age also has a secondary effect on migration: as a worker approaches the end of her lifespan, her value function approaches zero for any (n, x). It follows from (15) that the bilateral costs $\{\tau_{ni}\}$ play an increasingly large role in determining migration decisions. With $\tau_{nn} = 1 < \tau_{ni}$ for all $i \neq n$, it further follows that the old move less, even for a fixed rate of opportunity shocks, λ. This is consistent with the evidence that I will discuss in Section 3.

These conclusions are summarized in the following proposition.

Proposition 2. *The optimal migration rule along the* BGP*, $\mu_{ni}(a, x)$, favors nearby or centrally-located vibrant cities with high amenities and high consumption net of urban costs, with different weights placed on each feature by workers in different states. Nevertheless, at least some workers in each state move to each destination.*

City aggregates—Cities result from the balance of agglomeration and congestion forces. Net of differences in fundamentals, larger cities along a BGP are more productive because $\alpha > 0$.[41] Larger cities are also generically more vibrant because $\zeta > 0$.[42] It follows that, for a fixed investment rate, the model generates a city size earnings premium with both static

[39] Bilal and Rossi-Hansberg (2021) focus on the substitution between the location asset and another riskless asset when there is a borrowing constraint on the latter, which I do not consider here, but their specification of human capital accumulation is more stylized.

[40] Of course, the size and composition of each city is itself an equilibrium outcome determined by the migration choices of different types of workers.

[41] Allowing for heterogeneous fundamentals will make it possible in Section 3 to match observed patterns in U.S. cities, where there is not a strict ordering of cities' productivity and costs by size.

[42] Technically, it is possible for a city's vibrancy to decrease with its size if its power mean of human capital declines sufficiently quickly with size, i.e., if

$$\frac{d \ln \bar{x}_{n,\zeta}}{d \ln L_n} < -\frac{1}{\zeta}.$$

and dynamic components: a worker's earnings jump upon moving to a bigger city, and her earnings profile gets steeper because her returns to investment are higher in the more vibrant place. This is consistent with the evidence in Glaeser and Maré (2001), Baum-Snow and Pavan (2012), and Duranton and Puga (2023) for U.S. cities, which I will use to quantify the model in the next section. Workers trade off the benefits of the immediate earnings premium and the steeper earnings profile in bigger cities against higher urban costs (with elasticity θ) and lower average flow utility (with elasticity ϵ) operating through the idiosyncratic preference shocks. These conclusions are summarized in the following proposition.

Proposition 3. *Net of differences in city fundamentals, larger cities along a BGP are always more productive and more costly. In addition, larger cities are generically more vibrant. The model therefore generates a city size earnings premium with both static and dynamic components.*

2.3.2 Characterization of the growth rate

The growth rate γ is determined by integrating (in the sense of Lebesgue) the Kolmogorov forward equation (13) over the state space and leveraging two facts: (i) there is no net migration on a BGP (Equation (11)), and (ii) the marginal density with respect to age is uniform. With a bit of calculus, solving for γ yields the following proposition.

Theorem 1. *A balanced growth path characterized by the functions $\{v, \sigma, \mu, \phi\}$ has the growth rate*

$$\gamma = \frac{\sum_n \int_0^A \kappa[\sigma_n(a,x)] x^\beta X_n^{1-\beta}\, \phi_n(a,x)\, da}{\sum_n \int_0^A x\, \phi_n(a,x)\, da}, \quad \forall x \in \text{supp}(\phi). \tag{16}$$

Proof. See Appendix A.4. □

The expression in (16) is a natural extension of the representative-agent, one-location case from Uzawa (1965). In what follows we will build intuition for (16) by rebuilding the present model step-by-step atop Uzawa's.

Uzawa (1965) as a foundation—Uzawa's model can be summarized in one optimization problem:

$$V[z(0)] = \max_{s \in [0,1]} \int_0^\infty e^{-\rho t}[1 - s(t)] z(t)\, dt$$
$$\text{s.t. } \dot{z}(t) = \kappa[s(t)] z(t),$$

where κ is increasing and concave.[43] That is, the representative consumer splits her labor between learning, with fraction s, and producing, with fraction $1 - s$, in order to maximize the present discounted value of her output, subject to the specified learning technology. Under the assumption that $\kappa(1) < \rho < \kappa(0) + \kappa'(0)$, the optimal time allocation s^* is interior, unique, and implicitly defined by $\kappa'(s^*)(1 - s^*) = \rho - \kappa(s^*)$. Moreover, a unique balanced growth path exists, and its growth rate is $\gamma \equiv \kappa(s^*) < \rho$.[44]

But, since we generally expect higher ability workers to sort *into* bigger cities because of the supermodularity in consumption and learning, I leave this aside as a pathological case.

[43] This statement of Uzawa's model abstracts from physical capital and population growth.

[44] This underscores a point first made by Mulligan and Sala-i Martin (1993): In Uzawa (1965) and Lucas (1988), although a human capital externality in production or investment would affect the rate of balanced growth, such an externality is not *required* to sustain balanced growth. It is sufficient that the investment technology be linear in z.

Reintroduce the learning externality—The representative agent in Uzawa's economy now faces the law of motion

$$\dot{z} = \kappa(s) z^{\beta} Z^{1-\beta},$$

where z is her own human capital and Z is the aggregate stock, which the agent takes as given. Of course, in equilibrium the two must be the same, but because the agent does not internalize her effect on the aggregate, she underinvests. Indeed, her optimal time allocation s_1^* becomes implicitly defined by $\kappa'(s_1^*)(1 - s_1^*) = \rho - \beta\kappa(s_1^*)$, hence it is strictly less than s^*. By the monotonicity of κ, the growth rate, too, is lower than before.

Reintroduce the finite investment horizon—Suppose that instead of living indefinitely, the representative agent dies with a fixed hazard rate δ, after which she bequeaths her human capital to her replacement.[45] Then optimal time allocation s_2^* becomes implicitly defined by $\kappa'(s_2^*)(1 - s_2^*) = \rho + \delta - \beta\kappa(s_2^*)$, hence it is strictly less than s_1^*. With a shorter horizon over which to reap the returns from her human capital, she invests less. By the monotonicity of κ, the growth rate, too, is even lower.

Reintroduce worker heterogeneity—Now instead of a representative agent, let workers vary in their human capital. Let workers vary in their age, too, by imposing that workers die after a fixed length of time instead of at a fixed hazard rate. This implies that there will be a distribution of workers over the age-by-human capital state space, the density of which I denote by g. This distribution will evolve over time, so we need a Kolmogorov forward equation to specify precisely how it evolves. Its evolution will determine how the aggregate stock of human capital grows, because the stock must now be computed as an integral with respect to this distribution, $Z(t) = L \iint z\, g(a, z, t)\, dz\, da$.[46] Workers will choose different investment rates depending on their state, so—unlike in the last two steps—we cannot identify a single rate s_3^* that pins down the growth rate as $\kappa(s_3^*)$. But what we *can* do should be intuitive: the growth rate is a weighted average of returns to time invested, $\kappa[\sigma(a, x)]$, and indeed I find that the correct weighted average is

$$\gamma = \frac{\int_0^A \kappa[\sigma(a, x)] x^{\beta} X^{1-\beta}\, \phi(a, x)\, da}{\int_0^A x\, \phi(a, x)\, da}, \quad \forall x \in \operatorname{supp}(\phi),$$

where I use the same notation for detrended functions and variables as we have been using above.[47]

Reintroduce cities and migration—Finally, allow the workers to migrate between a finite set of cities and restrict the learning externality to be local in scope. Then the growth rate becomes as stated in (16). The growth rate is again a weighted average of the returns to time invested, but now more vibrant locations get more weight. Despite the fact that cities vary in their vibrancy, all cities grow at the same rate γ on the BGP because migration acts as a form of knowledge diffusion. As workers flow between cities according to (15), they bring their human capital with them.[48] So although there is no net migration on the BGP,

[45] This is not the same as specifying a finite lifespan for each agent, but that specification would imply heterogeneity, which I wait to introduce in the next step.

[46] This is the simplest possible stock measure. My definition of vibrancy takes the additional step of introducing curvature, governed by the parameter ζ, in order to adjust the weight placed on different features of the skill distribution.

[47] Note that the average is taken only across ages and must be equated for each level of relative human capital. This ensures that each quantile of the distribution grows at the common rate.

[48] Contrast this with models like that in Eaton and Eckstein (1997), which must assume an ad hoc technology diffusion across cities.

the gross migration of workers through the urban system spreads the gains from human capital accumulation across the whole system.

Summary—The preceding steps reveal a few key insights. Most importantly, they reveal that the spatial distribution of human capital *does* matter for aggregate growth. Indeed, the aggregate growth rate (16) is the weighted average of returns to investment in human capital across workers, with larger weight given to the learning that happens in more vibrant (roughly, bigger and more skilled) cities. Investment rates vary with workers' age, human capital, and location. Relative to an infinitely-lived planner, workers underinvest because (i) they have a shorter horizon over which to reap the benefits of their human capital and (ii) they do not internalize the local learning externality.

2.3.3 Well-posedness of a balanced growth path

Proposition 1 established that if a balanced growth path characterized by the functions $\{v, \sigma, \mu, \phi\}$ exists, it must have the growth rate specified in Equation (16). This left aside whether there exists an entire path $\{v, \sigma, \mu, \phi, \gamma\}$ that solves (11)–(15) and (16) uniquely.

To date, the well-posedness of solutions to this class of coupled partial differential equations has not been proven.[49] Nevertheless, I take the following measures to ensure that my results are well-behaved. First, I impose $\beta\kappa(1) < \rho < \beta\kappa(0) + \kappa'(0)$, which guarantees the existence of a unique solution in the Uzawa model with an externality described in Section 2.3.2.[50] Second, I show in Section 3 that a solution exists to the discretized version of the BGP system under my chosen parameters. Those parameters are chosen so that the BGP matches a number of moments in the data. Re-solving the model with those parameters from different initial conditions consistently recovers the same BGP, suggesting that the observed BGP is locally unique and stable. Third, when solving the policy counterfactual in Section 4, I follow Ahlfeldt et al. (2015) in selecting the closest counterfactual BGP by using the baseline BGP as the initial value for the solver.

3 Quantitative analysis

In this section I describe how I quantify and compute a BGP. I begin by describing the algorithm I use to compute a BGP given model primitives. I then detail how I quantify those primitives using U.S. data. Finally, I plot workers' decision rules and city characteristics in order to compare them to targeted and untargeted moments from the data.

3.1 Algorithm to compute a BGP

The algorithm, adapted from Achdou et al. (2022), is a fixed-point algorithm that iterates on a vector of length $2N+1$ containing the aggregate variables $\{\mathbf{L}, \mathbf{X}, \gamma\}$. Within each iteration, I use finite difference methods to solve the coupled partial differential equations (12) and (13) that characterize allocations along the BGP. This involves discretizing the continuous state

[49] In standard quantitative spatial models, the existence of a unique equilibrium is typically established by a condition on the agglomeration and congestion elasticities that translates as "congestion forces always dominate agglomeration forces" (Redding and Rossi-Hansberg, 2017). Such a condition cannot be stated here in part because the dynamic agglomeration force depends on the marginal distribution of human capital within each city, which is a nonparametric equilibrium object.

[50] With $\kappa(\sigma) = \frac{\bar{\kappa}}{\eta}\sigma^{\eta}$ for $\eta \in (0, 1)$, this translates to simply $\beta\bar{\kappa} < \rho\eta$ since $\kappa(0) + \kappa'(0) = \infty$.

space and leveraging the sparsity that comes from working in continuous time to compute and manipulate a large transition matrix.[51] Appendix C provides the full details.

Algorithm 1: Compute a BGP

Data: parameters $\{A, \rho, \lambda, \epsilon, \alpha, \zeta, \beta, \eta, \bar{\kappa}, x_{\text{scale}}\}$,
location fundamentals $\{B_n, T_n, p_n, \theta_n, \tau_{ni}\}$, and
relative productivity distribution of entrants $\underline{\phi}$

Result: BGP $\{v, \sigma, \mu, \phi; \gamma\}$

Initialize with guess $\{\mathbf{L}^0, \mathbf{X}^0, \gamma^0\}$. Denote iterations by $\iota = 0, 1, 2, \ldots$;

while *not converged* **do**

Given $\{\mathbf{L}^\iota, \mathbf{X}^\iota, \gamma^\iota\}$, solve the Hamilton-Jacobi-Bellman equation (12) using a finite difference method and calculate the decision rules $\sigma_n^\iota(a,x)$ and $\mu_{ni}^\iota(a,x)$; 1

Given $\sigma_n^\iota(a,x)$ and $\mu_{ni}^\iota(a,x)$, solve the Kolmogorov forward equation (13) for $\phi_n^\iota(a,x)$ using a finite difference method; 2

Given $\phi_n^\iota(a,x)$, compute the corresponding local populations and vibrancies 3

$$\tilde{L}_n^\iota = L \iint \phi_n^\iota(a,x)\, dx\, da, \qquad \tilde{X}_n^\ell = \left(L \iint x^\zeta \phi_n^\ell(a,x)\, dx\, da \right)^{\frac{1}{\zeta}}$$

for each n, and compute the implied growth rate

$$\tilde{\gamma}^\iota = \frac{\sum_n \int_0^A \kappa[\sigma_n^\iota(a,x)] x^\beta (X_n^\iota)^{1-\beta}\, \phi_n^\iota(a,x)\, da}{\sum_n \int_0^A x\, \phi_n^\iota(a,x)\, da}, \quad x \in \operatorname{supp}(\phi^\iota)$$

as defined in (16);

if $\{\tilde{\mathbf{L}}^\iota, \tilde{\mathbf{X}}^\iota, \tilde{\gamma}^\iota\}$ *close enough to* $\{\mathbf{L}^\iota, \mathbf{X}^\iota, \gamma^\iota\}$ **then**

converged;

else

construct $\{\mathbf{L}^{\iota+1}, \mathbf{X}^{\iota+1}, \gamma^{\iota+1}\}$ as a linear combination of the previous guess and the computed values;

end

end

3.2 Quantification

The **Data** line of Algorithm 1 lists the model primitives that must be set in order to calculate a BGP. To quantify them, I proceed in multiple steps. Throughout, I will make a number of choices that facilitate matching estimates from the human capital literature and Duranton and Puga (2023, Table 2). All data descriptions are in Appendix B.

Select cities and set congestion elasticities, $\{\mathcal{N}, \theta_n\}$

I start with the 378 Metropolitan Statistical Areas (MSAs) that had populations exceeding 75,000 people in the 2010 U.S. Census.[52] Because the number and sizes of cities matters for

[51] Rendahl (2022) demonstrates under what conditions similar computational speed can be achieved in discrete time.

[52] Throughout, I define MSAs using the February 2013 delineation of the Office of Management and Budget (OMB) whenever possible. Exceptions are noted explicitly in Appendix B.

workers' decisions in the model, it is important to represent the entire U.S. urban system. At the same time, the number of parameters to be estimated scales linearly with the number of cities, and the state space scales quadratically, so incorporating all 378 MSAs individually is prohibitively costly. Instead, I assign cities to groups based on metropolitan population in 2010. Any MSA that had a population greater than 2 million (of which there are 30) is in a group by itself. Below 2 million, groups are incremented by 500,000, so there are 4 such groups and 34 groups in total. The crucial restriction is that cities in the same group are forced to be identical. So Rochester, NY and Memphis, TN, both with 2010 populations between 1 million and 1.5 million, for example, are modeled as two copies of the "1.0M-1.5M" city. Accordingly, there are 378 cities in the model but only 34 types of cities among which workers choose.

I set the congestion elasticity in each city equal to the elasticity of housing supply estimated by Saiz (2010).[53] These estimates are determined primarily by land availability and land use regulations in each city. In groups with multiple cities, I use the population-weighted mean elasticity.

Set migration parameters, $\{\lambda, \epsilon, \tau_{ni}\}$, using ACS data

The 2011–15 American Community Survey (ACS) Migration Flow files aggregate over five years of surveys to provide counts of respondents who changed residence within the last year, cross-tabulated by age bin.[54] For each age bin and origin MSA, one can observe the count of residents that (i) did not move, (ii) moved within the same MSA, (iii) moved to each of the other possible destination MSAs, or (iv) moved to a non-metropolitan or foreign location.

The parameter λ can be computed directly from these data. Recall that migration opportunities arrive according to i.i.d. Poisson processes at rate λ. It follows that a fraction λ of the population will be allowed to move each year. Since within-MSA moves are observed separately from non-moves, I simply set λ equal to the overall fraction of urban residents that move.[55] In light of the many age-varying determinants of migration that are not modeled explicitly here, however, I now allow λ to vary with age, $\lambda(a)$, which decreases monotonically from 26.6% at age 20 to 7.6% at age 59.[56]

The parameter ϵ governs the migration elasticity with respect to a worker's value, which is the present discounted sum of her lifetime real consumption. I set $\epsilon = 3$ based on estimates of the migration elasticity with respect to real wages at decadal frequency from Diamond (2016, Table 5, pp.505–8).

[53] In 19 cities, Saiz (2010) data are missing. In those cases, I follow Hsieh and Moretti (2019) in imputing the elasticity from the relevant state average.

[54] Age is broken down into the following fifteen categories: 1 to 4 years, 5 to 17 years, 18 to 19 years, then five-year bins through age 74, and a final bin for 75 years and over. Consistent with the age restriction I impose later, I only consider flows by those from age 20 through age 59.

[55] This approach will match overall mobility rates in the data, but it likely understates the rate at which one might think workers have the *opportunity* to move because it conflates potential movers who choose to stay in their current residence with those unable to move. All qualitative results are robust to using higher values of λ.

[56] Consider that the likelihood of marriage, child-rearing, and homeownership are all typically increasing from young adulthood through at least middle age, and each increases the cost of moving. See, for example, Mincer (1978) on family ties in general, Costa and Kahn (2000) on the colocation problem of married couples specifically, and Oswald (2019) on homeownership. Kennan and Walker (2011) estimate a positive direct effect of age on moving costs in addition to its effect on moving decisions through their human capital channel.

To compute the bilateral migration frictions, I use the calibrated migration elasticity to invert the implied frictions from the flows I observe in the ACS. In particular, I construct a Head-Ries migration index

$$\frac{\bar{\mu}_{ni}\bar{\mu}_{in}}{\bar{\mu}_{nn}\bar{\mu}_{ii}} = \frac{\tau_{ni}^{-\epsilon}\tau_{in}^{-\epsilon}}{\tau_{nn}^{-\epsilon}\tau_{ii}^{-\epsilon}}, \tag{17}$$

where $\bar{\mu}_{ni}$ is the average likelihood over all ages and relative productivities of moving from n to i conditional on moving.[57] Under the assumption that $\tau_{nn} = 1$ for all n and $\tau_{ni} = \tau_{in}$ for all n, i, the index reduces to

$$-2\epsilon \ln \tau_{ni} = \ln \left(\frac{\bar{\mu}_{ni}\bar{\mu}_{in}}{\bar{\mu}_{nn}\bar{\mu}_{ii}} \right), \tag{18}$$

where the right-hand side is directly observed. I do this first at the level of individual cities, where $\tau_{nn} = 1$ is a tenable normalization. To aggregate to the group level, I use the flow-weighted average of bilateral costs for moves across cities within the same group to set the group-level τ_{nn}. So in the 30 groups with only one city, we still have $\tau_{nn} = 1$, but in the 4 groups with multiple cities, we have $\tau_{nn} > 1$. With the group-level own cost τ_{nn}, I then compute the group-level bilateral cost τ_{ni} using the formula in (17) while maintaining the assumption of symmetry, where the flows are now summed across groups.

Set general human capital investment parameters, $\{A, \rho, \beta, \eta, \underline{\phi}_n\}$

As highlighted in Section 2, my framework embeds the classic model of general human capital investment from Ben-Porath (1967) into a spatial general equilibrium. A large empirical literature has estimated structural versions of the Ben-Porath model.[58] Insofar as I can demonstrate that the additional structure of my framework is consistent with or independent of the Ben-Porath model, I can credibly use these estimates to calibrate some of my parameters.

My calibration draws most heavily from three sources. The first is Browning, Hansen, and Heckman (1999), which reviews structural estimates of the Ben-Porath model. The second is Heckman, Lochner, and Taber (1998), which extends the Ben-Porath model to allow for an up-front choice of total schooling (high school vs. college) and coarse heterogeneity in initial stocks of human capital (four groups). The authors estimate their model using data on white males in years 1979–1993 from the NLSY. The third is Huggett, Ventura, and Yaron (2006), which extends the Ben-Porath model to allow for a continuous distribution of initial stocks of human capital. The authors estimate their model using data on males from the PSID 1969–1992 family files. Collectively, these three sources demonstrate that the Ben-Porath model is consistent with observed patterns such as the concentration at young ages of time allocated to skill acquisition and the steepness of age-earnings profiles for people who choose high amounts of schooling. Moreover, when augmented to include worker heterogeneity, the Ben-Porath model does well in replicating trends in mean earnings and measures of earnings dispersion and skewness over most of the working life-cycle for a

[57] Head and Ries (2001) first proposed an index of this sort to back out a measure of trade costs from observed trade flows. Taking averages across ages and relative productivities and later aggregating to groups ensures that the matrix of flows is dense and limits concerns about overfitting (Dingel and Tintelnot, 2021).

[58] An even larger literature estimates the Mincer (1974) model, which was originally derived as an approximation of the Ben-Porath model. Heckman, Lochner, and Todd (2006) surveys this literature.

typical cohort as the cohort ages.[59]

Recall a typical worker's decision problem as specified in the Hamilton-Jacobi-Bellman equation (12). She takes city aggregates and the pace of growth as given, so in quantifying her investment problem we are free to consider these variables as exogenous. The horizon of her decision problem is determined by her lifespan, A, and her discount rate, ρ. I set $A = 40$, so the lifespan accords with ages 20 through 59. This choice is motivated by a few considerations. First, within this age band, a uniform age distribution is an adequate assumption: in the 2010 Census, there were roughly 20 million people in each five-year age bin up to age 60. Second, near the full Social Security retirement age of 67 there is a substantial fall in labor force participation that occurs for reasons not considered in my model. This suggests using an earlier terminal age. Third, it is consistent with Huggett, Ventura, and Yaron (2006), who emphasize the relative lack of data availability in the PSID outside this age range. For the discount rate, I set $\rho = 0.10$, which exceeds the typical annual discount rate set in most macroeconomic models, to account for missing curvature in the worker's flow utility function.[60] In particular, with linear utility, workers may be willing to accept negative flow payoffs in exchange for sufficient wage growth.[61] A higher-than-usual discount rate stands in for a finite intertemporal elasticity of substitution to curb this behavior.[62]

Her investment technology is characterized by two elasticities: β, which governs the marginal productivity in investment of a worker's current human capital, and η, which governs that of her time. I set $\beta = 0.8$, which accords with the best-fitting parametric model estimated by Huggett, Ventura, and Yaron (2006, Table 2, Panel B, Column 4) and falls just shy of the estimates in Heckman, Lochner, and Taber (1998, Table 1).[63] With panels of only a few thousand workers at most, few studies using U.S. data can reject the neutral Ben-Porath model in which $\eta = \beta$. One implication of this neutrality, however, is that human capital accumulation raises the marginal *cost* of human capital investment and the marginal *productivity* of human capital investment in the same proportion. It follows that the optimal accumulation would be independent of x. This contradicts evidence from De la Roca and Puga (2017) on a much larger panel of Spanish workers that the additional value of experience in bigger cities is stronger for those with higher ability. With this in mind, I set $\eta = 0.7$, which still falls within the range of estimates reviewed by Browning, Hansen, and Heckman (1999).

The preceding discussion passed over two sources of potential misspecification. First, alongside her investment, the typical worker in my model makes a simultaneous migration decision. Notice, though, that her migration decision interacts with her investment deci-

[59] Huggett, Ventura, and Yaron (2006) show that heterogeneity in learning ability, not just initial human capital, is required to generate a positive trend in earnings dispersion in their model. This is not true in my framework because of heterogeneity in returns to learning across locations: if higher-skilled workers sort sufficiently into vibrant environments, my model can generate a positive trend in earnings dispersion without incorporating individual heterogeneity in learning ability.

[60] Huggett, Ventura, and Yaron (2006) use linear utility with the equivalent of $\rho = 0.04$, but they calibrate to much lower wage growth (0.14% per year) and a positive depreciation rate (1.14% per year). Heckman, Lochner, and Taber (1998) use $\rho = 0.05$ and zero depreciation but specify an isoelastic utility function with coefficient of relative risk aversion equal to 0.9.

[61] This occurs, for example, if a worker sets $\sigma = 1$ because she must still pay the urban cost.

[62] The goal is not necessarily to rule out this behavior, however, because accepting negative flow payoffs in exchange for sufficient wage growth is itself a form of borrowing and saving that is otherwise precluded in this model.

[63] Heckman, Lochner, and Taber (1998) estimate $\beta = 0.832$ (S.E. 0.253) for high school and $\beta = 0.871$ (S.E. 0.343) for college-educated workers.

sion only through the former's effect on the scale of the marginal value of human capital, $\partial_x v_n(a, x)$. That effect is itself mediated by the arrival rate of migration opportunities, λ. Accordingly, I expect the misspecification error from ignoring this interaction to be small. Second, the worker's investment technology includes the vibrancy of her current city, so it is $\frac{\bar{\kappa}}{\zeta}\sigma^{\zeta}x^{\beta}X_n^{1-\beta}$ rather than simply $\frac{\bar{\kappa}}{\zeta}\sigma^{\zeta}x^{\beta}$.[64] This difference would be wholly irrelevant if there were a uniform vibrancy $X_n \equiv X$ across locations, in which case the missing $X^{1-\beta}$ would simply be subsumed in the estimate of $\bar{\kappa}$, or if there was no sorting on skill across locations, in which case matching worker panels pooled across locations would still yield an unbiased estimate of β. The main concern is that workers with high levels of human capital sort (albeit imperfectly) into more vibrant places, in which case existing estimates of β would be biased upward. With that in mind, I verify that my main results are robust to using lower values for β and η.

It remains to parameterize the initial human capital density for each city, $\underline{\phi}_n$. Huggett, Ventura, and Yaron (2006) demonstrate that a Ben-Porath model with a lognormal distribution of initial human capital stocks provides a good fit of key patterns in the U.S. earnings distribution. Accordingly, I set $\underline{\phi}_n$ to be a discretized lognormal density. I truncate the distribution at its fifth and ninety-fifth percentiles to ensure that its support lies within a finite state space and has a strictly positive minimum. I set its coefficient of variation to 0.468 from Huggett, Ventura, and Yaron (2006, Table 5, Panel B, Column 4), which corresponds to a neutral human capital model with elasticity $\beta = 0.8$ and accumulation starting at age 20. To set the mean in each city, I turn to Heckman, Lochner, and Taber (1998, Table 2), who estimate the initial level of human capital separately for high-school and college-educated workers. They estimate an initial premium of up to 50% for college-educated workers. I use the weighted average of these initial levels, weighting by the college share in a given city from the 2011–15 ACS five-year estimates, to set the mean for that city up to a choice of scale, which will be estimated later.[65] Finally, I set the total mass of entrants in each city to match observed population shares of 15–19 year olds in the 2010 ACS one-year sample.

Estimate $\{\alpha, \zeta, x_{\text{scale}}, T_n, B_n, p_n, \bar{\kappa}\}$ using a minimum distance estimator

The remaining parameters are estimated jointly to minimize the distance between the model and the data regarding five sets of moments: (i) panel estimates of the returns from big city experience from Duranton and Puga (2023, Table 2), (ii) total employment of wage and salary workers in each city per the 2013 BEA Regional Accounts, (iii) the total wage bill in each city per the 2013 BEA Regional Accounts, (iv) constant housing expenditure shares across all cities, (v) a 2% annual growth rate.[66] Although the parameters are estimated jointly, their identification is rather intuitive.

The first three parameters, $\{\alpha, \zeta, x_{\text{scale}}\}$, are primarily identified by the panel estimates from Duranton and Puga (2023, Table 2). Duranton and Puga (2023) follow De la Roca

[64] The restriction that the elasticity on the vibrancy is $1-\beta$ rather than a generic $\tilde{\beta}$ is mostly immaterial: one can simply toggle the value of ζ, the inverse elasticity of vibrancy with respect to city size, to affect the strength of city characteristics in enhancing workers' human capital investments.

[65] Previous estimates suggest that sorting on unobserved ability is negligible after controlling for education (Baum-Snow and Pavan, 2012; De la Roca, Ottaviano, and Puga, 2023).

[66] In static spatial frameworks with a discrete number of agent types, it is common for there to be a one-to-one mapping from the observed data on the endogenous variables of the model to the structural fundamentals of the model. This mapping can then be inverted to identify the unique values of the estimated structural fundamentals that exactly rationalize the observed data as an equilibrium (Redding and Rossi-Hansberg, 2017). Such is not the case here because the endogenous variables depend on the whole underlying spatial distribution of human capital, which is unknown.

and Puga (2017) in first estimating an individual earnings regression on a panel of workers from the restricted-access NLSY Geocodes (1979–2012):

$$\ln y_{nt}^{j} = a_n + a_j + a_t + \sum\nolimits_n b_n e_{nt}^{j} + \mathbf{C}_t^{j}\mathbf{b} + \varepsilon_{nt}^{j},$$

where a_n is a city fixed effect, a_j is a worker fixed effect, a_t is a time fixed effect, e_{nt}^{j} is the experience acquired by worker j in city n up until time t, $\mathbf{C}_t^{j}$ is a vector of time-varying individual and job characteristics, the scalar b_i and the vector $\mathbf{b}$ are parameters, and ε_{nt}^{j} is an error term.[67] They estimate that, controlling for worker and city characteristics, a year of experience in a city of more than 5 million residents (based on 2010 Census counts) yields 1.14% higher earnings than would a year in a city with between 2 and 5 million residents.[68] In my model, this corresponds to the moment condition

$$\left(\frac{X_{5\text{mil}}}{X_{2\text{mil}}}\right)^{1-\beta} = \left(\frac{L_{5\text{mil}}^{\frac{1}{\zeta}}\bar{x}_{5\text{mil},\zeta}}{L_{2\text{mil}}^{\frac{1}{\zeta}}\bar{x}_{2\text{mil},\zeta}}\right)^{1-\beta} = 1.0114,$$

where $X_{5\text{mil}}$ is the city size-weighted mean vibrancy among cities with more than 5 million residents and $X_{2\text{mil}}$ is that among cities with between 2 and 5 million residents. Given a value for β, we can think of this moment as primarily identifying the shape parameter of vibrancy, ζ.

In the second step, Duranton and Puga (2023) regress the estimated city fixed effects on city population to obtain a value for the static agglomeration elasticity, $\tilde{\alpha}$:

$$\hat{a}_n = \tilde{\alpha}\ln L_n + \varepsilon_n.$$

Because the city fixed effect captures local fundamentals that may be correlated with city size, they instrument for city size with a number of geographic characteristics and historical population values.[69] Their estimate is $\tilde{\alpha} = 0.0452$. In my model, estimating the city fixed effect a_n alongside individual and time fixed effects gives it the structural interpretation $a_n = \ln(T_n L_n^{\alpha})$, which is the log "wage" for human capital in city n. Although the fundamental productivity T_n may be correlated with city size L_n, instrumenting for city size as Duranton and Puga (2023) do suffices to identify α. It follows that α can actually be identified outside the minimum distance scheme as $\alpha = \tilde{\alpha} = 0.0452$.

In the final step, Duranton and Puga (2023) re-estimate the second regression after adding to the left-hand side the differential value of local experience in each city, valued at the average local experience, $\bar{e}$:

$$\hat{a}_n + \hat{b}_n\bar{e} = (\tilde{\alpha} + \tilde{\varsigma})\ln L_n + \varepsilon_n.$$

[67] Worker controls include firm tenure and its square along with indicators for two-digit sector and occupation. To ensure enough observations per city, Duranton and Puga (2023) group cities by size. Above 2 million residents, each city is in a group by itself, as I have done. Below 2 million residents, their groups are strictly finer than mine. Specifically, between 800,000 and 2 million, groups are incremented by 100,000. Between 600,000 and 800,000, they are incremented by 50,000. Finally, between 75,000 and 600,000, they are incremented by 25,000. See De la Roca, Ottaviano, and Puga (2023) for further description of the sample restrictions.

[68] See Duranton and Puga (2023, Table 2, Column 3).

[69] Specifically, they use (i) the percentage of the area within 30-kilometers of the city center that has slopes greater than 15% and the percentage covered by wetlands, (ii) the inverse hyperbolic sine of city population in 1850 and 1920, (iii) the inverse hyperbolic sine of distance to the Eastern Seaboard, and (iv) the number of heating degree days per year.

This gives an estimate of the dynamic agglomeration elasticity, $\tilde{\alpha}+\tilde{\varsigma}$. For average experience $\bar{e} = 8.4$, they estimate $\tilde{\alpha} + \tilde{\varsigma} = 0.0770$, so $\tilde{\varsigma} = 0.0318$. We can use this estimate to identify the scale of human capital, x_{scale}, because the static component of the city size earnings premium, $\hat{a}_n$, is independent of this scale but the dynamic component, $\hat{b}_n$, is not.

Identification of the remaining scale parameters goes roughly as follows. For a given wage bill, variation in employment identifies local amenities, B_n, which are restricted to have a mean of one. Similarly, for given employment, variation in the wage bill identifies local exogenous productivity, T_n. For a given local productivity, T_n, I set the scale of local urban costs, p_n, such that the expenditure share on local accommodations is constant across cities at 62% of potential income (Diamond, 2016).[70] Finally, the scale of the returns to time invested in human capital, $\bar{\kappa}$, is identified through Equation (16). This parameter can be separately identified from the scale of human capital, x_{scale}, because Equation (16) is invariant to this scale.

3.3 Quantitative predictions

The quantified model matches its aggregate targets exactly: city employment, city value added from labor, and aggregate growth are all as in the data. In particular, city employment, which is my measure of city size, follows a Pareto distribution and approximates Zipf's law, as demonstrated in Figure 1. The New York metropolitan area is the largest with more than 9 million workers in 2013, followed by Los Angeles with just under 6 million and Chicago with roughly 4.5 million. The four groups of small cities have mean employment of, roughly, 900,000; 600,000; 325,000; and 90,000, respectively.

Table 1 collects the parameter values that do not vary by location. The next section will discuss the city fundamentals.

Table 1: Quantification Results

Param.	Description	Value	Source
ϵ	Migration elasticity	3	Diamond (2016)
$\lambda(a)$	Mobility rates by age	$[0.076, 0.266]$	2011–15 ACS Migration Flow files
ρ	Discount rate	0.10	
A	Maximum age	40	Huggett, Ventura, and Yaron (2006)
β	Investment elasticity of own skill	0.8	Huggett, Ventura, and Yaron (2006)
η	Investment elasticity of time	0.7	Browning, Hansen, and Heckman (1999)
α	Static agglomeration elasticity	0.045	Duranton and Puga (2023)
ζ	Shape of vibrancy	6.228	MDE
x_{scale}	Scale of human capital	49.127	MDE
$\bar{\kappa}$	Scale of investment technology	0.139	MDE

City fundamentals and vibrancies

Figures 2–5 show the city fundamentals and vibrancies implied by my quantitative procedure. On each map, each marker refers to an MSA in that location. The size of each marker is proportional to total equilibrium employment in that city. The color of each marker varies

[70] In particular, $p_n = (0.62)T_n L_n^{\alpha-\theta_n}$. That expenditure on housing (and non-tradables more broadly) varies little across cities is consistent with evidence from Davis and Ortalo-Magné (2011), Moretti (2013b), and Diamond (2016). Along the baseline BGP, average time spent working exceeds 95%, so we lose little by setting expenditure shares relative to potential income instead of actual income.

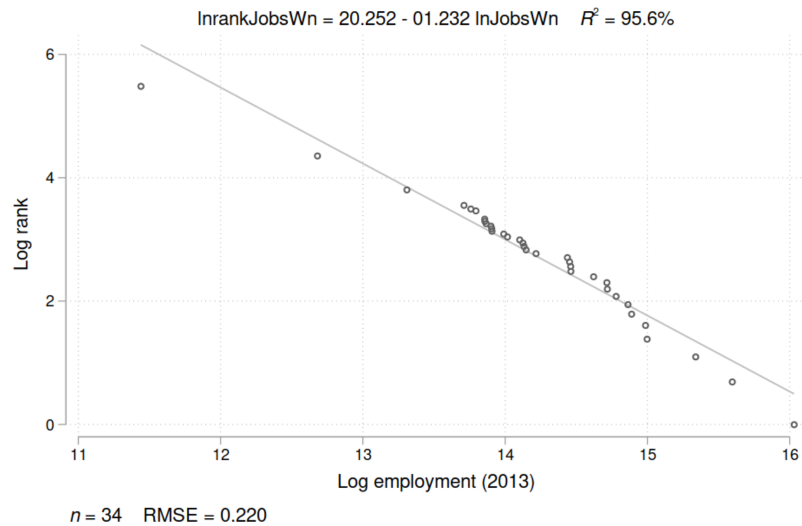

Figure 1: Log rank plotted against log city employment in 2013 for the 34 city groups in the sample. For groups with multiple cities, the rank shown is the median rank in the full distribution of its constituent cities.

with whatever city value is being shown, with higher values always represented by darker colors. Cities assigned to the same group are forced to be identical, so their markers will be of the same size and color by definition. Next to each map, a scatterplot shows the values for the thirty largest cities against their log size.

Figure 2 shows the local fundamental productivity, T_n, in each city. Unsurprisingly, fundamental productivity is positively correlated with size, because productive places attract workers. That said, cities like San Francisco, Washington, and Boston stand out relative to their size. Variation in T_n captures variation in local institutions, infrastructure, and capital stocks in general, but it is also how the model rationalizes pre-existing clusters like Silicon Valley or Boston's Kendall Square and Route 128.

Figure 3 shows the scale of urban costs, $p_n L_n^{\theta_n}$, in each city. Consistent with well-known patterns in the data, big cities are more costly, accounting for both higher rents and higher commuting costs.[71] Some of the most fundamentally productive cities, like San Francisco, are also the most expensive for two reasons. First, many of the most productive U.S. cities are also the most restrictive for urban land use, which translates into a high congestion elasticity, θ_n. Second, in order to match a constant budget share of urban costs across cities, p_n scales linearly with T_n in the quantitative procedure. This is a natural relationship, though, insofar as urban costs partly capture the opportunity costs of commuting, which vary with the average wage in the city.

Figure 4 shows the local amenity, B_n, in each city. Among the largest cities, warm-weather sites like Los Angeles, Miami, and Dallas stand out as having high amenities. Weather is clearly not the lone determinant, though: average amenities in smaller cities are substantially lower than those in big cities, regardless of climate. This suggests that

[71]See, among others, Combes, Duranton, and Gobillon (2019) and the review by Duranton and Puga (2020).

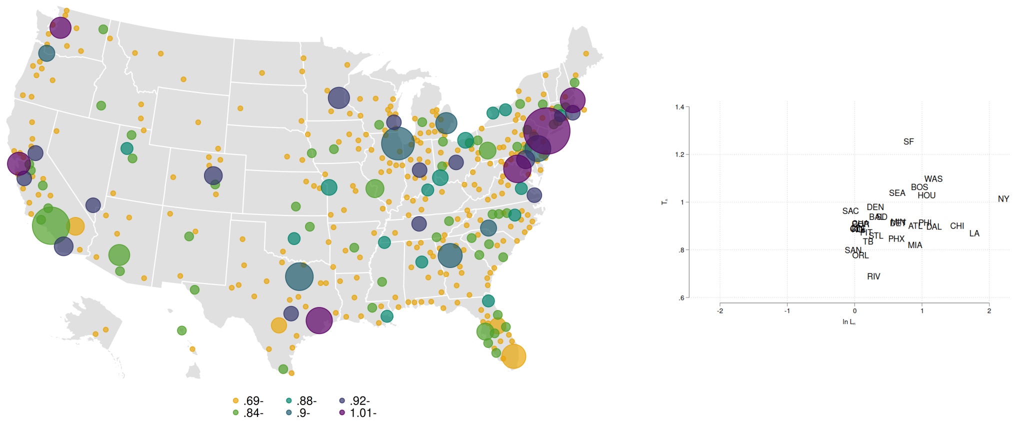

Figure 2: The map in the left panel shows the local fundamental productivity, T_n, of each of the 378 cities represented in the model. The scatterplot in the right panel shows, for the largest thirty cities, the level of each city's fundamental productivity against the natural log of its size.

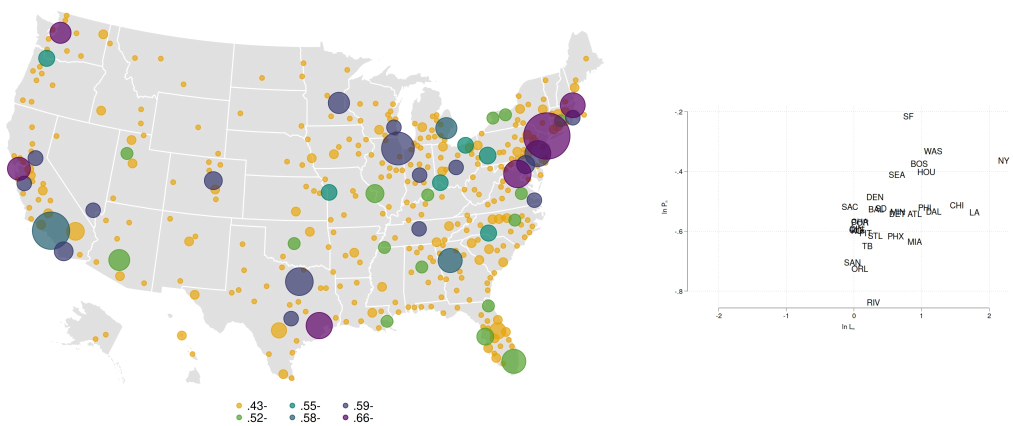

Figure 3: The map in the left panel shows the local scale of urban costs, $P_n = p_n L_n^{\theta_n}$, of each of the 378 cities represented in the model. The scatterplot in the right panel shows, for the largest thirty cities, the natural log of each city's urban cost against the natural log of its size.

something like consumption amenities—say, the quality or quantity of restaurants, entertainment, community spaces, and so on, which are typically higher in bigger cities—are also being captured by B_n. Consistent with this hypothesis, I find that, excepting San Francisco, my measure of amenities is positively correlated with the total amenity value estimated by Albouy (2016), who infers this value from U.S. data on wages, housing rents, and taxes through the lens of a spatial equilibrium model.[72] While Albouy (2016) infers that San Francisco has the highest total amenity value among U.S. cities, though, my measure ranks it poorly relative to other large cities. This is likely due to my method of inferring urban costs, which enforces a constant expenditure share on local accommodations across all workers and cities. To rationalize San Francisco's relatively small size given its high fundamental productivity, then, my procedure infers low amenities there.

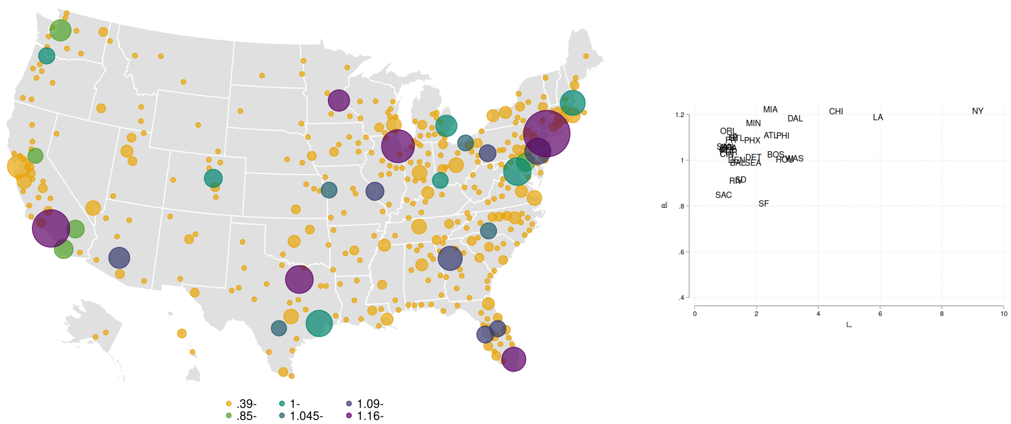

Figure 4: The map in the left panel shows the local amenity, B_n, of each of the 378 cities represented in the model. The scatterplot in the right panel shows, for the largest thirty cities, the level of each city's amenity against the natural log of its size.

Finally, Figure 5 shows the vibrancy, $X_n = L_n^{\frac{1}{\zeta}} \bar{x}_{n,\zeta}$, in each city. Given the parameter values in Table 1, the variation in city size dominates that in skill composition, so vibrancy is strongly positively correlated with city size. Nevertheless, differences in skill composition are still important: mean human capital in New York exceeds that in Detroit, for example, by nearly 9%. One city—Riverside, CA—sticks out as especially vibrant for its size. This is so despite a relatively low mean initial level of human capital per its observed college share and low fundamental productivity per its observed mean wage. As I explain in the next section, Riverside's high vibrancy is the result of a low opportunity cost of learning coupled with migration frictions.

[72]For groups with multiple cities, I use the population-weighted mean of his total amenity value. Note that Albouy (2016) defines cities based on the 1999 OMB definitions of metropolitan statistical areas, so there is some unavoidable mismatch between our results. Excepting San Francisco, the raw correlation between our amenity values across my 34 city groups is 22%.

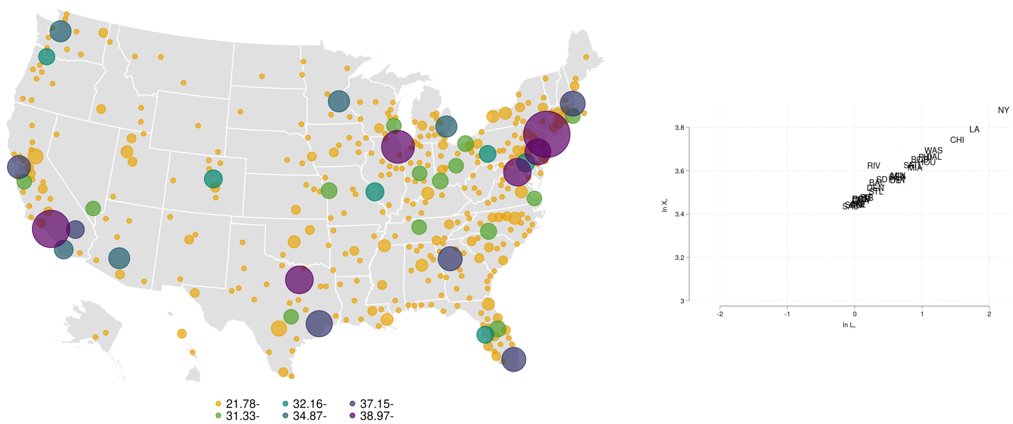

Figure 5: The map in the left panel shows the local vibrancy, $X_n = L_n^{\frac{1}{\zeta}} \bar{x}_{n,\zeta}$, of each of the 378 cities represented in the model. The scatterplot in the right panel shows, for the largest thirty cities, the natural log of each city's vibrancy against the natural log of its size.

Worker decision rules

In Figures 6–8, I turn to workers' decision rules. Each figure comprises four panels, one for each of four selected ages throughout the life cycle. Within each panel, I plot the relevant function in each city across the range of relative human capital levels that lie within the support of the stationary distribution. To limit visual clutter, I highlight two cities—New York (blue) and Detroit (orange)—and show the rest in gray.

Figure 6 shows the optimal investment rate, $\sigma_n(a, x)$. As demonstrated in Proposition 1, the optimal investment rate is non-increasing in both age and relative human capital within a given city. This is consistent with the observed concentration at young ages of time allocated to skill acquisition (Mincer, 1974, 1997). Across cities, however, the investment rate varies ambiguously: larger cities generally feature a higher opportunity cost of investment because they afford higher flow utility per unit of human capital, $B_n T_n L_n^{\alpha}$, but they also feature higher returns to investment because they have higher vibrancy, X_n. Relative to those in most cities, investment rates in New York are high, whereas those in Detroit are middling. Investment rates are especially high in one particular city: Riverside. This is because both its amenity, B_{RIV}, and its fundamental productivity, T_{RIV}, are quite low for its size, but its vibrancy is, if anything, high for its size. As such, the model predicts that workers in Riverside would invest relatively heavily in their human capital so long as they are stuck in a low-amenity, low-productivity environment.

Figure 7 shows the optimal accumulation rate, $h_n(a, x) = \frac{\bar{\kappa}}{\eta}[\sigma_n(a, x)]^{\eta} x^{\beta} X_n^{1-\beta}$. Although a worker's optimal investment rate decreases with her relative human capital, her optimal accumulation rate does not necessarily decrease because higher-skilled workers are also better learners. Because $\eta < \beta$, her optimal accumulation rate is in fact increasing with her relative human capital. Nevertheless, her accumulation declines with age as her investment rate declines. Across cities, workers generally accumulate human capital more quickly in more vibrant places, consistent with the moments I matched from Duranton and Puga (2023, Table 2). Riverside again stands apart because of its especially high investment rates for its size. Otherwise, cities are ranked roughly in order of their size, with New York

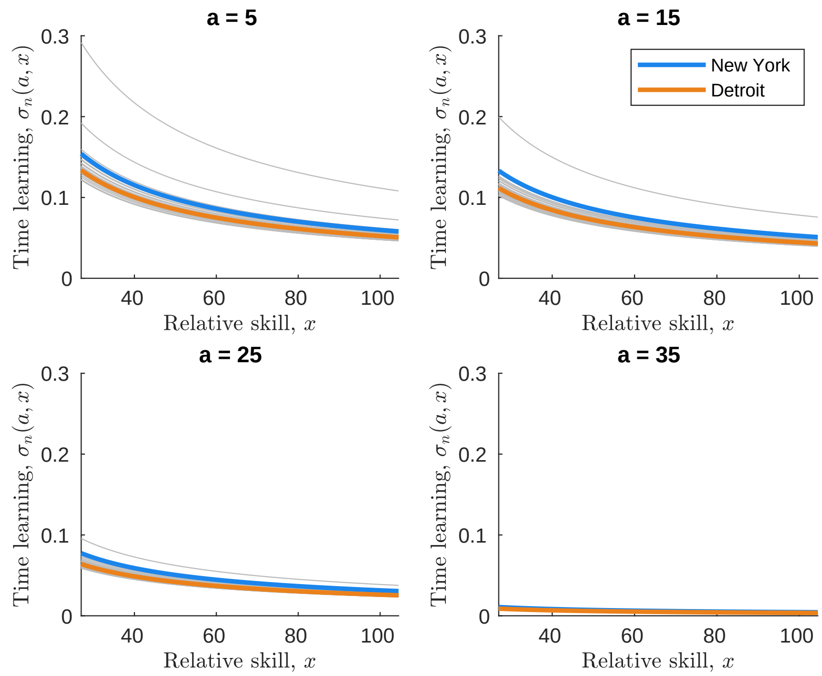

Figure 6: Each of the four panels plots the optimal time allocated to learning across relative productivity levels for a different age, identified in the title. There are 34 lines in each panel, one for each city group, with two cities highlighted: New York (blue) and Detroit (orange). All other cities are plotted in gray to save on visual clutter.

at the top.

Figure 8 shows the optimal drift of relative human capital, $d_n(a,x) = h_n(a,x) - \gamma x$. Although a worker's absolute level of human capital, z, does not depreciate, her relative level, x, does depreciate as the economy grows at rate γ. Her relative level only increases on net if her accumulation outpaces this shift in the distribution. The young typically invest heavily enough in their human capital to gain relative to the trend, as evidenced by the lines in the top two panels lying largely above the dashed line at zero. Such is not the case for the old, who let their relative level depreciate even still five years before retirement, as seen in the last panel. Regardless of age, accumulation relative to the trend is decreasing. Putting all of this together, the lines in Figure 8 allow us to visualize the components of the growth rate in Equation (16). For each relative level x within the support of the density ϕ, the growth rate is the number γ that makes the average drift $d_n(a,x)$ across all ages and cities—that is, across all panels and all lines within a panel—equal to zero.

Moreover, notice that at any age a in any city n, the value $\tilde{x}_{a,n}$ at which $d_n(a,\tilde{x}_{a,n})$=0 is a *sink*: a worker will continue to accumulate on net up to $\tilde{x}_{a,n}$, but past that level she would let her relative human capital depreciate. This suggests that the detrended stationary distribution has finite support, as we will confirm below. Contrast this result with other idea-based growth models that require a distribution with a fat tail or an external source

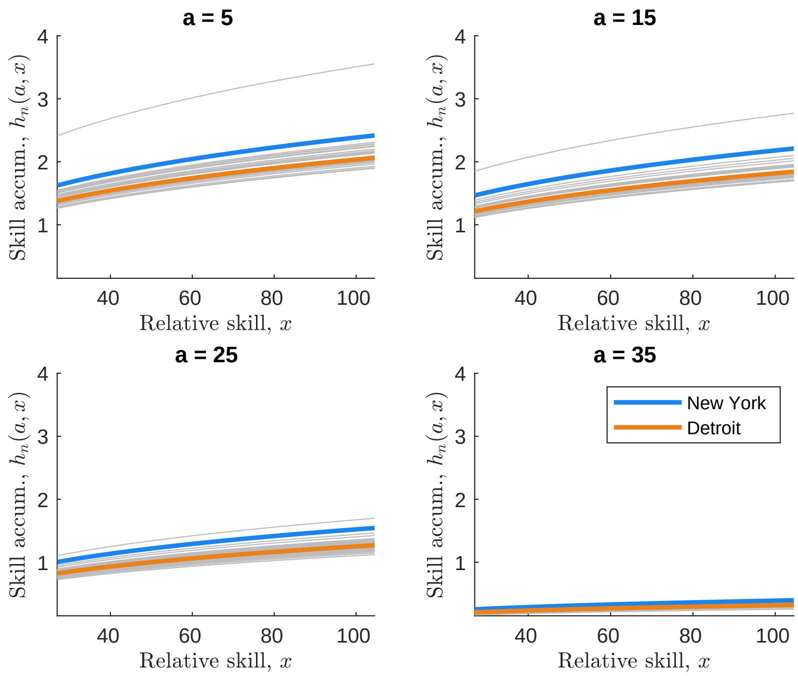

Figure 7: Each of the four panels plots the optimal accumulation of human capital, $h_n(a,x) = \kappa[\sigma_n(a,x)]x^\beta X_n^{1-\beta}$, across relative productivity levels for a different age, identified in the title. There are 34 lines in each panel, one for each city group, with two cities highlighted: New York (blue) and Detroit (orange). All other cities are plotted in gray to save on visual clutter.

to sustain growth.[73] The key difference here is that human capital accumulation occurs by producing potentially new knowledge, not simply by imitating others' knowledge, so steady-state growth can occur despite there being only finite knowledge in the economy at any particular date.

Figure 9 shows the optimal annual flows from each origin to each destination, $\lambda(a)L_n(a)\mu_{ni}(a,x)$, at four select ages, $a \in \{5, 15, 25, 35\}$. Rows represent the origin of a flow, with origin city groups ranked from top to bottom by the size of their constituent cities. Columns represent the destination of a flow, with destination city groups ranked from left to right by the size of their constituent cities. With any given year, workers are far more likely to stay in their current city than move away, so I omit the flows along the diagonal from the color scale to highlight the variation across destinations. Each panel has the same color scale, so looking across panels (first left to right, then top to bottom) we observe from the darkening palette that the young are more likely to migrate. This is due largely to the decline in frequency of migration opportunities, $\lambda(a)$, as workers age, but also in part to the decline in the value function, $v_i(a,x)$, relative to the size of the bilateral migration frictions, τ_{ni}, as workers age. With any given panel, flows below the diagonal run from smaller to bigger cities. Young workers (first panel) move on net to larger cities, with total flows below the diagonal exceeding those above the diagonal, albeit only by 2%. A particularly strong channel runs from

[73]See, for example, Lucas and Moll (2014) and Perla and Tonetti (2014).

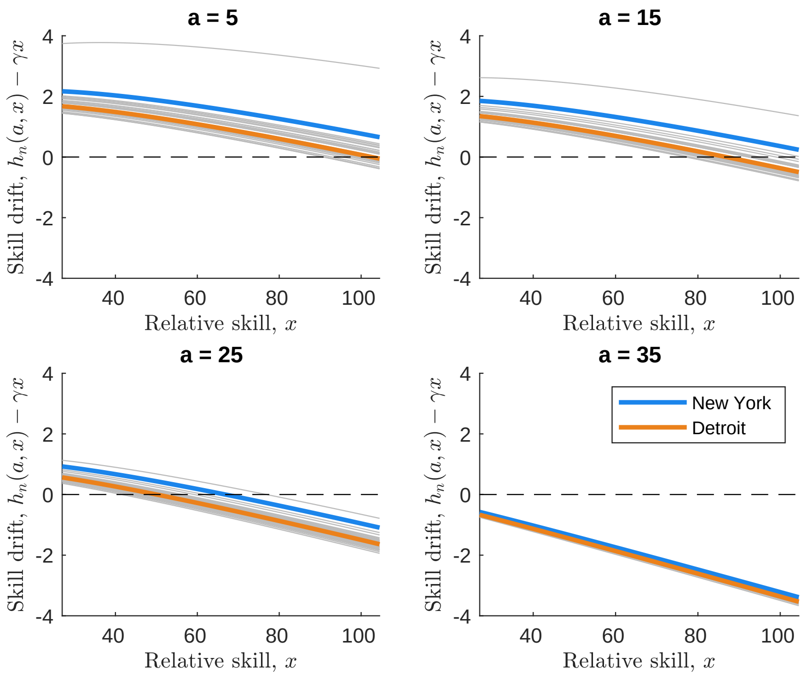

Figure 8: Each of the four panels plots the optimal drift of human capital, $d_n(a,x) = \kappa[\sigma_n(a,x)]x^{\beta}X_n^{1-\beta} - \gamma x$, across relative productivity levels for a different age, identified in the title. There are 34 lines in each panel, one for each city group, with two cities highlighted: New York (blue) and Detroit (orange). All other cities are plotted in gray to save on visual clutter.

Riverside to its larger neighbor, Los Angeles, but also to its smaller neighbor, San Diego, both of which have better fundamentals than Riverside.

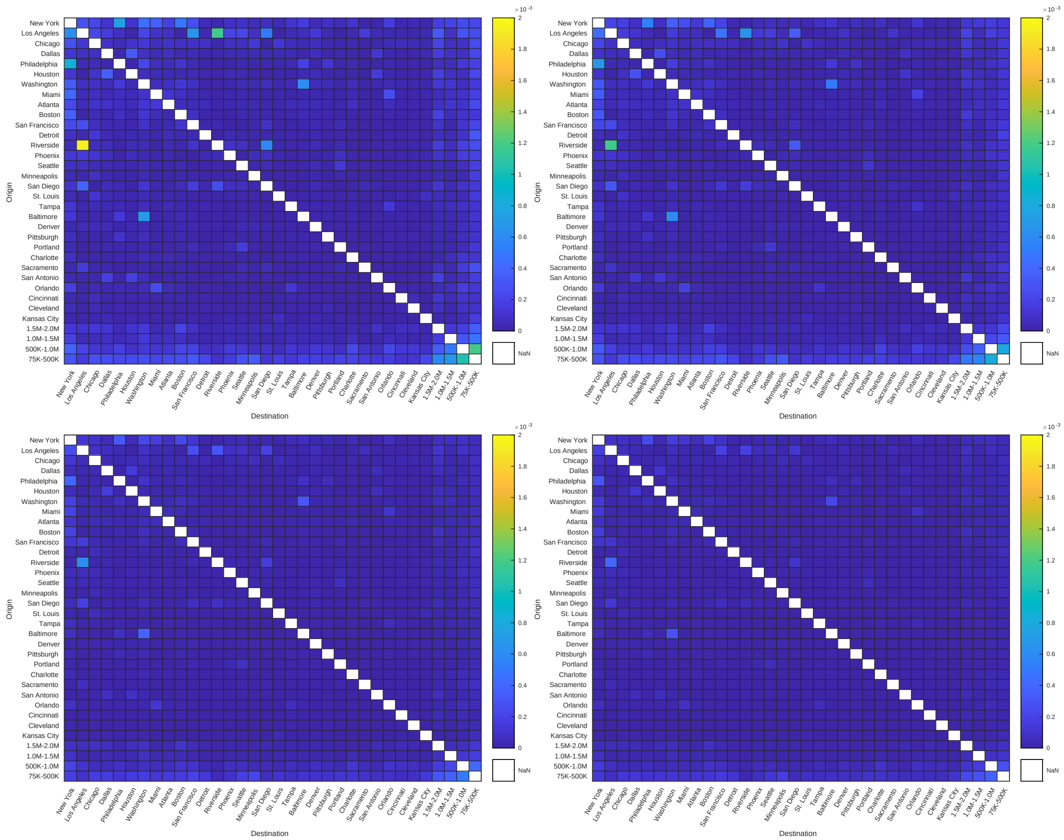

Figure 9: Each of the four panels contains a heatmap of annual flows (in millions of workers) across all bilateral pairs of the 34 city groups for a different age (in order left to right, then top to bottom: $a = 5$, $a = 15$, $a = 25$, $a = 35$). Rows represent the origin of a flow; columns, the destination. Own flows are omitted from the color scale because they would mask all other variation. Cities are ordered by size from top to bottom and left to right. Migration rates are *not* cross-tabulated with relative productivity in this figure.

City skill and age distributions

Each city's human capital density inherits its initial shape and mass from its distribution of entrants, $\underline{\phi}_n$, which is lognormal with a city-specific mean and mass as described in Section 3.2. In solving for the density as a cohort ages, however, recall that I proceed non-parametrically, simply following the evolution implied by the Kolmogorov forward equation. Nor do I target any moments of the distribution except the total mass of workers in each city.

Nevertheless, the resulting distributions are quite reasonable. Figure 10 shows the density of human capital within each city. Looking across panels, the density in each city first shifts to the right as a cohort ages and accumulates human capital. Near the end of the working life, the right tail recedes as workers let their relative human capital depreciate. These patterns are consistent with the concavity of the cross-sectional earnings distribution targeted in the human capital literature (Mincer, 1974).

Within any given panel, the relative placement of the lines indicates which cities are more or less skilled. At any age, for example, New York is not only larger than Detroit, because the blue line lies above the orange one, but it is also more skilled, because relatively more of its mass is in the upper range of the supported skill levels. Excluding Riverside, which has an artificially high mean level of relative skill as we have discussed above, the overall most skilled cities are, in order, New York, Los Angeles, Washington, Chicago, Boston, and San Francisco. The least skilled are Sacramento, Cincinnati, Cleveland, Tampa, and San Antonio.

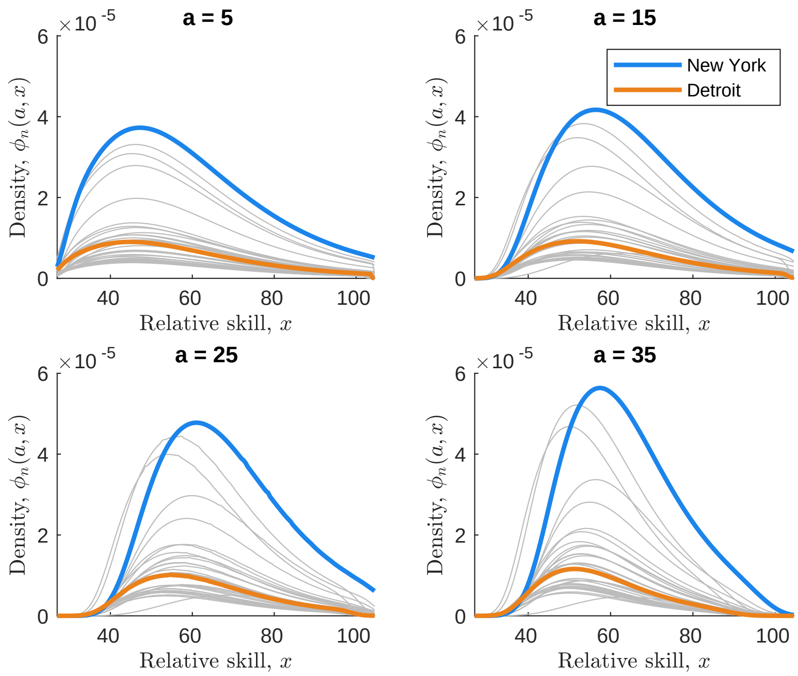

Figure 10: Each of the four panels plots the density $\phi_n(a,x)$ across relative productivity levels for a different age, identified in the title. There are 34 lines in each panel, one for each city group, with two cities highlighted: New York (blue) and Detroit (orange). All other cities are plotted in gray to save on visual clutter.

Figure 11 shows the marginal age density in each city normalized by the city's total population share, $\phi_n(a) \equiv \frac{1}{\ell_n} \int \phi_n(a,x)\,dx$. Normalizing by population share allows us to compare age demographics across cities on a common scale. A key distinction is between cities for which the marginal age density slopes up and those for which it slopes down. In an environment without migration frictions, an upward-sloping density here would imply that a particular city is disproportionately preferred by the old; a downward-sloping one, preferred by the young. But with migration frictions, the proper intuition is instead that the places most preferred by the young are those with the steepest upward slopes at young ages. These are the cities to which young people flock at their earliest migration opportunities. The marginal density of New York, for instance, is upward-sloping. The seven steepest lines up to $a = 10$ (that is, those starting below 0.02) are Charlotte, Washington, San Francisco, Minneapolis, Seattle, Denver, and Boston. These cities are relatively productive and skilled even controlling for size. Cities with the steepest declines up to $a = 10$, by contrast, are

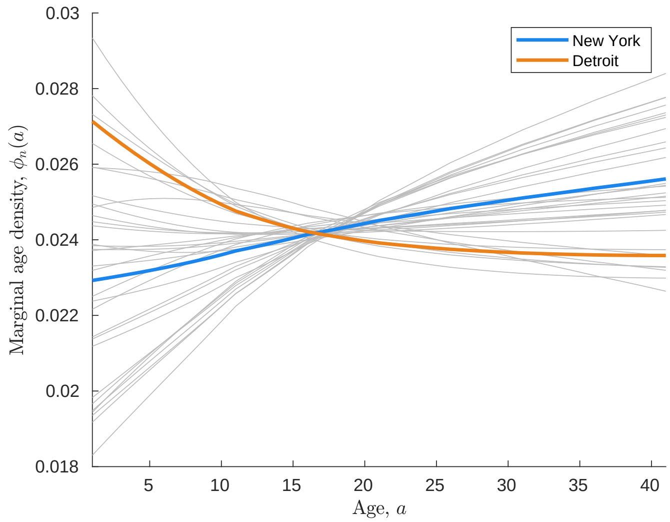

Figure 11: This figures plots the marginal density of age within each city, normalized by the city's total population share. There are 33 lines in the figure, one for each city group except Riverside, with two cities highlighted: New York (blue) and Detroit (orange). All other cities are plotted in gray to save on visual clutter. Riverside is omitted because it decreases convexly from 0.0458 to 0.0141, which means it would have dominated the variation across the other lines if shown on the same scale.

Riverside, Sacramento, San Antonio, Los Angeles, and Detroit. Two of these five cities, Sacramento and San Antonio, are among the five least-skilled cities in the economy, and Detroit and Riverside have poor fundamentals relative to other cities of similar sizes. These are the places that young people leave at their earliest opportunities. Places that are desirable to the young are also generally desirable to the old, which is why nearly all of the lines in Figure 11 are monotonic. Still, the slope for every city declines with age, consistent with the declining mobility of workers as they approach retirement.

An example individual history

Following De la Roca and Puga (2017), a final helpful lens through which to view the predictions of the model is to plot the evolution of earnings for workers in cities of different size. Consider a worker born with mean initial human capital in Detroit that will receive an opportunity shock immediately upon entry at age 20 ($a = 0$) and another after 10 years of experience ($a = 10$).[74] At each opportunity, she has a strictly positive probability of moving to any city in the set $\mathcal{N}$, but let us focus on two options: Detroit, with between 2 and 5 million residents, and New York, with more than 5 million. At her first opportunity, the worker can either stay in Detroit or move to New York. At her second opportunity, if she had previously moved to New York, she can remain there or return to Detroit; if she had stayed in Detroit, let us suppose that she chooses to stay there again.

Accordingly, we consider three potential paths for this worker, plotted in Figure 12. Along the first path, shown in solid orange, the worker is always in Detroit. Along the

[74] Of course, these shocks are unbeknownst to her, but with a continuum of agents we observers are assured to find some worker with this realization of shocks.

second, shown in solid blue, the worker immediately moves to New York and remains there. In the third, shown in solid blue up to age 30 ($a = 10$) and thereafter in dashed orange, she first moves to New York but later returns to Detroit. The left panel shows the evolution of her earnings premium relative to always working in Detroit for each of the three potential paths (hence why the solid orange line is at zero). The right panel shows the same for her underlying level of human capital.

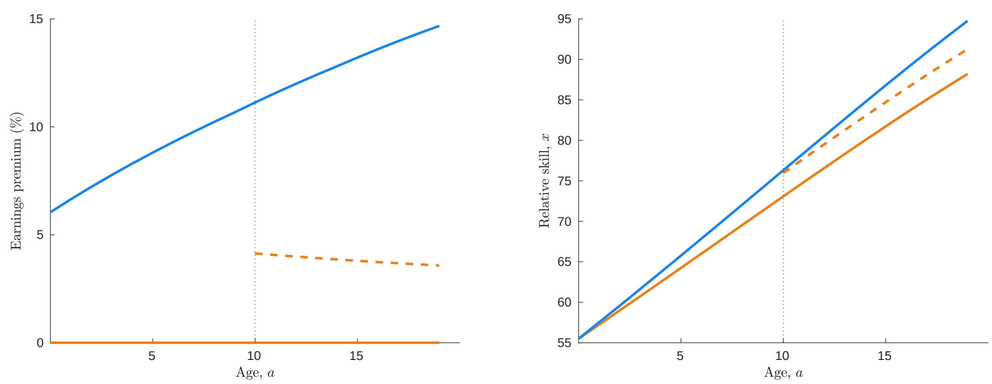

Figure 12: This figure illustrates three possible sample paths of a worker born with mean initial human capital in Detroit that receives opportunity shocks at ages 20 ($a = 0$) and 30 ($a = 10$). The left panel plots her earnings premium relative to always staying in Detroit against her age; the right panel, her relative productivity against her age. The solid orange line represents the path in which she chooses to stay in Detroit at both opportunities. The solid blue line represents her path after age 20 if she chooses to move to New York at her first opportunity and remain there. The dashed orange line represents her path after age 30 if, after moving to New York at age 20, she chooses to return to Detroit at her second migration opportunity.

For the version of the worker in New York, her profile of relative earnings has an intercept and a slope component. The intercept captures the percentage difference in earnings between the version of her working in New York and that of her working in Detroit. This is calculated as the percentage difference between her "wage" in New York, L^{α}_{NY}, and that in Detroit, L^{α}_{DET}, after controlling for the fundamental productivity in each location, since we are interested in the portion of the earnings premium that varies with city size. The slope component captures the rising gap in earnings between these versions of herself as they each accumulate experience in a different city. This is calculated by tracing out her accumulation in New York, $h_{\text{NY}}(a, x)$, and that in Detroit, $h_{\text{DET}}(a, x)$. Figure 12 shows that a worker in New York initially earns 6% more than a worker in Detroit, controlling for differences in fundamental productivity, and this gap nearly doubles, so that after 10 years the difference in earnings reaches almost 11%.

The dashed lines in Figure 12 illustrate the portability of the learning advantages of bigger cities. Up until year 10, the relative earnings profile of the version of our worker who begins in New York and then relocates is the same as that of the version who always works in New York as captured by the solid blue line discussed above. At that point, she relocates to Detroit, and her relative earnings drop as a result of the Detroit wage replacing the New York wage. Nevertheless, the worker is able to retain roughly 5% higher earnings after relocating to Detroit as a result of the more valuable experience she accumulated over 10 years in New York.

The right panel of Figure 12 demonstrates why this occurs. The same worker accumulates human capital at a faster rate in New York than she would in Detroit, so the version of herself that moves to New York has, after 10 years of experience, accumulated roughly 5% more human capital than the version of herself that stayed in Detroit. The final version of herself that moves back to Detroit after 10 years will therefore have a permanently higher level of human capital than the version that had always stayed in Detroit.

4 Policy exercise

As an application of the model, I now consider a commonly-proposed place-based policy: relaxing land use regulations in the "brain hubs" of New York and San Francisco. This exercise is subject to an important caveat: the effect of the policy is stated as a comparison *across* balanced growth paths without regard for the speed or cost of transition to the new path. In the presence of migration frictions, the transition would be interesting to study, but computing it falls outside the methods developed in this paper.

4.1 Relaxing land use regulations in the Brain Hubs

Recall that the congestion elasticity, θ_n, is set to match estimates from Saiz (2010) of the inverse housing supply elasticity in each city. This elasticity has two main determinants: land availability and land use regulations as measured by the Wharton Residential Land Use Regulatory Index (WRI).[75] According to the WRI, land use restrictions in New York and San Francisco are among the tightest in the country. The elasticity of housing supply due to land use regulations in New York is at the ninety-sixth percentile of the nationwide distribution of the housing supply elasticity; that in San Francisco, the ninety-ninth percentile.

In this policy exercise, I focus on a change in the congestion elasticity, θ_n, holding fixed the scale of urban costs, p_n, in both New York and San Francisco. Specifically, I set land use regulations in both cities equal to the level observed in the median US city. I then recompute the congestion elasticity in each city, holding fixed the amount of land available.[76] With relaxed land use regulations, both cities can accommodate more workers at the same level of urban costs.

Figures 13–15 show how city sizes, costs, and vibrancies change. Unsurprisingly, New York and San Francisco grow in size, with New York expanding by 16.9% and San Francisco by 4.9% as shown in Figure 13. All other cities shrink, but all do so by less than 3%. Most of the reallocation is away from large cities in the middle of the country, like Atlanta, Kansas City, and Dallas, or from cities close to New York, like Philadelphia.

One consequence of this spatial reallocation is that urban costs fall everywhere, as shown in Figure 14. The inflows to New York and San Francisco, though large, are small enough so as not to offset the direct reduction in costs from the policy. Meanwhile, all other cities experience reduced congestion as they shrink.

Another consequence is that the brain hubs become more vibrant, as shown in Figure 15. New York becomes 13.4% more vibrant; San Francisco, 2.9% more vibrant. Unlike with size, though, which decreases for all other cities in response to the policy, the vibrancy of some other large cities also increases. Philadelphia, Miami, and Boston each become at least 1% more vibrant. Additional gains cluster along the East Coast (Washington and

[75] The WRI is based on detailed surveys of municipalities in 2007. It is designed to have a zero mean and unit standard deviation. See Gyourko, Saiz, and Summers (2008) for further details.

[76] Hsieh and Moretti (2019) and Martellini (2022) consider similar counterfactuals in static or stationary settings.

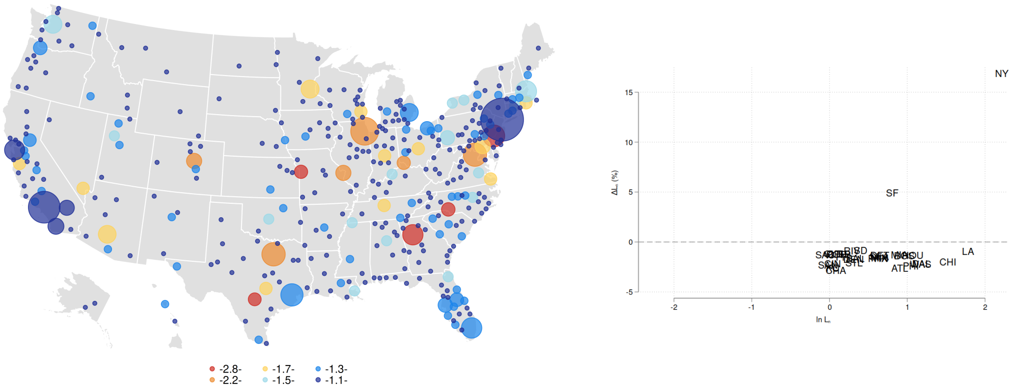

Figure 13: The left panel maps each city's percentage change in population relative to the competitive allocation. The right panel plots that percentage change against the natural log of each city's population in the competitive allocation.

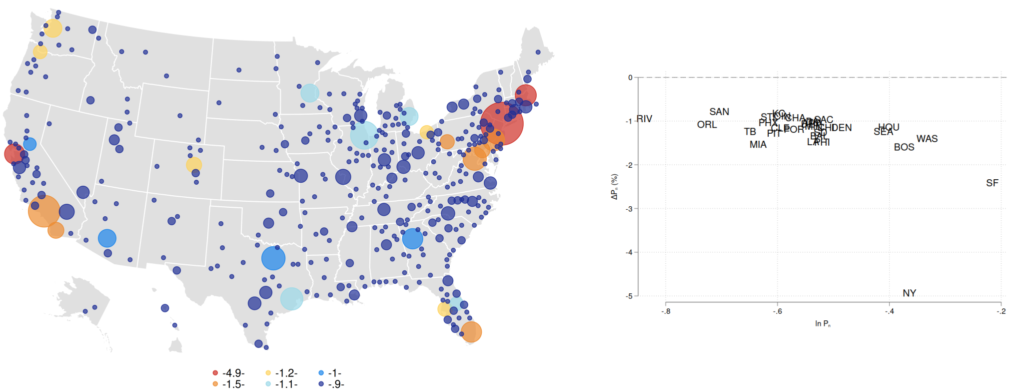

Figure 14: The left panel maps each city's percentage change in total urban cost relative to the competitive allocation. The right panel plots that percentage change against the natural log of each city's urban cost in the competitive allocation.

Baltimore), in Southern California (Los Angeles, Riverside, and San Diego), and in Central Florida (Tampa and Orlando). For these cities to shrink and yet become more vibrant, it must be that they are either attracting or producing more human capital. The same must be true even for the brain hubs: holding the skill distribution fixed within each city, one would expect a 1% increase in size to yield only a $\frac{1}{\zeta}\%$ increase in vibrancy.

We can confirm this by examining the underlying skill distributions in different cities. Figure 16 shows the skill distribution in three cities: New York, San Francisco, and Detroit. The solid lines show the three distributions along the baseline BGP. The dashed lines show the three distributions along the counterfactual BGP. As a result of the policy, we see a pronounced rightward shift in New York's distribution, especially at older ages. A much less pronounced rightward shift also occurs in San Francisco. Detroit, which was not targeted by the policy, sees essentially no change in its skill distribution. Changes in other untargeted cities range from the imperceptible to the slightly noticeable in Philadelphia, which experienced the largest deviation between its change in size (-2.42%) and change in

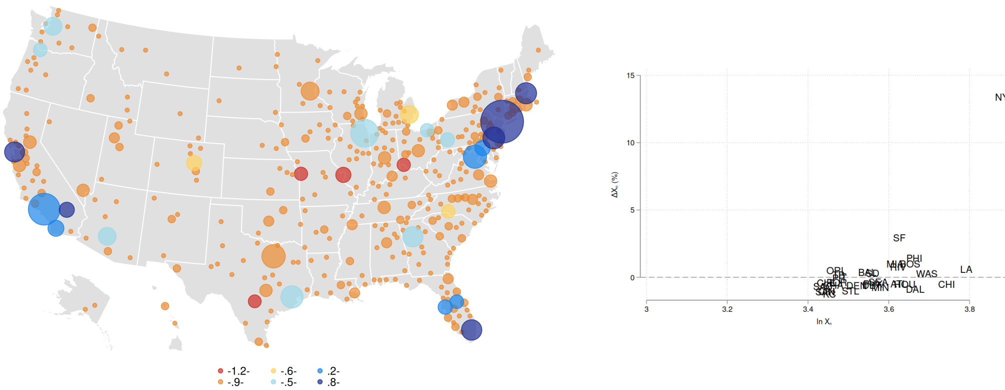

Figure 15: The left panel maps each city's percentage change in vibrancy relative to the competitive allocation. The right panel plots that percentage change against the natural log of each city's vibrancy in the competitive allocation.

vibrancy (+1.42%) of all untargeted cities.

The large shift in New York compared to the minimal shifts in untargeted cities suggests that the response to the policy is not primarily through skill-biased migration. A cheaper New York is not simply syphoning off high-skilled workers from other cities. Instead, workers in New York must be accumulating human capital at a faster rate. We see this in Figure 17, which shows the optimal accumulation, $h_n(a, x)$, in the same three cities as before. The dashed blue line lies well above the solid blue line, which means workers in New York earn much higher returns on their human capital investments after the policy. The same goes for workers in San Francisco, but to a lesser extent.

Despite the noticeable effects in New York, ultimately I find that relaxing land use regulations in the brain hubs increases the aggregate growth rate only by 13 basis points to $\gamma^{\text{BH}} = 2.13\%$. A few factors mitigate the aggregate effect. First, the policy only induces moderate employment reallocation. New York and San Francisco become cheaper and so attract more workers, but in general equilibrium this makes all other cities cheaper, too, as they shrink, thereby attenuating the pull of the brain hubs. Second, although human capital accumulation speeds up in the brain hubs, the diffusion of this new knowledge is slowed by migration frictions. Freer mobility would likely lead to larger gains. Nevertheless, this exercise demonstrates that simply relaxing land use regulations in two cities could have a permanent, positive effect on aggregate growth.

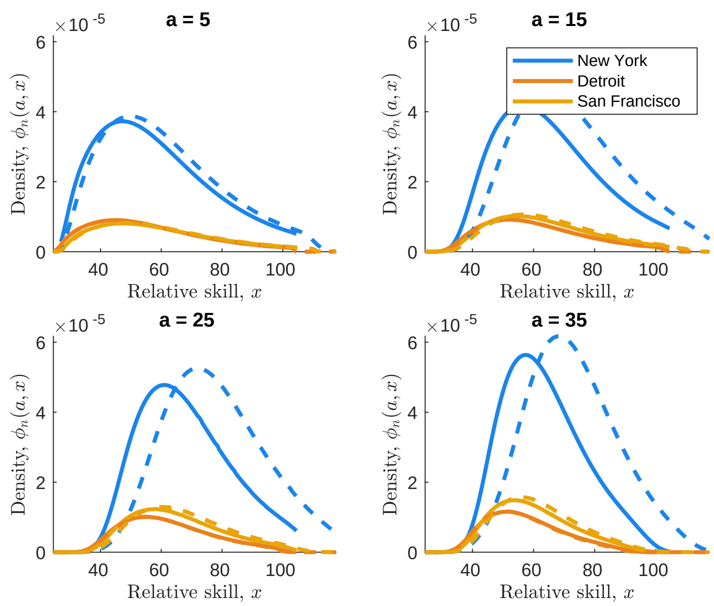

Figure 16: Each of the four panels plots the density $\phi_n(a, x)$ across relative productivity levels for a different age, identified in the title. There are 6 lines in each panel, two each for New York, San Francisco, and Detroit. Solid lines denote distributions on the baseline BGP. Dashed lines denote distributions on the counterfactual BGP.

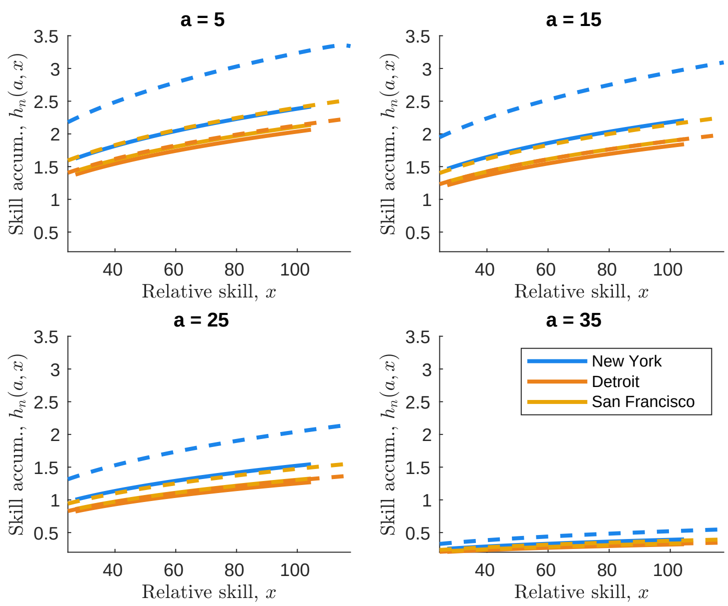

Figure 17: Each of the four panels plots the accumulation $h_n(a, x)$ across relative productivity levels for a different age, identified in the title. There are 6 lines in each panel, two each for New York, San Francisco, and Detroit. Solid lines denote accumulations on the baseline BGP. Dashed lines denote accumulations on the counterfactual BGP.

5 Conclusion

In this paper, I develop a quantitative spatial endogenous growth model in which heterogeneous workers make forward-looking migration decisions and human capital investments over the life cycle. Local externalities in human capital investment drive both agglomeration and growth. I show that the growth rate depends on the spatial distribution of human capital, making it sensitive to spatial policies.

I quantify the model such that it jointly rationalizes a host of empirical facts about American cities in the cross-section and American workers over the life cycle. City sizes approximately follow Zipf's law, and bigger cities are generally more productive, more expensive, and more skilled. Consistent with recent panel estimates, workers learn more in big cities, and the young migrate disproportionately to bigger and more-skilled cities because of this. All the while, the economy as a whole follows a balanced growth path with 2% annual growth.

Using the quantified model, I evaluate a commonly-proposed place-based policy: relaxing land use restrictions in the "brain hubs" of New York and San Francisco. I find that, along the new balanced growth path, both cities are larger and more skilled, with New York especially so. This is not primarily the result skill-biased migration, as if the brain hubs simply syphoned off skilled workers from other cities. Instead, the rate of human capital accumulation in both cities increases, meaning that the spatial policy helps *produce* more skilled workers. Ultimately, I find that the aggregate growth rate for the entire economy increases by 13 basis points in response to the policy.

I expect my framework to serve as a basis for future work. My framework is well-suited to evaluate the long-run effects of policies that shift workers around the urban system, like the \$11 billion regional technology hub program funded through the American Rescue Plan and the CHIPS and Science Act, which aims to create "regional growth clusters" outside of the existing coastal hubs (EDA, 2021; CHIPS, 2022).[77] It could also be used to estimate the long-run effects of remote work, to reexamine the Great Divergence (Moretti, 2013a), or to trace out the effects of place-based policies for macro-development (Duranton and Venables, 2021).

It would be particularly useful to extend my framework along several margins. First, I abstracted from the consumption-savings problem because Ben-Porath workers maximize the present value of earnings in the absence of a labor-leisure decision and liquidity constraints. When workers face liquidity constraints, though, Bilal and Rossi-Hansberg (2021) have argued that workers may treat their location as an asset by moving to cheaper places with lower future returns as a form of borrowing. Including a consumption-savings problem with liquidity constraints alongside the migration and investment problems would allow for this effect. Second, I defined vibrancy as a single index of knowledge at the city level to which resident workers have continuous access, but one may consider how occupational propinquity or urban form affect the strength of knowledge spillovers or the frequency of face-to-face interactions (Atkin, Chen, and Popov, 2022). Finally, I restricted my solutions to balanced growth paths, but it would be useful to solve for the transition between balanced growth paths in order to compute the speed and welfare effects of the transition in the presence of migration frictions.

[77] Atkinson, Muro, and Whiton (2019) and Gruber and Johnson (2019) advocate for this type of policy in response to recent trends in the concentration of U.S. innovation-sector jobs.

References

2021. “FY 2021 American Rescue Plan Act Build Back Better Regional Challenge Notice of Funding Opportunity.” Economic Development Administration, Tech. Rep. EDA-HDQ-ARPBBB-2021-2006976.

2022. “FACT SHEET: CHIPS and Science Act Will Lower Costs, Create Jobs, Strengthen Supply Chains, and Counter China.” The White House, Tech. rep.

Achdou, Yves, Pierre Cardaliaguet, François Delarue, Alessio Porretta, and Filippo Santambrogio. 2020. *Mean field games*. No. 2281 in CIME Foundation Subseries. Springer International Publishing.

Achdou, Yves, Jiequn Han, Jean-Michel Lasry, Pierre-Louis Lions, and Benjamin Moll. 2022. “Income and wealth distribution in macroeconomics: A continuous-time approach.” *Review of Economic Studies* 89 (1):45–86.

Ahlfeldt, Gabriel M., Stephen J. Redding, Daniel M. Sturm, and Nikolaus Wolf. 2015. “The economics of density: Evidence from the Berlin Wall.” *Econometrica* 83 (6):2127–2189.

Albouy, David. 2016. “What are cities worth? Land rents, local productivity, and the total value of amenities.” *Review of Economics and Statistics* 98 (3):477–487.

Alvarez, Fernando E., Francisco J. Buera, and Robert E. Lucas, Jr. 2014. “Idea flows, economic growth, and trade.”

Artuç, Erhan, Shubham Chaudhuri, and John McLaren. 2010. “Trade shocks and labor adjustment: A structural empirical approach.” *American Economic Review* 100 (3):1008–45.

Atkin, David, Keith Chen, and Anton Popov. 2022. “The returns to face-to-face interactions: Knowledge spillovers in Silicon Valley.” National Bureau of Economic Research, Working Paper 30147.

Atkinson, Robert D., Mark Muro, and Jacob Whiton. 2019. “The case for growth centers: How to spread tech innovation across America.” Brookings Institution, Research report.

Balboni, Clare. 2019. “In harm’s way? Infrastructure investments and the persistence of coastal cities.”

Baum-Snow, Nathaniel and Ronni Pavan. 2012. “Understanding the city size wage gap.” *Review of Economic Studies* 79 (1):88–127.

Behrens, Kristian, Gilles Duranton, and Frédéric Robert-Nicoud. 2013. “Household sorting in the city.”

Behrens, Kristian and Frédéric Robert-Nicoud. 2015. “Agglomeration theory with heterogeneous agents.” In *Handbook of Regional and Urban Economics*, vol. 5A, edited by Gilles Duranton, John Vernon Henderson, and William C. Strange, chap. 4. Elsevier, 171–245.

Ben-Porath, Yoram. 1967. “The production of human capital and the life cycle of earnings.” *Journal of Political Economy* 75 (4):352–365.

Berliant, Marcus and Masahisa Fujita. 2008. "Knowledge creation as a square dance on the Hilbert cube." *International Economic Review* 49 (4):1251–1295.

Berliant, Marcus, Robert R. Reed, and Ping Wang. 2006. "Knowledge exchange, matching, and agglomeration." *Journal of Urban Economics* 60 (1):69–95.

Berry, Christopher R. and Edward L. Glaeser. 2005. "The divergence of human capital levels across cities." *Papers in Regional Science* 84 (3):407–444.

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

Bilal, Adrien and Esteban Rossi-Hansberg. 2021. "Location as an asset." *Econometrica* 89 (5):2459–2495.

Black, Duncan and Vernon Henderson. 1999. "A theory of urban growth." *Journal of Political Economy* 107 (2):252–284.

Breinlich, Holger, Gianmarco I. P. Ottaviano, and Jonathan R. W. Temple. 2014. "Regional growth and regional decline." In *Handbook of Economic Growth*, vol. 2, edited by Philippe Aghion and Steven N. Durlauf, chap. 4. Elsevier, 683–779.

Browning, Martin, Lars Peter Hansen, and James J. Heckman. 1999. "Micro data and general equilibrium models." In *Handbook of Macroeconomics*, vol. 1, chap. 8. Elsevier, 543–633.

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

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

Cai, Sheng, Lorenzo Caliendo, Fernando Parro, and Wei Xiang. 2023. "Mechanics of spatial growth." National Bureau of Economic Research, Working Paper 30579.

Caicedo, Santiago, Robert E. Lucas, Jr, and Esteban Rossi-Hansberg. 2019. "Learning, career paths, and the distribution of wages." *American Economic Journal: Macroeconomics* 11 (1):49–88.

Caliendo, Lorenzo, Maximiliano Dvorkin, and Fernando Parro. 2019. "Trade and labor market dynamics: General equilibrium analysis of the China trade shock." *Econometrica* 87 (3):741–835.

Carlsen, Fredrik, Jørn Rattsø, and Hildegunn E. Stokke. 2016. "Education, experience, and urban wage premium." *Regional Science and Urban Economics* 60 (1):39–49.

Combes, Pierre-Philippe, Gilles Duranton, and Laurent Gobillon. 2019. "The costs of agglomeration: House and land prices in French cities." *Review of Economic Studies* 86 (4):1556–1589.

Costa, D. L. and M. E. Kahn. 2000. "Power couples: Changes in the locational choice of the college educated, 1940-1990." *Quarterly Journal of Economics* 115 (4):1287–1315.

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

Davis, Morris A., Jonas D. M. Fisher, and Toni M. Whited. 2014. "Macroeconomic implications of agglomeration." *Econometrica* 82 (2):731–764.

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

De la Roca, Jorge, Gianmarco I. P. Ottaviano, and Diego Puga. 2023. "City of dreams." *Journal of the European Economic Association* 21 (2):690–726.

De la Roca, Jorge and Diego Puga. 2017. "Learning by working in big cities." *Review of Economic Studies* 84 (1):106–142.

Desmet, Klaus, Dávid Kristián Nagy, and Esteban Rossi-Hansberg. 2018. "The geography of development." *Journal of Political Economy* 126 (3):903–983.

Desmet, Klaus and Esteban Rossi-Hansberg. 2010. "On spatial dynamics." *Journal of Regional Science* 50 (1):43–63.

———. 2014. "Spatial development." *American Economic Review* 104 (4):1211–1243.

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

Dingel, Jonathan I. and Felix Tintelnot. 2021. "Spatial economics for granular settings." National Bureau of Economic Research, Working Paper 27287.

Donald, Eric, Masao Fukui, and Yuhei Miyauchi. 2023a. "Optimal dynamic spatial policy."

———. 2023b. "Unpacking aggregate welfare in a spatial economy."

Duranton, Gilles and Diego Puga. 2004. "Micro-foundations of urban agglomeration economies." In *Handbook of Regional and Urban Economics: Cities and Geography*, vol. 4, edited by John Vernon Henderson and Jacques-Francois Thisse, chap. 48. Elsevier, 2063–2117.

———. 2015. "Urban land use." In *Handbook of Regional and Urban Economics*, vol. 5, edited by Gilles Duranton, J. Vernon Henderson, and William C. Strange, chap. 8. Elsevier, 467–560.

———. 2020. "The economics of urban density." *Journal of Economic Perspectives* 34 (3):3–26.

———. 2023. "Urban growth and its aggregate implications." *Econometrica* 91 (6):2219–2259.

Duranton, Gilles and Anthony J. Venables. 2021. "Place-based policies: Principles and developing country applications." In *Handbook of Regional Science*, edited by M. M. Fischer and P. Nijkamp. Berlin: Springer, 1009–1030.

Dvorkin, Maximiliano A. 2023. "Heterogeneous agents dynamic spatial general equilibrium." Tech. Rep. 2023-005A.

Eaton, Jonathan and Zvi Eckstein. 1997. "Cities and growth: Theory and evidence from France and Japan." *Regional Science and Urban Economics* 27 (4):443–474.

Fajgelbaum, Pablo D. and Cecile Gaubert. 2020. "Optimal spatial policies, geography, and sorting." *Quarterly Journal of Economics* 135 (2):959–1036.

Farrokhi, Farid. 2021. "Skill, agglomeration, and inequality in the spatial economy." *International Economic Review* 62 (2):671–721.

Gabriel, Antonio and Robert E. Lucas, Jr. 2019. "Idea flows, economic growth and human capital."

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

Gaubert, Cecile, Patrick M. Kline, Damián Vergara, and Danny Yagan. 2024. "Place-based redistribution." National Bureau of Economic Research, Working Paper 28337.

Giannone, Elisa. 2022. "Skill-biased technical change and regional convergence."

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

———. 2011. *Triumph of the city: How our greatest invention makes us richer, smarter, greener, healthier, and happier.* New York: Penguin.

Glaeser, Edward L. and Joshua D. Gottlieb. 2008. "The economics of place-making policies." *Brookings Papers on Economic Activity* 39 (1):155–253.

Glaeser, Edward L. and David C. Maré. 2001. "Cities and skills." *Journal of Labor Economics* 19 (2):316–342.

Glaeser, Edward L. and Albert Saiz. 2004. "The rise of the skilled city." *Brookings-Wharton Papers on Urban Affairs* 1:47–105.

Gould, E. D. 2007. "Cities, workers, and wages: A structural analysis of the urban wage premium." *Review of Economic Studies* 74 (2):477–506.

Greaney, Brian. 2023. "Homeownership and the distributional effects of uneven regional growth."

Gruber, Jonathan and Simon Johnson. 2019. *Jump-starting America.* PublicAffairs.

Gyourko, Joseph, Albert Saiz, and Anita Summers. 2008. "A new measure of the local regulatory environment for housing markets: The Wharton residential land use regulatory index." *Urban Studies* 45 (3):693–729.

Head, Keith and John Ries. 2001. "Increasing returns versus national product differentiation as an explanation for the pattern of U.S.-Canada trade." *American Economic Review* 91 (4):858–876.

Heckman, James J. 1976. "A life-cycle model of earnings, learning, and consumption." *Journal of Political Economy* 84 (4):S11–S44.

Heckman, James J., Lance Lochner, and Christopher Taber. 1998. "Explaining rising wage inequality: Explorations with a dynamic general equilibrium model of labor earnings with heterogeneous agents." *Review of Economic Dynamics* 1 (1):1–58.

Heckman, James J., Lance J. Lochner, and Petra E. Todd. 2006. "Earnings functions, rates of return and treatment effects: The Mincer equation and beyond." In *Handbook of the Economics of Education*, vol. 1, edited by Eric A. Hanushek and Finis Welch, chap. 7. Elsevier, 307–458.

Hsieh, Chang-Tai and Enrico Moretti. 2019. "Housing constraints and spatial misallocation." *American Economic Journal: Macroeconomics* 11 (2):1–39.

Huggett, Mark, Gustavo Ventura, and Amir Yaron. 2006. "Human capital and earnings distribution dynamics." *Journal of Monetary Economics* 53 (2):265–290.

Jacobs, Jane. 1969. *The economy of cities*. New York: Vintage Books.

Jones, Charles I. 1999. "Growth: With or without scale effects?" *AEA Papers and Proceedings* 89 (2):139–144.

———. 2005. "Growth and ideas." In *Handbook of Economic Growth*, vol. 1, edited by Philippe Aghion and Steven N. Durlauf, chap. 16. Elsevier, 1063–1111.

Judd, Kenneth. 1998. *Numerical methods in economics*. Cambridge, Mass: MIT Press.

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

Kleinman, Benny, Ernest Liu, and Stephen J. Redding. 2023. "Dynamic spatial general equilibrium." *Econometrica* 91 (2):385–424.

Klette, Tor Jakob and Samuel Kortum. 2004. "Innovating firms and aggregate innovation." *Journal of Political Economy* 112 (5):986–1018.

Kredler, Matthias. 2018. "Lecture notes on the Kolmogorov forward equation."

Lasry, Jean-Michel and Pierre-Louis Lions. 2007. "Mean field games." *Japanese Journal of Mathematics* 2 (1):229–260.

Ljungqvist, Lars and Thomas J. Sargent. 2018. *Recursive macroeconomic theory*. Cambridge, MA: The MIT Press, fourth ed.

Lucas, Robert E., Jr. 1988. "On the mechanics of economic development." *Journal of Monetary Economics* 22 (1):3–42.

Lucas, Robert E., Jr. 2009. "Ideas and growth." *Economica* 76 (301):1–19.

Lucas, Robert E., Jr and Benjamin Moll. 2014. "Knowledge growth and the allocation of time." *Journal of Political Economy* 122 (1):1–50.

Luttmer, Erzo G. J. 2012. "Eventually, noise and imitation implies balanced growth." Federal Reserve Bank of Minneapolis, Working Paper 699.

Marshall, Alfred. 1890. *Principles of economics*. London: MacMillan, eighth ed.

Martellini, Paolo. 2022. "Local labor markets and aggregate productivity."

Mincer, Jacob. 1974. *Schooling, experience, and earnings.* National Bureau of Economic Research.

———. 1978. "Family migration decisions." *Journal of Political Economy* 86 (5):749–773.

———. 1997. "The production of human capital and the life cycle of earnings: Variations on a theme." *Journal of Labor Economics* 15 (1, Part 2):S26–S47.

Monras, Joan. 2020. "Immigration and wage dynamics: Evidence from the Mexican Peso Crisis." *Journal of Political Economy* 128 (8):3017–3089.

Moretti, Enrico. 2013a. *The new geography of jobs.* Mariner Books.

———. 2013b. "Real wage inequality." *American Economic Journal: Applied Economics* 5 (1):65–103.

Mulligan, Casey B. and Xavier Sala-i Martin. 1993. "Transitional dynamics in two-sector models of endogenous growth." *Quarterly Journal of Economics* 108 (3):739–773.

Oswald, Florian. 2019. "The effect of homeownership on the option value of regional migration." *Quantitative Economics* 10 (4):1453–1493.

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

Perla, Jesse, Christopher Tonetti, and Michael E. Waugh. 2021. "Equilibrium technology diffusion, trade, and growth." *American Economic Review* 111 (1):73–128.

Peters, Michael. 2022. "Market size and spatial growth—evidence from Germany's post-war population expulsions." *Econometrica* 90 (5):2357–2396.

Redding, Stephen J. and Esteban Rossi-Hansberg. 2017. "Quantitative spatial economics." *Annual Review of Economics* 9 (1):21–58.

Rendahl, Pontus. 2022. "Continuous vs. discrete time: Some computational insights." *Journal of Economic Dynamics and Control* 144.

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

Rosen, Sherwin. 1976. "A theory of life earnings." *Journal of Political Economy* 84 (4):S45–S67.

Rossi-Hansberg, Esteban. 2019. "Geography of growth and development." In *Oxford Research Encyclopedia of Economics and Finance*. Oxford University Press.

Rossi-Hansberg, Esteban, Pierre-Daniel Sarte, and Felipe Schwartzman. 2021. "Cognitive hubs and spatial redistribution." National Bureau of Economic Research, Working Paper 26267.

Rossi-Hansberg, Esteban and Mark L. J. Wright. 2007. "Urban structure and growth." *Review of Economic Studies* 74 (2):597–624.

Rudik, Ivan, Gary Lyn, Weiliang Tan, and Ariel Ortiz-Bobea. 2022. "The economic effects of climate change in dynamic spatial equilibrium."

Ruggles, Steven, Sarah Flood, Ronald Goeken, Megan Schouweiler, and Matthew Sobek. 2022. "IPUMS USA: Version 12.0."

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

Sanders, Carl and Christopher Taber. 2012. "Life-cycle wage growth and heterogeneous human capital." *Annual Review of Economics* 4 (1):399–425.

Sennewald, Ken and Klaus Wälde. 2006. "Itô's lemma and the Bellman equation for Poisson processes: An applied view." *Journal of Economics* 89 (1):1–36.

Sjaastad, Larry A. 1962. "The costs and returns of human migration." *Journal of Political Economy* 70 (5):80–93.

Stokey, Nancy L. and Robert E. Lucas, Jr. 1989. *Recursive methods in economic dynamics.* Cambridge, MA: Harvard University Press.

Uzawa, Hirofumi. 1965. "Optimum technical change in an aggregative model of economic growth." *International Economic Review* 6 (1):18–31.

Wang, Zhi. 2016. "Wage growth, ability sorting, and location choice at labor-force entry: New evidence from US Census data." *Journal of Urban Economics* 96 (1):112–120.

Appendix

A Theory

A.1 City structure and urban cost

This appendix provides microfoundations for my specification of urban costs. I build on Behrens, Duranton, and Robert-Nicoud (2013) and Duranton and Puga (2015), who extend the canonical monocentric city model to environments with heterogeneous workers. Relative to the canonical model, there are two main differences: (i) utility will not equalize across all workers, only across those of the same type; (ii) different groups will have different willingness to pay for the same location, so the overall bid-rent curve will be formed by the upper envelope of the type-specific bid-rent curves.

In each city, production takes place at a single point called the central business district (CBD). On either side of the CBD there is a line comprised of identical residences of unit length. Workers commute from their residence to the CBD at a cost. The cost is in units of time, so commuting is more expensive for highly-productive workers. In particular, I assume that commuting from a residence at distance ℓ from the CBD requires $\vartheta_n \ell^{\theta_n}$ units of time, so the commuting cost for a worker with productivity z residing at distance ℓ is $(\vartheta_n T_n) z \ell^{\theta_n}$. The rent at distance ℓ in city n at date t is denoted $r_n(\ell, t)$.

Each worker chooses her residence by trading off commuting costs and land rents. She is allowed to move freely within the city at every date, so her choice of residence is static and separable from her dynamic decision problem specified in the main text.[78] Her objective is to minimize her total urban cost at each date, which yields the Alonso-Muth condition

$$\theta_n \vartheta_n T_n z \ell^{\theta_n - 1} = -\partial_\ell r_n(\ell, t). \tag{A.1}$$

A *residential equilibrium* at date t consists of a rent gradient $r_n(\ell, t)$ and an assignment function $\mathcal{L}_n(z, t)$ such that the Alonso-Muth condition (A.1) holds for all (z, ℓ) and all workers are allocated to a residence.

The Alonso-Muth condition implies that rents fall with distance and that willingness to pay for a residence at a given distance is increasing in human capital. It follows from the

[78] The only concern would be if the worker's time allocation is forced to a corner. This never occurs in the quantified model.

second fact that workers will sort perfectly such that more productive workers will be closer to the CBD. With a slight abuse of notation, let $G_n(z,t)$ denote the marginal distribution of human capital in city n at date t and $g_n(z,t)$ its density.[79] The supply of residential land between ℓ and $\ell + d\ell$ is simply $2d\ell$. For a proposed assignment function $\mathcal{L}$ with inverse $\mathcal{L}^{-1}$, there is a density $g_n(\mathcal{L}^{-1}(\ell),t)$ of residents with human capital between $\mathcal{L}^{-1}(\ell)$ and $\mathcal{L}^{-1}(\ell) - \frac{d\mathcal{L}^{-1}(\ell)}{d\ell}d\ell$. Equating supply and demand yields the relationship

$$-g_n(\mathcal{L}^{-1}(\ell),t)\frac{d\mathcal{L}^{-1}(\ell)}{d\ell}d\ell = 2d\ell,$$

which, by the inverse function theorem, simplifies to

$$-\frac{d\mathcal{L}(z)}{dz} = \frac{1}{2}g_n(z,t).$$

Noting that the city edge will be at distance $\frac{1}{2}L_n(t)$, one can integrate this expression over z to get the equilibrium assignment function

$$\mathcal{L}_n(z,t) = \frac{L_n(t)}{2}[1 - G_n(z,t)]. \tag{A.2}$$

So workers are spread across the linear city according to their productivity quantiles, with the most productive closest to the CBD.

To solve for the equilibrium bid-rent curve, I integrate the Alonso-Muth condition (A.1) over the line using the assignment function (A.2) to substitute for z. I normalize rent at the city edge to zero, so

$$r_n(\ell,t) = \theta_n \vartheta_n T_n \int_{\ell}^{L_n(t)/2} G_n^{-1}\left(1 - \frac{2l}{L_n(t)}, t\right) l^{\theta_n - 1}\, dl. \tag{A.3}$$

In general, the equilibrium bid-rent curve inherits the lack of a closed-form solution from the human capital distribution. Nevertheless, one can identify an important trait of the total urban cost $(\vartheta_n T_n) z \ell^{\theta_n} + r_n(\ell,t)$, namely that it can be detrended by the same rate γ as the rest of the model. This is because workers are assigned to distances in accordance with their productivity quantiles, all of which grow at the same constant rate γ by definition of the BGP. One could substitute this specification of the urban cost into the main text and go on to solve for the BGP without hitting any theoretical obstructions.

I do not proceed along this route for a computational reason: it is much faster and easier to iterate over a small set of location characteristics than a large one. In the main text, workers need only consider the vibrancy and population of each city, which has dimension $2N$. Under the specification above, however, they would need to consider all vibrancies and rent gradients. The latter are continuous functions over the space of detrended productivities; their set, when discretized with M_x points, consists of NM_x values. Since sorting by skill within the city is not a focus of this paper, I want to avoid this computational burden.

Accordingly, I further suppose that within each city there is a local government who collects land rents and redistributes them to current residents. Note that the local government can target residents by their productivity without distorting their choice of distance from the CBD. The only constraint on the local government is that it cannot run a deficit;

[79] That is, $g_n(z,t) = (L/L_n(t)) \int_0^A g_n(a,z,t)\, da$ and $G_n(z,t) = \int^z g_n(\zeta,t)\, d\zeta$.

for simplicity, surpluses are simply burned.[80] In order to generate the urban cost specified in the main text, the government of city n must impose the reimbursement schedule

$$\varrho_n(z,t) = r_n(\mathcal{L}_n(z,t),t) - (\vartheta_n T_n) z \left[\left(\frac{L_n(t)}{2} \right)^{\theta_n} - \mathcal{L}_n(z,t)^{\theta_n} \right] \tag{A.4}$$

This reimbursement schedule is feasible: the local government just hands back each worker's land rent minus a non-negative amount. All workers are weakly worse off than they would be without the transfers, and more so for higher-skilled ones, but this is of no consequence for the questions of this paper.[81]

A.2 Derivation of the Hamilton-Jacobi-Bellman equation

I start by fully specifying the sequence of idiosyncratic preference shocks and (utility-denominated) migration costs that accrue to a worker given her particular location history. I then recast the worker's sequence problem in recursive form. Along the way, I establish the main lemma: the worker's value function is homogeneous of degree one in the product of all her *past* shocks and costs, making them irrelevant for current decisions. This result allows us to obtain the Hamilton-Jacobi-Bellman equation (5) in the main text and other isomorphic representations akin to what is often seen in the discrete-time literature on optimal migration. I conclude this section by deriving the optimal investment rule.

A.2.1 Specification of a worker's location history

The location history of an individual ω at date t is a set

$$\mathcal{H}^\omega(t) = \{n_0^\omega, I^\omega(t), \mathbf{H}_b^\omega(t), \mathbf{H}_n^\omega(t)\}$$

with the following elements:

- her birth location n_0^ω;
- her count of migration opportunities $I^\omega(t) \in \mathbb{N}$, which evolves as a Poisson process with arrival rate λ;
- an $I^\omega(t) \times N$ matrix $\mathbf{H}_b^\omega(t)$ recording her idiosyncratic preference draws, which are drawn i.i.d. from the Fréchet distribution;
- an $I^\omega(t) \times N$ binary Markov matrix $\mathbf{H}_n^\omega(t)$ recording her location choices, which are optimal controls that take into account her idiosyncratic preference draws.

Let $\mathbf{N}$ be the column vector of location labels $[1, 2, \ldots, N]$. The set of locations that individual ω has chosen through date t is the set of elements in the vector $\mathbf{n}^\omega(t) \equiv \mathbf{H}_n^\omega(t)\mathbf{N}$, which we denote by n_ι^ω for $\iota = 1, 2, \ldots, I^\omega(t)$. For any date $t' < t$, her location at t' is encoded as

$$n^\omega(t') = \begin{cases} \mathbf{e}^\omega(t',t)'\mathbf{n}^\omega(t) & \text{if } I^\omega(t) > 0 \\ n_0^\omega & \text{else} \end{cases}$$

[80] Alternatively, one could suppose that they are surrendered to a national pool and redistributed equally to all workers as a lump sum so that they do not distort any choices.

[81] A helpful analogy is to consider how we often introduce subsidies into New Keynesian models to correct for inefficient production when dealing with the inefficiency would only be a nuisance for the question at hand.

where $\mathbf{e}^{\omega}(t',t)$ is the elementary column vector of length $I^{\omega}(t)$ with a one in entry $I^{\omega}(t')$. The function $n^{\omega}(t)$ is continuous from the right. The law of motion for an individual's location is therefore

$$dn^{\omega}(t) = [n^{\omega}(t) - n^{\omega}(t_{-})]\, dI^{\omega}(t).$$

where $n^{\omega}(t_{-})$ is the limit of $n^{\omega}(t)$ from the left.

Her flow utility at date t is the product of her current local amenity, her consumption, and her history of permanent preference shifters and migration costs. Define

$$\begin{aligned}\hat{\mathbf{b}}^{\omega}(t) &\equiv \mathbf{H}_{n}^{\omega}(t)'\mathbf{H}_{b}^{\omega}(t)\\ \hat{\boldsymbol{\tau}}^{\omega}(t) &\equiv \mathbf{H}_{n^{-}}^{\omega}(t)'\mathbf{H}_{n}^{\omega}(t)\mathbf{T}\end{aligned}$$

where $\mathbf{H}_{n^{-}}^{\omega}(t)$ is the *origin* matrix, which is just the one-step lag of the choice matrix $\mathbf{H}_{n}^{\omega}(t)$,

$$\mathbf{H}_{n^{-}}^{\omega}(t) \equiv \begin{bmatrix}\mathbf{e}'_{n_0^{\omega}}\\ \mathbf{0}'_N\\ \vdots\\ \mathbf{0}'_N\end{bmatrix}_{I^{\omega}(t)} + \begin{bmatrix}0 & \dots & \dots & \dots & 0\\ 1 & 0 & \dots & \dots & 0\\ 0 & 1 & 0 & \dots & 0\\ \vdots & \ddots & \ddots & \ddots & \vdots\\ 0 & \dots & 0 & 1 & 0\end{bmatrix}\mathbf{H}_{n}^{\omega}(t),$$

and $\mathbf{T} \equiv [\tau_{nm}]_{N\times N}$ is the matrix of bilateral migration costs. Both $\hat{\mathbf{b}}^{\omega}(t)$ and $\hat{\boldsymbol{\tau}}^{\omega}(t)$ are column vectors of length $I^{\omega}(t)$ with typical elements $\hat{b}_{\iota}^{\omega} = b_{n_{\iota}^{\omega}}^{\omega}$ and $\hat{\tau}_{\iota}^{\omega} = \tau_{n_{\iota-1}^{\omega},n_{\iota}^{\omega}}$, respectively. With this notation, flow utility can be written as

$$U_n[c;\mathcal{H}^{\omega}] = \left(\prod_{\iota=1}^{I^{\omega}(t)} \frac{\hat{b}_{\iota}^{\omega}}{\hat{\tau}_{\iota}^{\omega}}\right) u_n(c)$$

where $u_n(c) = B_n c$ is the amenity-adjusted flow of the consumption good.

A.2.2 Recasting the sequence problem in recursive form

Under the assumption that location-specific preference shocks are permanent in the particular sense defined in the previous section, the value function of an individual ω of age a with human capital z and history $\mathcal{H}$ in location n at date t is

$$\tilde{V}_n(a,z,t;\mathcal{H}) = \max_{s} \mathbb{E}\left[\int_t^{t+A-a} e^{-\rho r}\left(\prod_{\iota=1}^{I^{\omega}(r)} \frac{\hat{b}_{\iota}^{\omega}}{\hat{\tau}_{\iota}^{\omega}}\right) B_n c_n(s,z,r)\, dr \;\middle|\; \begin{matrix} a(t)=a\\ z(t)=z\\ n(t)=n\\ \mathcal{H}(t)=\mathcal{H}\end{matrix}\right] \tag{A.5}$$

subject to the budget constraint

$$c_n(s,z,r) = \left[T_n L_n(r)^{\alpha}(1-s) - p_n L_n(r)^{\theta_n}\right] z$$

and the laws of motion

$$\begin{aligned}da &= dr\\ dz &= \left[\kappa(s) z^{\beta} Z_n^{1-\beta}\right]\, dr\end{aligned}$$

and those for $n(r)$ and $\mathcal{H}(r)$ specified in the previous section. The general form of the corresponding Hamilton-Jacobi-Bellman equation (HJB) is

$$\rho\tilde{V}_n(a,z,t;\mathcal{H}) = \max_s \left\{ \left(\prod_{\iota=1}^{I^\omega(t)} \frac{\hat{b}_\iota^\omega}{\hat{\tau}_\iota^\omega} \right) u_n(s,z,t) + \frac{1}{dt}\mathbb{E}\, dV_{n(t)}[a(t),z(t),t;\mathcal{H}(t)] \right\}$$

subject to the same constraints. Following Sennewald and Wälde (2006, §2.4), we can rewrite the HJB in three steps.

1. Use a change of variables to write

$$\begin{aligned} d\tilde{V}_n(a,z,t;\mathcal{H}) = \{&\partial_t\tilde{V}_n(a,z,t;\mathcal{H}) + \partial_a\tilde{V}_n(a,z,t;\mathcal{H})[\dot{a}] + \partial_z\tilde{V}_n(a,z,t;\mathcal{H})[\dot{z}]\}\, dt \\ &+ [\tilde{V}_{n_+}(a,z,t;\mathcal{H}_+) - \tilde{V}_n(a,z,t;\mathcal{H})]\, dI(t), \end{aligned}$$

   where $\mathcal{H}_+$ is the limit of the worker's history from the right, incorporating the optimal location choice at the next migration opportunity, n_+.

2. Using $\mathbb{E}\, dI(t) = \lambda\, dt$ and that $\dot{a} = 1$,

$$\begin{aligned} \mathbb{E}\, d\tilde{V}_n(a,z,t) = \{&\partial_t\tilde{V}_n(a,z,t;\mathcal{H}) + \partial_a\tilde{V}_n(a,z,t;\mathcal{H}) + \partial_z\tilde{V}_n(a,z,t;\mathcal{H})[\dot{z}]\}\, dt \\ &+ \lambda[\tilde{V}_{n_+}(a,z,t;\mathcal{H}_+) - \tilde{V}_n(a,z,t;\mathcal{H})]\, dt. \end{aligned}$$

3. Divide by dt.

The resulting HJB is

$$\begin{aligned} \rho\tilde{V}_n(a,z,t;\mathcal{H}) = \max_s \Bigg\{ &\left(\prod_{\iota=1}^{I^\omega(t)} \frac{\hat{b}_\iota^\omega}{\hat{\tau}_\iota^\omega} \right) u_n(s,z,t) \\ &+ \partial_z\tilde{V}_n(a,z,t;\mathcal{H})\left[\kappa(s)z^\beta Z_n^{1-\beta}\right] \Bigg\} \\ &+ \partial_t\tilde{V}_n(a,z,t;\mathcal{H}) + \partial_a\tilde{V}_n(a,z,t;\mathcal{H}) \\ &+ \lambda\left[\tilde{V}_{n_+}(a,z,t;\mathcal{H}_+) - \tilde{V}_n(a,z,t;\mathcal{H})\right], \end{aligned} \tag{A.6}$$

and we have the following lemma by construction.

Lemma 2. *The value function $\tilde{V}_n(a,z,t;\mathcal{H})$ specified in (A.5) is homogeneous of degree one in the agent's location history up to time t from the left.*

Lemma 2 states that the scale of the value function can be renormalized after each migration opportunity without affecting future decisions, even though those decisions are forward-looking. The key is that once a particular preference shock is realized and its associated city is chosen, $\hat{b}_\iota^\omega$, it multiplies *all future utility flows*, regardless of subsequent investments or migration decisions. The same goes for migration frictions, $\hat{\tau}_\iota^\omega$.[82] Since

[82]Because flow utility is a linear function of skill, migration frictions act like a particular form of skill depreciation in which labor productivity is reduced but investment productivity is not. Likewise, idiosyncratic preference shocks can be reinterpreted as a particular kind of place-specific productivity shocks. See Bryan and Morten (2019) for a model of location choice developed under this interpretation.

the scale of utility is irrelevant for decisions, we can arbitrarily renormalize the scale of a worker's history up to the present moment, which we will do from now on when we write the worker's problem.[83] Bilal (2023, §IV.A.3–4), whose model is likewise in continuous time, uses a similar specification for preference shocks, but he ignores bilateral migration costs. Desmet, Nagy, and Rossi-Hansberg (2018) also use a similar specification in discrete time, but they require the additional assumption that bilateral migration costs are, in a particular sense, decomposable and reversible so that the migration decision problem reduces to a static one.[84]

The final maneuver is to take expectations with respect to the Fréchet taste shock from the contemporaneous migration opportunity. Letting $V_n(a, z, t)$ denote the value function that incorporates said expectation, we have,

$$\begin{aligned}\rho V_n(a,z,t) = \max_s \left\{u_n(s,z,t) + \partial_z V_n(a,z,t)\left[\kappa(s) z^\beta Z_n^{1-\beta}\right]\right\} \\ + \partial_t V_n(a,z,t) + \partial_a V_n(a,z,t) \\ + \lambda\left\{\mathbb{E}_{b^\omega}\left[\max_i \frac{b_i^\omega}{\tau_{ni}} V_i(a,z,t)\right] - V_n(a,z,t)\right\}.\end{aligned}$$

With the usual algebra, we find that

$$\mathbb{E}_{b^\omega}\left[\max_i b_i^\omega \frac{V_i(a,z,t)}{\tau_{ni}}\right] = \left(\sum_i \tau_{ni}^{-\epsilon} V_i(a,z,t)^\epsilon\right)^{1/\epsilon},$$

and that the optimal migration shares are

$$m_{ni}(a,z,t) \equiv \mathbb{P}\left\{\arg\max_k b_k^\omega \frac{V_k(a,z,t)}{\tau_{nk}} = i\right\} = \frac{\tau_{ni}^{-\epsilon} V_i(a,z,t)^\epsilon}{\sum_k \tau_{nk}^{-\epsilon} V_k(a,z,t)^\epsilon}.$$

It follows that

$$\mathbb{E}_{b^\omega}\left[\max_i b_i^\omega \frac{V_i(a,z,t)}{\tau_{ni}}\right] = m_{ni}(a,z,t)^{-\frac{1}{\epsilon}} \frac{V_i(a,z,t)}{\tau_{ni}}, \quad \text{any } i. \tag{A.7}$$

We emphasize that the right-hand side holds for any city i. This makes sense because the HJB is agnostic about *which specific location* attains the maximum.

Equation (A.7) yields two particularly useful specifications. The first is analogous to what is used in Caliendo, Dvorkin, and Parro (2019) and related models:[85]

$$\mathbb{E}_{b^\omega}\left[\max_i b_i^\omega \frac{V_i(a,z,t)}{\tau_{ni}}\right] = m_{nn}(a,z,t)^{-\frac{1}{\epsilon}} V_n(a,z,t).$$

[83] By the same logic, we can further introduce a permanent preference reduction at each migration opportunity to keep the mean equal to one.

[84] Specifically, they assume that $m_{ni} = m_n^{\text{out}} m_i^{\text{in}}$ for all n, i and $m_{nn} = 1$ for all n, which together imply $m_n^{\text{out}} = 1/m_n^{\text{in}}$. That is, the cost of migrating *into* a location is always fully recouped upon migrating *out of* it, so there is no sense in which one location provides greater "migration access" than another.

[85] See also, for example, Artuç, Chaudhuri, and McLaren (2010) and Balboni (2019). Notice that each of these papers directly specifies type-1 extreme value preference shocks and migration costs as additive with the worker's *value function* (or its log). The resulting migration decision problem, conditional on having the opportunity to move, is then isomorphic to the one presented here. Balanced growth requires a specification that is multiplicative in levels (as I specify here) or additive in logs (as in, e.g., Monras, 2020).

With this specification, only the current location appears in the HJB; in fact, the entire migration block can be brought over to the left-hand side, appearing analogously to an adjustment of the discount rate. While simple, this specification appears to divorce the HJB, for which the identity of the location choice is inconsequential, from the Kolmogorov forward equation, for which it is crucial to track cross-city flows. Accordingly, I prefer a second specification:

$$\mathbb{E}_{b^\omega}\left[\max_i b_i^\omega \frac{V_i(a,z,t)}{\tau_{ni}}\right] = \sum_i m_{ni}(a,z,t)^{\frac{\epsilon-1}{\epsilon}} \frac{V_i(a,z,t)}{\tau_{ni}}.$$

The HJB can then be written as

$$\begin{aligned}\rho V_n(a,z,t) = \max_s &\left\{u_n(s,z,t) + \partial_z V_n(a,z,t)\left[\kappa(s) z^\beta Z_n^{1-\beta}\right]\right\} \\ &+ \partial_t V_n(a,z,t) + \partial_a V_n(a,z,t) \\ &+ \lambda \sum_i m_{ni}(a,z,t)\left[\frac{m_{ni}(a,z,t)^{-\frac{1}{\epsilon}}}{\tau_{ni}} V_i(a,z,t) - V_n(a,z,t)\right]\end{aligned}$$

where we are free to bring $V_n(a,z,t)$ inside the brackets of the last line because $\sum_i m_{ni}(a,z,t) = 1$ for all (n,a,z,t). The HJB, stated this way, has an alternative interpretation as the value function of a representative agent in state (n,a,z,t) who maximizes with the set of *continuous* control variables $\{s_n(a,z,t), \mathbf{m}_{n\cdot}(a,z,t)\}$ subject to the constraint $\sum_i m_{ni}(a,z,t) = 1$.[86]

Lemma 3. *A representative agent with state (n,a,z,t) maximizes the value function*

$$\begin{aligned}\rho V_n(a,z,t) = \max_{s,\mathbf{m}} &\left\{u_n(s,z,t) + \partial_z V_n(a,z,t)\left[\kappa(s) z^\beta Z_n^{1-\beta}\right]\right\} \\ &+ \partial_t V_n(a,z,t) + \partial_a V_n(a,z,t) \\ &+ \lambda \sum_i m_{ni}\left[\frac{m_{ni}^{-\frac{1}{\epsilon}}}{\tau_{ni}} V_i(a,z,t) - V_n(a,z,t)\right]\end{aligned}$$

subject to the constraint $\sum_i m_{ni} = 1$ for all n.

Proof. Take first order conditions with respect to m_{ni} for all i:

$$\begin{aligned}0 &= \lambda\left(\frac{\epsilon-1}{\epsilon}\right) m_{ni}^{-\frac{1}{\epsilon}} \frac{V_i}{\tau_{ni}} - \Lambda, \quad \forall n,i \\ 0 &= 1 - \sum_i m_{ni}, \quad \forall n\end{aligned}$$

where Λ is the Lagrange multiplier on the constraint. It follows that

$$m_{ni} = K\left(\frac{V_i}{\tau_{ni}}\right)^\epsilon, \quad \forall n,i.$$

for some constant K. The constraint $\sum_i m_{ni} = 1$ then implies that

$$m_{ni} = \frac{\tau_{ni}^{-\epsilon} V_i}{\sum_k \tau_{nk}^{-\epsilon} V_k}, \quad \forall n,i.$$

just as in Equation (7). □

[86]See Rudik et al. (2022, Prop. 1) and Donald, Fukui, and Miyauchi (2023a, Lemma 1) for related results.

A.2.3 Optimal investment rule

Consider the maximization over a worker's raw labor allocated to investment, s, in the HJB equation (5). With $k(s) \equiv \frac{\bar{\kappa}}{\eta} s^\eta$, the first-order condition with respect to s is

$$0 = -B_n T_n L_n(t)^\alpha z + \partial_z V_n(a,z,t) \bar{\kappa} s^{\eta-1} z^\beta Z_n(t)^{1-\beta}.$$

Solving for s yields the rule

$$s_n(a,z,t) = \left(\frac{\bar{\kappa}\, \partial_z V_n(a,z,t)}{B_n T_n L_n(t)^\alpha} \left(\frac{Z_n(t)}{z} \right)^{1-\beta} \right)^{\frac{1}{1-\eta}}. \tag{A.8}$$

The corresponding optimal drift of the worker's human capital is then

$$h_n(a,z,t) = \kappa[s_n(a,z,t)] z^\beta Z_n(t)^{1-\beta}. \tag{A.9}$$

A.3 Derivation of the Kolmogorov forward equation[87]

The Hamilton-Jacobi-Bellman equation derived above encodes a controlled stochastic process for the state of an individual agent. The coupled Kolmogorov forward equation answers the following question: given this controlled stochastic process for each agent and some initial distribution of agents $g_n(a,z,0)$, what is the density function $g_n(a,z,t)$ over the state space at each point in time $t > 0$?

Following Kredler (2018), we will derive the Kolmogorov forward equation by first considering the discrete-time evolution of a moment of an arbitrary test function. Fix a continuously differentiable function f that vanishes on the boundaries of the state space. Consider the moment

$$M_f(t) = \sum_n \iint f_n(a,z) g_n(a,z,t)\, da\, dz.$$

If we want to approximate the change in M_f over a small time interval Δ, we can do so in two ways:

1. If we already knew the density, we could simply approximate the change in M_f by linearly extrapolating the density at each point in the state space:

$$M_f(t+\Delta) = \sum_n \iint f_n(a,z) \frac{\partial g_n(a,z,t)}{\partial t} \Delta\, da\, dz.$$

2. Given the controlled drifts and jumps of agents at each point in the state space, we can instead start by deriving the expected change in f:

$$\begin{aligned} df_n(a,z) &= \frac{\partial f}{\partial a}\frac{\partial a}{\partial t}\Delta + \frac{\partial f}{\partial z}\frac{\partial z}{\partial t}\Delta + p(\Delta) \sum_i p\{n \to i\}[f_i(a,z) - f_n(a,z)] \\ &= \frac{\partial f}{\partial a}\Delta + \frac{\partial f}{\partial z} h_n(a,z,t)\,\Delta + \lambda \sum_i m_{ni}(a,z,t)[f_i(a,z) - f_n(a,z)]\,\Delta \end{aligned}$$

[87] I thank Tommaso Santini for helpful comments on a previous version of this derivation.

where the second line substitutes in the controlled stochastic processes and uses the small-Δ approximation $p(\Delta) = 1 - e^{-\lambda\Delta} \approx \lambda\Delta$ for the probability of a migration opportunity arriving. Then integrate over all points in the state space, and for each point integrate over all possible paths crossing that point at t to get

$$M_f(t+\Delta) = \sum_n \iint g_n(a,z,t)\left[\frac{\partial}{\partial a} f_n(a,z) + h_n(a,z,t)\frac{\partial}{\partial z} f_n(a,z) + \lambda \sum_i m_{ni}(a,z,t)[f_i(a,z) - f_n(a,z)]\right]\Delta\, da\, dz.$$

Finally, integrate by parts, recalling that f vanishes on the boundaries, to get

$$M_f(t+\Delta) = \sum_n \iint f_n(a,z)\left[-\frac{\partial}{\partial a} g_n(a,z,t) - \frac{\partial}{\partial z}[h_n(a,z,t) g_n(a,z,t)] + \lambda \sum_i m_{in}(a,z,t)[g_i(a,z,t) - g_n(a,z,t)]\right]\Delta\, da\, dz,$$

where the second line just rearranges terms.[88]

Now set the two expressions equal, divide by Δ, and note that, because f was arbitrary, we must have, at every point in the state space,

$$\partial_t g_n(a,z,t) = -\frac{\partial}{\partial a} g_n(a,z,t) - \frac{\partial}{\partial z}[h_n(a,z,t) g_n(a,z,t)] + \lambda \sum_i m_{in}(a,z,t)[g_i(a,z,t) - g_n(a,z,t)],$$

which is the Kolmogorov forward equation (9).

A.4 Derivation of the balanced growth path

A.4.1 Detrending the equilibrium conditions

The goal in this section is to derive the detrended equations at the end of Section 2.2.4. To that end, recall the Hamilton-Jacobi-Bellman equation (5) and the Kolmogorov forward equation (9). Recall also Definition 2, which states that a balanced growth path is a number γ and a tuple of functions (v, σ, μ, ϕ) on $\mathcal{N} \times [0, A] \times \mathbb{R}_+$ such that

$$\begin{aligned} V_n(a,z,t) &= e^{\gamma t}\ \ v_n(a,x) \\ s_n(a,z,t) &= \sigma_n(a,x) \\ m_n(a,z,t) &= \mu_n(a,x) \\ g_n(a,z,t) &= e^{-\gamma t}\phi_n(a,x) \end{aligned}$$

for all (n,a,z,t) and (V,s,m,g) is an equilibrium with the initial condition $g_n(a,z,0) = \phi_n(a,z)$, where $x \equiv z e^{-\gamma t}$ denotes productivity relative to trend.

Start with the Hamilton-Jacobi-Bellman equation. We want to show that (i) the expression $e^{\gamma t}$ can be factored out of each of its terms and that (ii) the expression that

[88] Note that $\sum_n \sum_i g_n m_{ni} f_i = \sum_i \sum_n g_i m_{in} f_n$ simply by relabeling.

remains after dividing through by $e^{\gamma t}$ is independent of time. We will go term-by-term through the derivation:

$$
\begin{aligned}
V_n(a,z,t) &= e^{\gamma t} v_n(a,x), \\
\partial_t V_n(a,z,t) &= \gamma e^{\gamma t} v_n(a, z e^{-\gamma t}) - \gamma [z e^{-\gamma t}] e^{\gamma t}\, \partial_2 v_n(a, z e^{-\gamma t}) \\
&= e^{\gamma t} [\gamma v_n(a,x) - \gamma x\, \partial_x v_n(a,x)], \\
\partial_z V_n(a,z,t) &= e^{\gamma t} [\partial_z v_n(a, z e^{-\gamma t})] \\
&= e^{\gamma t} e^{-\gamma t}\, \partial_2 v_n(a, z e^{-\gamma t}) \\
&= \partial_x v_n(a,x), \\
Z_n(t) &= \left(L \iint \left(x e^{\gamma t}\right)^{\zeta} e^{-\gamma t} \phi_n(a,x)\, da\, \frac{dz}{dx}\, dx \right)^{1/\zeta} \\
&= e^{\gamma t} \left(L \iint x^{\zeta} \phi_n(a,x)\, da\, dx \right)^{1/\zeta} \\
&\equiv e^{\gamma t} X_n, \\
s_n(a,z,t) &= \left(\frac{\bar{\kappa}\, \partial_z V_n(a,z,t)}{B_n T_n L_n^{\alpha}} \left(\frac{Z_n(t)}{z} \right)^{1-\beta} \right)^{\frac{1}{1-\eta}} \\
&= \left(\frac{\bar{\kappa}\, \partial_x v_n(a,x)}{B_n T_n L_n^{\alpha}} \left(\frac{X_n}{x} \right)^{1-\beta} \right)^{\frac{1}{1-\eta}} \\
&\equiv \sigma_n(a,x), \\
\left[\kappa(s) z^{\beta} Z_n(t)^{1-\beta}\right] &= \left[\kappa(\sigma) (x e^{\gamma t})^{\beta} (e^{\gamma t} X_n)^{1-\beta}\right], \\
&= e^{\gamma t} \left[\kappa(\sigma) x^{\beta} X_n^{1-\beta}\right], \\
\partial_a V_n(a,z,t) &= e^{\gamma t} \partial_a v_n(a,x), \\
m_{ni}(a,z,t) &= \frac{\tau_{ni}^{-\epsilon} [e^{\gamma t} v_i(a,x)]^{\varepsilon}}{\sum_k \tau_{nk}^{-\epsilon} [e^{\gamma t} v_k(a,x)]^{\varepsilon}} \\
&= \frac{\tau_{ni}^{-\epsilon} v_i(a,x)^{\varepsilon}}{\sum_k \tau_{nk}^{-\epsilon} v_k(a,x)^{\varepsilon}} \\
&\equiv \mu_{ni}(a,x).
\end{aligned}
$$

All together, we see that $e^{\gamma t}$ can be factored out of the Hamilton-Jacobi-Bellman equation, leaving us with

$$
\begin{aligned}
(\rho - \gamma) v_n(a,x) = B_n &\left(T_n L_n^{\alpha} [1 - \sigma_n(a,x)] - p_n L_n^{\theta_n} \right) x \\
&+ \partial_x v_n(a,x) \left(\kappa[\sigma_n(a,x)] x^{\beta} X_n^{1-\beta} - \gamma x \right) \\
&\quad + \lambda \sum\nolimits_i \mu_{ni}(a,x) \left[\frac{\mu_{ni}(a,x)^{-\frac{1}{\epsilon}}}{\tau_{ni}} v_i(a,x) - v_n(a,x) \right] \\
&\qquad + \partial_a v_n(a,x),
\end{aligned}
$$

with terminal condition $v_n(A, x) = 0$ for all (n, x).

Now turn to the Kolmogorov forward equation. We want to show that (i) the expression $e^{-\gamma t}$ can be factored out of each of its terms and that (ii) the expression that remains after dividing through by $e^{-\gamma t}$ is independent of time. We will again proceed term-by-term:

$$\begin{aligned}
g_n(a,z,t) &= e^{-\gamma t}\phi_n(a,x), \\
\partial_t g_n(a,z,t) &= -\gamma e^{-\gamma t}\phi_n(a,x) - \gamma x e^{-\gamma t}\partial_x \phi_n(a,x), \\
\frac{\partial}{\partial z}[h_n(a,z,t)g_n(a,z,t)] &= \frac{\partial}{\partial z}\left(\frac{\partial z}{\partial x}\frac{\partial x}{\partial z}\right)[e^{\gamma t}\kappa(\sigma)x^\beta X_n^{1-\beta}e^{-\gamma t}\phi_n(a,x)], \\
&= e^{-\gamma t}\left(\partial_x[\kappa(\sigma)x^\beta X_n^{1-\beta}\phi_n(a,x)]\right), \\
\partial_a g_n(a,z,t) &= e^{-\gamma t}\partial_a\phi_n(a,x), \\
\underline{g}(z,t) &\equiv e^{-\gamma t}\underline{\phi}(x),
\end{aligned}$$

where the last line is a valid definition of $\underline{\phi}$ by Assumption 1. All together, we see that $e^{-\gamma t}$ can be factored out of the Kolmogorov forward equation, leaving us with

$$\begin{aligned}
-\gamma[\phi_n(a,x) + x\,\partial_x\phi_n(a,x)] = -\partial_x[h_n(a,x)\phi_n(a,x)] \\
- \lambda[1-\mu_{nn}(a,x)]\phi_n(a,x) \\
+ \lambda\sum_{i\neq n}\mu_{in}(a,x)\phi_i(a,x) \\
- \partial_a\phi_n(a,x)
\end{aligned} \tag{A.10}$$

or, after rearranging,

$$\begin{aligned}
0 = -\partial_x[(h_n(a,x) - \gamma x)\phi_n(a,x)] - \lambda[1-\mu_{nn}(a,x)]\phi_n(a,x) \\
+ \lambda\sum_{i\neq n}\mu_{in}(a,x)\phi_i(a,x) - \partial_a\phi_n(a,x)
\end{aligned}$$

with the slight abuse of notation

$$h_n(a,x) \equiv \kappa[\sigma_n(a,x)]x^\beta X_n^{1-\beta}.$$

A.4.2 Properties of the value function

Lemma 1. *The value function $v_n(a,x)$ is decreasing in age, a, and increasing in relative human capital, x. Moreover, the second partial $\partial_{x,x}v_n(a,x)$ is always non-positive, as is the cross partial $\partial_{a,x}v_n(a,x)$.*

Proof. The value function $v_n(a,x)$ does not have a closed form, but one can nonetheless deduce the four claims. This argument takes as given that, for any given n, there are no kinks in the value function and all derivatives up to second order exist.

(i) Consider two workers, one with state (n,a,x) and another with state (n,a',x), where $a' < a$. Let the first worker use her optimal investment and migration policies so that she achieves the value $v_n(a,x)$. Let the second worker copy the investment and migration decision of the first worker. The second worker attains a value $\tilde{v}_n(a',x) > v_n(a,x)$ because he has extra life over which to reap returns from his accumulated skill. If he used his own optimal policies, he would attain the value $v_n(a',x) \geq \tilde{v}_n(a',x)$.

(ii) Consider two workers, one with state (n, a, x) and another with state (n, a, x'), where $x < x'$ and $a < A$. Let the first worker use her optimal investment and migration policies so that she achieves the value $v_n(a, x)$. Let the second worker copy the investment and migration decision of the first worker. The second worker attains a value $\tilde{v}_n(a, x') > v_n(a, x)$ because, for any level of investment, he will earn more income and reap higher returns to his investment. If he used his own optimal policies, he would attain the value $v_n(a, x') \geq \tilde{v}_n(a, x')$. When $a = A$, $v_n(A, x) = 0$ for all x, so $\partial_x v_n(A, x) = 0$.

(iii) The concavity of the value function in human capital follows from backwards induction by repeatedly applying the argument used in Stokey and Lucas (1989, Theorem 4.8) to the discrete-time analogue of the value function and taking limits as $\Delta t \to 0$. Because the human capital production function is strictly concave in current human capital, the worker's constraint set is convex. Because, in addition, her earnings are linear in her human capital, it is easy to verify that her flow utility is jointly concave in (i) her current human capital and (ii) her human capital after investing for a period of length Δt. Finally, the terminal value function $v_n(A, x) \equiv 0$ is trivially (weakly) concave in human capital.

(iv) This follows from (i) and (ii) by backward induction from the marginal value of human capital at the maximum age: $\partial_x v_n(A, x) = 0$ for all (n, x).

□

A.4.3 Derivation of the growth rate

The growth rate γ is determined by integrating (in the sense of Lebesgue) the Kolmogorov forward equation over the state space and leveraging two facts: (i) there is no net migration on a BGP (Equation (11)), and (ii) the marginal density with respect to age is uniform. I now provide a formal proof.

Theorem 1. *A balanced growth path characterized by the functions $\{v, \sigma, \mu, \phi\}$ has the growth rate*

$$\gamma = \frac{\sum_n \int_0^A \kappa[\sigma_n(a, x)] x^\beta X_n^{1-\beta} \, \phi_n(a, x) \, da}{\sum_n \int_0^A x \, \phi_n(a, x) \, da}, \quad \forall x \in \operatorname{supp}(\phi). \tag{16}$$

Proof. Let $\mathcal{X} \equiv [x_{\text{supp}}, \infty) \subset \mathbb{R}_{++}$, where x_{supp} is any arbitrary relative productivity in the support of ϕ. Although the support of ϕ is endogenous, Assumption 1 guarantees that $x_{\text{supp}} > 0$ for any permissible choice. Moreover, because the vibrancy is well-defined, it must be that $\lim_{x\to\infty} x \, \phi_n(a, x) = 0$ for any (n, a), hence $\lim_{x\to\infty} h_n(a, x) \, \phi_n(a, x) = 0$ for any (n, a) as well. Both limit statements are trivial whenever the support of ϕ has a finite upper bound.

Now consider the detrended Kolmogorov forward equation as specified in (A.10). Integrate both sides over the space $\mathcal{N} \times [0, A] \times \mathcal{X}$ and make use of the following identities:

$$\sum_n \iint \phi_n(a, x) \, dx \, da = 1 - \Phi(x_{\text{supp}})$$

$$\sum_n \iint \lambda \sum_{i \neq n} \mu_{in}(a, x) \phi_i(a, x) \, dx \, da = \sum_n \iint \lambda [1 - \mu_{nn}(a, x)] \phi_n(a, x) \, dx \, da$$

$$\sum_n \iint \partial_a \phi_n(a,x)\, da\, dx = 0,$$

where the first follows because ϕ is a probability density; the second, because there is no net migration on a BGP; the third, because the marginal distribution of age is uniform. Now isolating γ yields

$$\gamma = \frac{\sum_n \iint \partial_x [h_n(a,x)\, \phi_n(a,x)]\, dx\, da}{1 - \Phi(x_{\text{supp}}) + \sum_n \iint x\, \partial_x \phi_n(a,x)\, dx\, da}.$$

Integrating the numerator with respect to x yields

$$\begin{aligned}\sum_n \iint \partial_x [h_n(a,x)\phi_n(a,x)]\, dx\, da &= \sum_n \int \left[h_n(a,x)\phi_n(a,x)\Big|_{\mathcal{X}} \right] da \\ &= -\sum_n \int h_n(a, x_{\text{supp}})\phi_n(a, x_{\text{supp}})\, da\end{aligned}$$

by the Fundamental Theorem of Calculus. Integrating the second term of the denominator by parts with respect to x yields

$$\begin{aligned}\sum_n \iint x\, \partial_x \phi_n(a,x)\, dx\, da &= \sum_n \int \left[x\phi_n(a,x)\Big|_{\mathcal{X}} - \int \phi_n(a,x)\, dx \right] da \\ &= -\sum_n \int x_{\text{supp}}\phi_n(a, x_{\text{supp}})\, da - [1 - \Phi(x_{\text{supp}})].\end{aligned}$$

Substituting these expressions back into the expression for γ yields

$$\gamma = \frac{\sum_n \int_0^A h_n(a, x_{\text{supp}})\, \phi_n(a, x_{\text{supp}})\, da}{\sum_n \int_0^A x_{\text{supp}}\, \phi_n(a, x_{\text{supp}})\, da}.$$

Since $x_{\text{supp}} \in \text{supp}(\phi)$ was arbitrary, this proves the claim. $\square$

B Data

B.1 Decennial Census and American Community Survey

Unless otherwise specified, data at the level of Metropolitan Statistical Areas (MSAs) use the February 2013 OMB definitions of MSAs.

2010 Decennial Census

From the 2010 Decennial Census, I collect the population count of each MSA. I use 2010 population counts only to group cities together, as described in the main text.

2010 ACS 1-year sample

From the 2010 ACS 1-year sample, I compute the share of all 15–19 year olds living in each MSA, with MSAs defined according to the December 2009 OMB definitions. These data are downloaded from IPUMS (Ruggles et al., 2022).

2011-15 ACS 5-year sample

From the 2011–15 ACS 5-year sample, I compute empirical c.d.f. of age and wages within each MSA among workers who are employed full-time, which I define by (i) being employed, (ii) working at least 35 hours per week, and (iii) working at least 40 weeks per year. I also compute the college share among those employed full-time in each MSA, where I define college as completing at least four years of post-secondary education. These data are downloaded from IPUMS (Ruggles et al., 2022).

2011–15 ACS 5-year Migration Flow files

The ACS 5-year Migration Flow files comprise period estimates that measure where people lived when surveyed ("current residence") and where they lived one year prior ("residence one year ago"). The data are collected continuously over a five-year period in order to provide a large enough sample for estimates in smaller geographies. The flow estimates represent the annual number of movers between MSAs for the five-year period over which data was collected.

The 2011–15 release is cross-tabulated by demographic characteristics: age, sex, race, and Hispanic origin. I only use the information on age. Age is broken down into the following fifteen categories: 1 to 4 years, 5 to 17 years, 18 to 19 years, 20 to 24 years, 25 to 29 years, 30 to 34 years, 35 to 39 years, 40 to 44 years, 45 to 49 years, 50 to 54 years, 55 to 59 years, 60 to 64 years, 65 to 69 years, 70 to 74 years, and 75 years and over. To be consistent with the rest of my quantitative procedure, I restrict my attention to flows of people between age 20 and age 59.

B.2 BEA Regional Accounts

From the BEA Regional Accounts, I collect the 2013 local area personal income files at the county level and aggregate them to the MSA level. My primary measure of city size is counts of wage and salary jobs. My primary measure of city output is wages and salaries in thousands of current U.S. dollars. My results are robust to using broader measures, like total population for city size and gross metro product for city output. Data on gross metro product are collected directly at the MSA level.

C Computation

C.1 Set up

A balanced growth path of the economy can be represented by the following system of equations that I aim to solve numerically:

$$\begin{aligned}(\rho-\gamma)v_n(a,x) = {} & B_n[T_n L_n^{\alpha}(1-\sigma) - p_n L_n^{\theta_n}]x + [\partial_x v_n(a,x)]d_n(a,x) + \partial_a v_n(a,x) \\ & + \lambda \sum\nolimits_i \mu_{ni}(a,x)\,[\xi_{ni}(a,x)v_i(a,x) - v_n(a,x)]\end{aligned} \tag{C.1}$$

$$\begin{aligned}0 = {} & -\partial_x[d_n(a,x)\phi_n(a,x)] - \partial_a \phi_n(a,x) \\ & - \lambda\phi_n(a,x) + \lambda \sum\nolimits_i \mu_{ni}(a,x)\phi_i(a,x)\end{aligned} \tag{C.2}$$

$$X_n = \left(L \iint x^{\zeta}\, \phi_n(a,x)\, dx\, da \right)^{\frac{1}{\zeta}}, \quad \forall n \in \mathcal{N} \tag{C.3}$$

$$L_n = L \iint \phi_n(a,x)\, dx\, da, \quad \forall n \in \mathcal{N} \tag{C.4}$$

$$\gamma = \frac{\sum_n \int_0^A [h_n(a,x)\,\phi_n(a,x)]\, da}{\sum_n \int_0^A [x\,\phi_n(a,x)]\, da}, \quad \forall x \in \operatorname{supp}(\phi) \tag{C.5}$$

with terminal condition $v_n(A,x) = 0$ and initial condition $\phi_n(0,x) = \frac{1}{A}\underline{\phi}_n(x)$ for any (n,x), where

$$d_n(a,x) = h_n(a,x) - \gamma x$$
$$h_n(a,x) = \kappa[\sigma_n(a,x)]x^{\beta} X_n^{1-\beta}$$

are optimal drift and skill accumulation, respectively, with optimal time spent learning $\sigma_n(a,x)$ the solution to

$$\kappa'(\sigma) = \frac{B_n T_n L_n^{\alpha}}{\partial_x v_n(a,x)} \left(\frac{x}{X_n}\right)^{1-\beta}, \tag{C.6}$$

and where

$$\mu_{ni}(a,x) = \frac{\tau_{ni}^{-\epsilon} v_i(a,x)^{\epsilon}}{\sum_k \tau_{nk}^{-\epsilon} v_k(a,x)^{\epsilon}}$$

$$\xi_{ni}(a,x) = \frac{\mu_{ni}(a,x)^{-\frac{1}{\epsilon}}}{\tau_{ni}}$$

are the optimal migration shares and selection bonuses, respectively. Relative to the systems of equations presented in Achdou et al. (2022, Num. Appx.), the new components here are the endogeneity of the discrete state transition, finite lives, and growth. Ben Moll describes how to incorporate deterministic finite lives in this note.

C.2 Overview of algorithm

Below is a copy of the overview shown in Section 3.1.

Data: parameters $\{A, \rho, \lambda, \epsilon, \alpha, \zeta, \beta, \eta, \bar{\kappa}, x_{\text{scale}}\}$,
location fundamentals $\{B_n, T_n, p_n, \theta_n, \tau_{ni}\}$, and
relative productivity distribution of entrants $\underline{\phi}$

Result: BGP $\{v, \sigma, \mu, \phi; \gamma\}$

Initialize with guess $\{\gamma^0, \mathbf{L}^0, \mathbf{X}^0\}$. Denote iterations by $\ell = 0, 1, 2, \ldots$;

while *not converged* **do**

Given $\{\gamma^\ell, \mathbf{L}^\ell, \mathbf{X}^\ell\}$, solve the Hamilton-Jacobi-Bellman equation (12) using a finite difference method and calculate the policy functions $\sigma_n^\ell(a,x)$ and $\mu_{ni}^\ell(a,x)$; **1**

Given $\sigma_n^\ell(a,x)$ and $\mu_{ni}^\ell(a,x)$, solve the Kolmogorov forward equation (13) for $\phi_n^\ell(a,x)$ using a finite difference method; **2**

Given $\phi_n^\ell(a,x)$, compute the corresponding local populations and vibrancies **3**

$$\tilde{L}_n^\ell = L \iint \phi_n^\ell(a,x)\, dx\, da, \qquad \tilde{X}_n^\ell = \left(L \iint x^\zeta \phi_n^\ell(a,x)\, dx\, da \right)^{\frac{1}{\zeta}}$$

for each n, and compute the implied growth rate

$$\tilde{\gamma}^\iota = \frac{\sum_n \int_0^A \kappa[\sigma_n^\iota(a,x)] x^\beta (X_n^\iota)^{1-\beta}\, \phi_n^\iota(a,x)\, da}{\sum_n \int_0^A x\, \phi_n^\iota(a,x)\, da}, \quad x \in \text{supp}(\phi^\iota)$$

as defined in (16);

if $\{\tilde{\gamma}^\ell, \tilde{\mathbf{L}}^\ell, \tilde{\mathbf{X}}^\ell\}$ *close enough to* $\{\gamma^\ell, \mathbf{L}^\ell, \mathbf{X}^\ell\}$ **then**

converged;

else

construct $\{\gamma^{\ell+1}, \mathbf{L}^{\ell+1}, \mathbf{X}^{\ell+1}\}$ as a linear combination of the previous guess and the computed values;

end

end

Note that in place of iterating on the vector $\{\gamma^\ell, \mathbf{L}^\ell, \mathbf{X}^\ell\}$ toward a fixed point, one could instead solve for the fixed point using any of a number of nonlinear solvers. See Judd (1998, Ch.5) for a review of the underlying algorithms.

C.3 Solving the Hamilton-Jacobi-Bellman equation

To solve the Hamilton-Jacobi-Bellman equation, I use a finite difference method: the *semi-implicit upwind scheme.* I discretize the state space, denoting the grid points by x_i, $i = 1, \ldots, I$ and a_j, $j = 1, \ldots, J$ with $\Delta x = x_{i+1} - x_i$ and $\Delta a = a_{j+1} - a_j$ for all i, j. I approximate the functions (v, ϕ) at each of these $N \times I \times J$ grid points and denote

$$v_{i,n}^j = v_n(x_i, a_j) \quad \text{and} \quad \phi_{i,n}^j = \phi_n(x_i, a_j).$$

I approximate the derivatives with respect to a and x with either a forward- or backward-difference approximation:

$$\partial_x^F v_{i,n}^j := \frac{v_{i+1,n}^j - v_{i,n}^j}{\Delta x},$$

$$\partial_x^B v_{i,n}^j := \frac{v_{i,n}^j - v_{i-1,n}^j}{\Delta x},$$

$$\partial_a^F v_{i,n}^j := \frac{v_{i,n}^{j+1} - v_{i,n}^j}{\Delta a},$$

$$\partial_a^B v_{i,n}^j := \frac{v_{i,n}^j - v_{i,n}^{j-1}}{\Delta a}.$$

An *upwind* scheme means that one approximates the derivative $\partial_x v_n(x_i, a_j)$ by $\partial_x^F v_{i,n}^j$ whenever the drift of the state variable x is **positive** and by $\partial_x^B v_{i,n}^j$ whenever it is **negative**. Because age only drifts forward, I always approximate $\partial_a v_n(a_i, x_j)$ by $\partial_a^F v_{i,n}^j$. The qualifier *semi-implicit* just describes the method of iterating on the value function.

Let $\mathbf{v}^j$ be the $N \times I$ column vector whose components are the values for both discretized age-a_j value functions. That is,

$$\mathbf{v}^j = [v_{1,1}^j, v_{2,1}^j, \ldots, v_{I,1}^j, v_{1,2}^j, \ldots, v_{I,2}^j, \ldots, v_{1,N}^j, \ldots, v_{I,N}^j]'.$$

Define the $N \times I$ vector u^j similarly. Then, for a given set $\{\gamma, \mathbf{L}, \mathbf{X}\}$, the value function can be found by solving

$$(\rho - \gamma)\mathbf{v}^j = u^{j+1} + \tilde{\mathbf{\Pi}}^{j+1}\mathbf{v}^j + \frac{\mathbf{v}^{j+1} - \mathbf{v}^j}{\Delta a} \tag{C.7}$$

by iterating backwards from the terminal condition $\mathbf{v}^J = \mathbf{0}$, where $\tilde{\mathbf{\Pi}}^j = \mathbf{\Pi}^j \odot \mathbf{\Xi}^j$ is the Hadamard product between the *transition matrix* $\mathbf{\Pi}^j$, defined by

$$\mathbf{\Pi}^j = \begin{bmatrix} \boldsymbol{\pi}_1^j & \mathbf{M}_{1,2}^j & \mathbf{M}_{1,3}^j & \cdots & \mathbf{M}_{1,N}^j \\ \mathbf{M}_{2,1}^j & \boldsymbol{\pi}_2^j & \mathbf{M}_{2,3}^j & \cdots & \mathbf{M}_{2,N}^j \\ \vdots & \mathbf{M}_{3,2}^j & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & \mathbf{M}_{N-1,N}^j \\ \mathbf{M}_{N,1}^j & \mathbf{M}_{N,2}^j & \cdots & \mathbf{M}_{N,N-1}^j & \boldsymbol{\pi}_N^j \end{bmatrix},$$

$$\mathbf{M}_{kn}^j = \begin{bmatrix} \lambda\mu_{1,kn}^j & 0 & \cdots & \cdots & \cdots & 0 \\ 0 & \lambda\mu_{2,kn}^j & 0 & \ddots & \ddots & \vdots \\ \vdots & 0 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & \cdots & \cdots & 0 & \lambda\mu_{I,kn}^j \end{bmatrix}$$

$$\boldsymbol{\pi}_n^j = \begin{bmatrix} \pi_{1,1}^j & \pi_{1,1}^{F,j} & 0 & \cdots & \cdots & 0 \\ \pi_{2,1}^{B,j} & \pi_{2,1}^j & \pi_{2,1}^{F,j} & 0 & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & 0 & \pi_{I-1,1}^{B,j} & \pi_{I-1,1}^j & \pi_{I-1,1}^{F,j} \\ 0 & \cdots & \cdots & 0 & \pi_{I,1}^{B,j} & \pi_{I,1}^j \end{bmatrix}$$

with the entries of $\boldsymbol{\pi}_n^j$ given by

$$\begin{aligned}
\pi_{i,n}^{B,j} &= -\frac{(d_{i,n}^{B,j})^-}{\Delta x} \\
\pi_{i,n}^{F,j} &= \frac{(d_{i,n}^{F,j})^+}{\Delta x} \\
\pi_{i,n}^{j} &= -\pi_{i,n}^{F,j} + \pi_{i,n}^{B,j} - \lambda(1-\mu_{i,nn}^j),
\end{aligned}$$

and the *selection matrix* $\boldsymbol{\Xi}^j$, given by

$$\boldsymbol{\Xi}^j = \begin{bmatrix} \boldsymbol{\xi}_1^j & \boldsymbol{\xi}_{1,2}^j & \boldsymbol{\xi}_{1,3}^j & \cdots & \boldsymbol{\xi}_{1,N}^j \\ \boldsymbol{\xi}_{2,1}^j & \boldsymbol{\xi}_2^j & \boldsymbol{\xi}_{2,3}^j & \cdots & \boldsymbol{\xi}_{2,N}^j \\ \vdots & \boldsymbol{\xi}_{3,2}^j & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & \boldsymbol{\xi}_{N-1,N}^j \\ \boldsymbol{\xi}_{N,1}^j & \boldsymbol{\xi}_{N,2}^j & \cdots & \boldsymbol{\xi}_{N,N-1}^j & \boldsymbol{\xi}_N^j \end{bmatrix}, \qquad \boldsymbol{\xi}_{kn}^j = \begin{bmatrix} \xi_{1,kn}^j & 0 & \cdots & \cdots & \cdots & 0 \\ 0 & \xi_{2,kn}^j & 0 & \ddots & \ddots & \vdots \\ \vdots & 0 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & \cdots & \cdots & 0 & \xi_{I,kn}^j \end{bmatrix}, \quad k \neq n$$

$$\boldsymbol{\xi}_n^j = \begin{bmatrix} \frac{\pi_{1,n}^j + \lambda(\xi_{1,nn}^j - 1)\mu_{1,nn}^j}{\pi_{1,n}^j} & 0 & \cdots & & 0 \\ 0 & & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \ddots & 0 \\ 0 & & \cdots & 0 & \frac{\pi_{I,n}^j + \lambda(\xi_{I,nn}^j - 1)\mu_{I,nn}^j}{\pi_{I,n}^j} \end{bmatrix}$$

The transition matrices $\boldsymbol{\Pi}^j$, $j = 1, 2, \ldots, J$ encode the evolution of the stochastic process (a_t, x_t, n_t). Note, in particular, that $\boldsymbol{\Pi}^j$ satisfies all the usual properties of a Poisson transition matrix: all rows sum to zero, all diagonal elements are non-positive, and all off-diagonal elements are non-negative.

Returning to (C.7), note that it can be recast so that $\mathbf{v}^j$ appears as the solution to a matrix equation:

$$\begin{aligned}
\mathbf{B}^{j+1}\mathbf{v}^j = \mathbf{b}^{j+1}, \quad \text{where} \quad \mathbf{B}^{j+1} &= \left(\frac{1}{\Delta a} + \rho - \gamma\right)\mathbf{I} - \tilde{\boldsymbol{\Pi}}^{j+1}, \\
\mathbf{b}^{j+1} &= \mathbf{u}^{j+1} + \frac{1}{\Delta a}\mathbf{v}^{j+1}.
\end{aligned} \tag{C.8}$$

This matrix equation can be solved efficiently with sparse matrix routines.

Let us summarize the sub-algorithm to solve the Hamilton-Jacobi-Bellman equation. For a given set $\{\gamma, \mathbf{L}, \mathbf{X}\}$, iterate backwards from the terminal conditions $\mathbf{v}^J = \mathbf{0}$ and $\delta_x^F v_{i,n}^J = 0$ as follows:

1. compute σ^{j+1} from

$$\sigma_{i,n}^{j+1} = \min\left\{1, \max\left\{\left[\left(\frac{\bar{\kappa}\partial_x v_{i,n}^{j+1}}{B_n T_n L_n^\alpha}\right)\left(\frac{X_n}{x}\right)^{1-\beta}\right]^{\frac{1}{1-\eta}}, 0\right\}\right\}$$

and μ^{j+1} from

$$\mu_{i,kn}^{j+1} = \frac{\tau_{kn}^{-\epsilon}(v_{i,n}^{j+1})^{\epsilon}}{\sum_{n'} \tau_{kn'}^{-\epsilon}(v_{i,n'}^{j+1})^{\epsilon}};$$

2. compute $\mathbf{u}^{j+1}$ and $\tilde{\mathbf{\Pi}}^{j+1}$;
3. find $\mathbf{v}^j$ from $\mathbf{B}^{j+1}\mathbf{v}^j = b^{j+1}$;
4. stop once you compute $\mathbf{v}^1$.

C.4 Solving the Kolmogorov forward equation

I now turn to the solution of (C.2) subject to (C.4). I discretize these as follows:

$$0 = -[d_{i,n}^j \phi_{i,n}^j]' - \lambda(1-\mu_{i,n}^j)\phi_{i,n}^j + \lambda \sum_{n' \neq n} \mu_{i,n}^j \phi_{i,n'}^j - \frac{\phi_{i,n}^{j+1} - \phi_{i,n}^j}{\Delta a} \tag{C.9}$$

$$\frac{1}{A} = \sum_n \sum_i \phi_{i,n}^j \Delta x. \tag{C.10}$$

I will approximate the derivative $-[d_{i,n}\phi_{i,n}]'$ by

$$-\frac{(d_{i,n}^{F,j})^+ \phi_{i,n}^j - (d_{i-1,n}^{F,j})^+ \phi_{i-1,n}^j}{\Delta x} - \frac{(d_{i+1,n}^{B,j})^- \phi_{i+1,n}^j - (d_{i,n}^{B,j})^- \phi_{i,n}^j}{\Delta x}$$

where $(d_{i,n}^F)^+$ is the forward approximation for skill accumulation that comes from the final step of the HJB iterative procedure. It follows that I can approximate the Kolmogorov forward equation by

$$0 = \pi_{i+1,n}^{B,j}\phi_{i+1,n}^j + \pi_{i,n}^j \phi_{i,n}^j + \pi_{i-1,n}^{F,j}\phi_{i-1,n}^j + \lambda \sum_i \mu_{i,n}\phi_{i,m}^j + \frac{\phi_{i,n}^{j+1} - \phi_{i,n}^j}{\Delta a},$$

which can be written in matrix notation as

$$\mathbf{0} = \left(\mathbf{\Pi}^j\right)^T \boldsymbol{\phi}^j - \frac{\boldsymbol{\phi}^{j+1} - \boldsymbol{\phi}^j}{\Delta a} \tag{C.11}$$

where

$$\boldsymbol{\phi}^j = \left[\phi_{1,1}^j, \dots, \phi_{I,1}^j, \phi_{1,2}^j, \dots, \phi_{I,2}^j, \dots, \phi_{1,N}^j, \dots, \phi_{I,N}^j\right]'$$

Note that (C.11) can be solved either explicitly,

$$\boldsymbol{\phi}^{j+1} = \Delta a \left(\mathbf{\Pi}^j\right)^T \boldsymbol{\phi}^j + \boldsymbol{\phi}^j,$$

or implicitly,

$$\boldsymbol{\phi}^{j+1} = \left(\mathbf{I} - \Delta a \left(\mathbf{\Pi}^j\right)^T\right)^{-1} \boldsymbol{\phi}^j,$$

from the initial condition $\boldsymbol{\phi}^1 = \underline{\boldsymbol{\phi}}$. Note that I can use $\mathbf{\Pi}^j$ to solve for $\boldsymbol{\phi}^{j+1}$ because I've already solved for it when I did the value function iteration.

Note, too, that these schemes preserve mass: starting from any initial distribution $\boldsymbol{\phi}^1$ that sums to one, all future $\boldsymbol{\phi}^j$'s also sum to one. This follows from the fact that the rows of the transition matrices $\mathbf{\Pi}^j$ sum to zero. The implicit scheme is also guaranteed to preserve the positivity of $\boldsymbol{\phi}^j$ for arbitrary time steps Δa.