A Dynamic Model of Skill Formation and Worker Migration

Sunham Kim*
Purdue University

November 4, 2023
Job Market Paper

Please click here for the latest draft.

Abstract

Human capital is a key determinant of regional economic development. In the U.S., the varying skill production efficiencies of state-specific education systems and the dynamics of worker migration shape regional human capital, influencing economic outcomes at both the state and national levels. This paper develops a novel dynamic spatial general equilibrium model with overlapping generation framework in which heterogeneous individuals accumulate human capital and move across regions. Calibrated to U.S., the model illustrates how variations in education efficiency lead to substantial cross-state income disparities and shows that internal migration can notably boost output in states with lower education efficiencies such as North Dakota and Oklahoma. At the national level, free mobility of workers yields a 6.9% output gain. Moreover, the model suggests that variations in human capital account for 46.6% of the state variation per capita output. Applying the calibrated model to analyze the Obama Administration's Race to the Top initiative finds that the $4.1 billion grant spurred a 0.2% increase in U.S. GDP, mostly benefiting the grant-winning states and their neighbors. Additionally, alternative grant allocation experiments show that strategic reallocation of education grants, considering regional skill production efficiencies, could further enhance national GDP gains without necessarily worsening regional income disparities.

*All errors are my own. Contact: sunham-kim@purdue.edu

1. Introduction

Human capital serves as the cornerstone of regional economic development. (Lucas 1988; Hsieh and Klenow 2010) Education is the primary engine that generates this resource, while migration steers its spatial allocation, shaping the distribution of skills and knowledge across regions. Such dynamics are particularly pronounced in the U.S., which features free internal migration and a decentralized education system characterized by state-led policymaking and funding schemes.

Existing empirical evidence suggests that regions with higher education quality tend to produce better educational outcomes, influencing migratory decisions. Workers are drawn to areas offering higher wages, while parents also consider the quality of education, making human capital's reallocation a multifaceted decision. However, few models of human capital and economic development have yet to incorporate the interplay of endogenous skill formation and migration dynamics. It leaves a crucial question unresolved: How does a state's skill production efficiency influence its own human capital and extend its effects to neighboring states, thereby shaping the national economic landscape? This gap has also constrained the evaluation of regional education policies, leaving their spatial general equilibrium effects largely uncharted.

This paper addresses those concerns by developing a unified framework of a dynamic spatial general equilibrium model with overlapping generations of workers. In so doing, I make several contributions. A primary contribution is the elucidation of a mechanism where endogenous skill formation, coupled with migration choices, determines the spatial distribution of human capital, and thus regional economic outcomes. Second, my model quantifies the flux of human capital across spaces embodied in migration flows. Third, employing the model-inferred human capital stock, I decompose the regional income differences into contributions from differences of human capital stock and regional total factor productivity (TFP), while incorporating all previously mentioned dynamics. Fourth, leveraging the calibrated model, I evaluate the outcome of a federal grant that subsidizes state education. Finally, this paper introduces a novel, efficient algorithm to tackle the computational complexity of the model.

In the model, a finite number of states are characterized by efficiencies in producing final good and producing skills. The states are postulated by heterogeneous workers who make forward-looking human capital accumulation and migration decisions. Central to the model is the interplay between two channels: for instance, when workers make human capital investment decisions, they internalize the future possibility of migrating to high-wage states. Concurrently, these anticipated migration choices are influenced by each state's efficiency in skill production, reflecting workers' considerations for their expected offspring. In equilibrium, the markets for the final good and human capital clear in each state at all times, and while workers may migrate across states, the overall net flows balance out in the steady state.

The model addresses a computationally intensive, high-dimensional problem. It must consider various cohorts of workers with innate abilities, levels of education, and skill investments. The model tracks both states of birth and current residences to quantify the flow of human capital between regions. Additionally, it characterizes migration propensities for each potential pair of origin and destination states for different worker types. Although the framework can be applied to a range of geographical units, this analysis is limited to U.S. states and internal migration to manage the computational demands, which increase quadratically with the number of regions.

I quantify the model with data on U.S. states. First, I estimate the moving costs and migration elasticity by leveraging average regional wages and the 5-year cross-state migration flows from the Current Population Survey.

Second, I calibrate the regional fundamentals and parameters influencing the education decision to align with data moments such as population share, wage, and per capita output, ensuring all labor markets clear.

I find substantial variation in skill production efficiency across U.S. states, with a ratio of 1.23 between the 90th and 10th percentiles, indicating that the most efficient states are 23% more efficient than the least. Internal migration plays a crucial role in mitigating this disparity, reallocating human capital to optimize national output. Absent this mobility, the model estimates that the national GDP would be reduced by 6.9%, equating to a loss of $700 billion in 2000 dollars. This estimate stands as the first attempt to quantify the economic contributions of the state-level public goods in human-capital production, highlighting the decentralized U.S. education system's benefit from high internal mobility. It suggests that the decline in U.S. internal migration over recent decades could have long-run implications for future U.S. economic growth, via human capital production, deserves further investigation.

On the state level, this efficiency variance creates distinct winners and losers due to migration dynamics. Less efficient states benefit from importing human capital, which can bolster regional GDP per capita by up to 7%. In contrast, the more efficient states may experience up to a 3.8% decrease in GDP per capita due to a net outflow of human capital. These divergent outcomes illuminate a tradeoff between equity and efficiency in economic development, leading to an investigation into the potential impacts of place-based educational policies.

Using the calibrated model, I examine the impacts of Race to the Top (RTT) grant of Obama administration on the regional and national economies. The Department of Education distributed $4.1 billion in competitive grants to states that elected to participate, with the amounts awarded varying among the 18 states and the District of Columbia that received funds. To gauge the program's effects, I simulate a counterfactual scenario wherein the RTT grants are translated into enhancements in skill production efficiency. The results indicate that the RTT initiative augmented U.S. output by 0.2%, which corresponds to an increase of $21.5 billion, marking the first rigorous cost-benefit analysis of the program. The investment appears to yield high returns, partly due to the subsidy's spillover effects to neighboring states and the temporal amplification of benefits. Specifically, output per capita in grant-receiving states rose between 0.1% in Illinois and 0.92% in Tennessee, linked to the per-pupil size of the grant. Furthermore, adjacent states experienced output gains as well, albeit smaller, ranging from 0.03% in Oregon to 0.14% in Vermont.

Finally, in exploring alternative RTT grant allocation strategies, I find that the distribution methodology can significantly influence economic outcomes at both regional and national levels. Consider a hypothetical redistribution of the RTT funds exclusively to the five states with the lowest education quality: North Dakota, Oklahoma, New Mexico, Montana, and Mississippi. This allocation would result in a 0.13% boost to U.S. GDP—less than the increase observed under the actual RTT program's implementation—but would also narrow the income disparities among states. For instance, Oklahoma's per capita output ratio compared to New York, the wealthiest state, would rise from 0.56 to 0.65.

Conversely, directing the RTT grants to the top five states in terms of education quality—New Jersey, Nevada, Washington, New York, and Connecticut—would enhance U.S. GDP by 0.27%. This increase exceeds the gains from both the original RTT implementation and the alternative scenario focusing on the bottom states. Remarkably, this approach does not exacerbate income inequality; the output ratio of Oklahoma to New York increases slightly to 0.58. This can be attributed to the fact that subsidizing states with already high education quality could induce spillover benefits through migration. With these states strategically positioned—three on the

East Coast and two on the West—the potential for spillover effects is considerable, extending through neighboring regions and infusing the broader U.S. economy. This result illustrates that targeted investments in states with already high education quality can yield widespread economic benefits without necessarily deepening disparities. It also underscores the importance of geographical considerations in the distribution of federal education grants, as the spatial dynamics play a crucial role in amplifying the impact on the overall U.S. economic output.

Related Literature This paper relates to several veins of literature. Recent studies have advanced dynamic quantitative spatial general equilibrium models in international trade and economic geography, addressing labor market dynamics. Notable works include Artuç et al. (2010), Desmet et al. (2018), Caliendo et al. (2019), Fogli and Guerrieri (2019), Lyon and Waugh (2019), Giannone et al. (2023), Ferriere et al. (2021), Hsiao (2021), and Crews (2023). This paper extends these models by integrating worker skill heterogeneity and regional skill production differences, leading to endogenous skill acquisition dynamics. While it shares commonalities with Ferriere et al. (2021), which examines skill acquisition in response to trade shocks, this study centers on the spatial distribution of human capital and the subsequent impact on regional economic development. In contemporaneous work,Eckert and Kleineberg (2021) proposes a model in which residential and educational decisions result in geographic sorting; however, their mechanism of education choice is driven by taste shocks—a stark contrast to this paper's approach where high-skilled individuals' sorting towards higher education is explicitly modeled. Similarly, Hsiao (2021) explores education and migration within Indonesia using a spatial framework, yet this work differentiates itself by focusing on the U.S. and presenting a richer model of skill formation decisions. Building on Hsieh et al. (2019) and Bryan and Morten (2019), and situated within an extensive literature on selection, sorting, and migration (Roy (1951),Hanson and Liu (2021)), this study probes the interactions between mobility and skill formation, their consequences for regional economic development, and the implications for regional education policy at both local and national levels.[1]

This paper also relates to the literature that studies the U.S. spatial dispersion, such as Berry and Glaeser (2005), Diamond (2016), Ganong and Shoag (2017), Fajgelbaum and Gaubert (2020), and Rossi-Hansberg et al. (2019). (2019). My paper contributes to this literature by introducing a coupled consideration of skill acquisition and migratory choices, thereby enabling the quantification of regional human capital stock interstate human capital flows. I leverage the model-inferred labor stock to elucidate the role of human capital in explaining per capita income variation across states. Additionally, I study the impact of regional education policies on spatial dispersion, accounting for the general equilibrium forces. This analysis highlights how education policy could potentially serve as a tool to reduce spatial income dispersion over the long term.

In the domain of development accounting that investigates the role of human capital in income disparities, this paper closely aligns with the approach of Hanushek et al. (2017) who introduced new measures of human capital to account for cross-state migrations in the U.S. states. However, their work, based on data-driven measurement and accounting, differs from my research, which employs a micro-founded framework producing a consistent measure of human capital stocks and flows. This framework also incorporates multiple amplification mechanisms, and as a result, human capital explains 46.6% of cross-state income variations, nearly double the estimates

[1] Bertoli and Rapoport (2015) and Delogu et al. (2018) are other papers in the similar spirit that endogenize education and migration choices. While my paper exclusively focuses on internal migration in the U.S., both are in the context of international migration. A two-region model of Bertoli and Rapoport (2015) focuses on the role of migration networks. Delogu et al. (2018) study the impact of immigration restrictions such as visa policy on the world distribution of income.

provided by Hanushek et al. (2017). In conjunction with other macro-level studies on skill accumulation and income disparities such as those of Erosa et al. (2010), Manuelli and Seshadri (2014), Cubas et al. (2016), and Xiang and Yeaple (2021), my model adds endogenous migration as a novel amplification mechanism. This enhancement not only complements these models but also bridges the theoretical gap between accounting frameworks and empirical migration data through its accounting for the role of labor mobility in human capital accumulation. In this context, work of Klein and Ventura (2007, 2009) features a two-region model that similarly employs amplification and labor mobility in cross-country development accounting. While our frameworks share similarities, two key differences stand out: their model primarily emphasizes endogenous physical capital accumulation, whereas mine focuses on human capital; and my quantitative spatial model, capable of incorporating numerous regions, offers the flexibility to extend the model's applications well beyond development accounting.[2]

The empirical foundation of my model is anchored in two pivotal insights from the literature on migration incentives. First, the expected income is a primary determinant of migration decisions. High-skilled workers are more inclined to migrate and to make longer-distance moves (Borjas et al. 1992; Dahl 2002; Kennan and Walker 2011; Grogger and Hanson 2011; Young 2013; Amior 2019). This selective migration pattern also informs residential choices, with parents seeking neighborhoods that offer superior educational opportunities, thereby investing in their children's human capital, as analyzed by Bayer and Timmins (2007) and Nechyba (2006). Secondly, my model integrates findings on the enduring impact of neighborhood and educational

[2]There are many studies that use immigrant data in development accounting, such as Hendricks (2002), Schoellman (2012), Schoellman (2016), Hendricks and Schoellman (2018), Rossi (2022). Their goal is to disentangle the worker-embodied human capital (supply) and skill bias of technology (demand) using immigrants. None of amplification, migration, and interaction of the two are their interest.

quality on economic outcomes, highlighted in studies by Chetty et al. (2016) and Jackson et al. (2016). These impacts are incorporated through skill production efficiency, a core parameter that I calibrate.

2. Model

2.1 Environment

I consider an economy composed of multiple regions indexed by $k \in \mathcal{K}$. The economy is populated by a continuum of overlapping generations of workers who are altruistic towards their children. Each worker has one and only one child throughout their life, thus each generation has the identical population size. Workers possess different levels of human capital, which is characterized by skill and education. Skill is characterized by innate abilities ε and acquired knowledge H, which is a function of region of birth k_B, skill investment e. There are two education levels, **c**ollege and **n**on-college, indexed by $o \in \{c, n\}$. College education lets worker access to high-paying jobs that requires a college diploma. It however does not change one's skill. Individuals have to pay fixed cost χ and idiosyncratic cost z of cost, in terms of utility. Regions differ in four dimensions: amenities a^k, labor demand shifter A_o^k, the final good TFP Θ^k and the efficiency in skill production h^k. In each region, a representative firm produces non-tradable final good. All firms have a CES technology and demand labor from all education levels. Time is discrete and indexed by $t \in \{0, 1, 2, \cdots\}$.

2.2 Equilibrium

2.2.1 Households

There is a measure-one continuum of finitely-lived workers. Age is indexed by $q = \{1, 2, \dots, q_R\}$. Workers are endowed with an innate ability ε, of which distribution $F(\varepsilon)$ has a strictly positive support. They work for q_R periods and have a child at age $q_P < q_R$. Fertility is exogenous and deterministic thus one adult gives a birth to one child. Workers derive utility from consumption of the final good. Parents care about their children's lifetime utility as well, hence finance their skill investment.

Workers can move across regions every period. They receive region-specific taste shocks $u^{k'}$ at the end of each period and choose where to reside in next period. The taste shocks are iid across time and workers. Relocation from k to k' incurs education-specific moving cost $\tau^o_{kk'} \geq 0$ measured in utility, where the equality holds only if $k = k'$. Following ACM and Caliendo et al. (2019), I assume u is a vector of $\mathcal{K}$ independent shocks that follow Type I Extreme value distribution with zero mean and scale parameter ν. The taste shocks u and abilities draw ε are independent.

Workers' human capital is a combination of skill and education. Skill is again constituted by innate ability ε and acquired knowledge H. They draw innate abilities at birth from $\mathcal{LN}(0, \sigma_\varepsilon^2)$. It is not genetically inherited, i.e., the ability of a child and her parents are not correlated.[3] Parents Instead affect children's human capital through residential and skill investment decisions: parent's region of residence is children's region of education; parents finance children's skill investment by their income. They cannot borrow to finance for financing. A child with $q = 0$ is in the "education stage", and invest in skill, and choose whether or not to go to college, given the innate ability, region

[3]The "genetic inheritance" assumption is popular in the papers that use the overlapping generation setup to study intergenerational inequality, such as Erosa et al. (2010), Fogli and Guerrieri (2019), Ferriere et al. (2021), etc.

of birth, and parental transfer for education. Human capital is fixed over lifetime after the education stage.

The following skill production function, translates skill investment e to acquired knowledge for an individual born in region k_B.[4]

$$H(k_B; e) = h^{k_B} e^{\eta}. \tag{1}$$

In a broad sense, h^k is a parameter that captures educational quality of region k. The parameter $\eta \in (0, 1)$ is the elasticity of skill with respect to investment, which would capture diminishing returns. In sum, one's human capital stock is a product of innate ability and acquired human capital, and it is the very amount of effective labor supplied by the worker. Workers earn a market wage for each supplied unit of effective labor, thus labor income is a product of the market wage and effective labor:

$$I^k(k_B, \varepsilon, e, o) = w_o^k H\varepsilon = w_o^k (h^{k_B} e^{\eta})\varepsilon, \tag{2}$$

where o is the education level.

Households are forward-looking and have subjective discount rate $\beta \in (0, 1)$. The timeline for the household's problem and decisions is as follows:

1. Children arrive and their abilities vector ε' realizes. All households in region k know the economic condition of the region;
2. All adults work and earn the market wage;
3. Observing ε', parents choose their own consumption and children's skill investment. Children choose education level. Non-parent adults enjoy consumption;

[4]The log-linear specification follows HHJK, XY, and Erosa et al. (2010).

4. The region-specific taste shock realizes and region choices are made. The terminal age workers retire and exit from the labor market. Newborns enter the labor market as a worker with age 1.

I formalize the household's decision problem at each stage of their life-cycle. I denote the state variables fixed over time by $\mathrm{s} \equiv (k_B, \varepsilon, e, o)$ to ease the notation. Let $v^q(k, \mathrm{s})$ be the lifetime utility of a worker of age q at time t in region k, born in region k_B with ability draw ε, invested e for skill, and chose education o. Also, let $V^q(k, \mathrm{s}) \equiv \mathbb{E}v^q(k; \mathrm{s})$ be the expected lifetime utility where the expectation is taken over the region-specific taste shocks. The utility function $u(c)$ is CRRA, $u(c) = \frac{c^{1-\rho}-1}{1-\rho}$.[5]

Working Stage At the beginning of a period, adult workers observe the economic condition of all regions, work, and earn the market wage. They have the option to relocate then. The worker's problem formulated recursively as follows, where P^k is the final good price in region k.

$$v_t^q(k; \mathrm{s}) = u(c_t^k(\mathrm{s})) + a^k + \max_{k' \in \mathcal{K}} \left\{ u_t^{k'} - \tau_o^{k,k'} + \beta \mathbb{E} v_{t+1}^{q+1}(k'; \mathrm{s}) \right\} \quad (3)$$

subject to $c_t^k(\mathrm{s}) = I_t^k(\mathrm{s}) / P_t^k$.

The first and second terms constitutes the flow utility. The third term is the future value of the market. It depends on the idiosyncratic taste shock $u^{k'}$, education-specific mobility cost $\tau_o^{k,k'}$, and the expected future value function for destination labor market k'. As usual in the discrete choice model literature (ACM, CDP, etc.), the Type I extreme value distribution assumption of

[5]The utility function choice has a nontrivial implication on workers' decision rules on education investment and migration. See Appendix B.2.1.

$u^{k'}$ lets me rewrite (3) as follows:

$$V_t^q(k;\mathbf{s}) = u(c_t^k(\mathbf{s})) + a^k + \nu \log \left[\sum_{k'} \exp\left(\beta V_{t+1}^{q+1}(k',\mathbf{s}) - \tau_o^{k,k'}\right)^{\frac{1}{\nu}} \right] \tag{4}$$

$$s.t. \qquad c_t^k(\mathbf{s}) = I_t^k(\mathbf{s})/P_t^k.$$

I use the V notation for the rest of the paper unless ν is necessary. It delivers a migration probability from k to k' for a worker with (k, s) as

$$m_t^q(k,k';\mathbf{s}) = \frac{\exp\left[\beta V_{t+1}^{q+1}(k';\mathbf{s}) - \tau_o^{k,k'}\right]^{\frac{1}{\nu}}}{\sum_{k'\in\mathcal{K}} \exp\left[\beta V_{t+1}^{q+1}(k';\mathbf{s}) - \tau_o^{k,k'}\right]^{\frac{1}{\nu}}}. \tag{5}$$

The probability depends on one's human capital. I relegate the relationship between human capital and m to Section 2.3.

Education Stage ("pre-period") At $q = 0$, children start in their parent's region and draw innate ability ε. Observing the draw, parents finance (thus choose) the level of skill investment e for children. Children then choose whether to go to college, given (ε, e), subject to the fixed cost of college χ and idiosyncratic cost of college z. The region-specific taste shock realizes at the end of the period, then newborns choose the initial region to join as a worker with $q = 1$. Thus, the newborn's problem consists of two sub-problems. Let $V_t^{0+}(k_P, \varepsilon, e, o)$ be the expected lifecycle utility of a child who draw ability ε, spend e for human capital, born in k_p (p for parents), and choose education level o. Formally, the end-of-period problem is written as follows[6]:

$$V_t^{0+}(k_P, \varepsilon, e, o) = \nu \log \Big[\sum_{k'} \exp \Big(\beta V_{t+1}^1(k', \mathbf{s}) - \tau_o^{k_p, k'} \Big)^{\frac{1}{\nu}} \Big] \tag{6}$$

That is, children face the same region choice problem as the workers, but have no own consumption.[7] In the beginning of the education stage, the value of a child in k_p with (ε, e) who chooses her education level $o \in \{c, n\}$ is given as

$$V_t^{0-}(k_P, \varepsilon, e) = \max_{o \in \{c,n\}} \Big\{ V_t^{0+}(k_P, \varepsilon, e, c) - (\chi + z), V_t^{0+}(k_P, \varepsilon, e, n) \Big\}. \tag{7}$$

The child pays a cost of college $\chi + z$ in terms of utility. The first term χ is a fixed cost common to everyone. The second term z is a random cost, which distribution follows Logistic$(0, \sigma_z^2)$. The optimal education policy o is obtained from solving (7) and the initial labor market choice m^0 is obtained from (6). It determines the measure $\mu_t^{q=1}(k'; \mathrm{s})$ of type s workers in region k' at time t.

The last building blocks of the household's problem are the decision problems for retiring workers ($q = q_R$), parents ($q = q_P$), and the workers who will become parents ($q = q_P - 1$). Those are "special" periods in the working stage.

[6]As in the working stage, before solving for the expectation, the problem is written as

$$v_t^{0+}(k_P, \varepsilon, e, o) = \mathbb{E}\Big[\max_k \Big\{ u^k - \tau^{k_P, k} + \beta v_{t+1}^1(k; \mathbf{s}) \Big\} \Big]$$

[7]This assumption is common in the literature. (e.g. Erosa et al. (2010), Fogli and Guerrieri (2019), Eckert and Kleineberg (2021).) The parental consumption could be interpreted as a household consumption.

Terminal Period In the terminal period, a worker work, earn, consume, and retire:

$$V_t^{q_R}(k;\mathbf{s}) = u(I_t^k(\mathbf{s})/P_t^k) \tag{8}$$

Equation (4) is a simple a hand-to-mouth problem. The retiring workers work, earn, consume, then exit the labor market. They have no relocation choice.

Parental Period The parent's problem connects the children and adults. In the beginning of the parental period, children arrive with their innate ability ε'. Observing ε', parents face a tradeoff between their own consumption c and child's skill investment e'. Formally, they solve the following problem:

$$\begin{aligned} V_t^{q_P}(k,\varepsilon';\mathbf{s}) = \max_{c,e'} u(c_t^k(\mathbf{s})) + a^k + \nu \log \left[\sum_{k'} \exp\left(\beta V_{t+1}^{q_P+1}(k',\mathbf{s}) - \tau_o^{k,k'} \right)^{\frac{1}{\nu}} \right] \\ + \alpha V_t^{0-}(\varepsilon', e', k) \\ s.t. \qquad c_t^k(\mathbf{s}) + (e_t')^k(\mathbf{s}) = I_t^k(\mathbf{s})/P_t^k. \end{aligned} \tag{9}$$

Compared to Equation (4), Equation (9) has the fourth term, which is a value function of child. The parameter α governs the degree of alturism. Notice that the new ability draw ε' enters into the state space of the parents as they fully observe its level. On the other hand, In the meantime, the workers who will become parents make migration decision before knowing ε', as will soon be discussed.

Pre-parental Period Pre-parental adults are the workers who will become a parent in the following period. $(q = q_P - 1)$. They do know they will be a parent, observe child's ability pair ε', and make the skill investment decision in the following period. But they do not know the level of ε' thus take expectation over ε' in contrast to parents. As a result, their future value of the market is different from Equation (4), the problem of non-parent adults whose age is neither q_P nor q_R.

$$V_t^{q_P-1}(k;\mathbf{s}) = u(c) + \nu \log \left[\sum_{k'} \exp \left(\beta \mathbb{E}_{\varepsilon'} V_t^{q_P}(k', \varepsilon', \mathbf{s}) - \tau_o^{k,k'} \right)^{\frac{1}{\nu}} \right] \tag{10}$$
$$s.t. \qquad c_t^k(\mathbf{s}) = I_t^k(\mathbf{s})/P_t^k.$$

While the ability of (future) children is integrated out, $V_t^{q_P}(k';\mathrm{s})$ reflects the difference of $h^{k'}$ across regions. Conditioning on s, a worker is likely to relocate to a region with higher $h^{k'}$ which delivers higher $\mathbb{E}V_t^{q_P}$ via higher future value of the child V_0^{0-}, all others equal. In sum, the relocation decision account for both worker's own job prospect and child's expected lifetime utility, which is a function of human capital. The migration probability can be derived likewise to (5).

The worker's problem defined above jointly generates a set of policy functions as follows. For $1 \leq q \leq q_R$, $c_t^q(k;\mathbf{s})$ denote consumption; For $0 \leq q \leq q_R - 1$: $m_t^q(k, k';\mathbf{s})$ denotes migration probability from k to k'; For $q = q_P$, $(e')_t^{q_P}(\varepsilon', k;\mathbf{s})$ denotes skill investment for children; For $q = 0$, $o(\varepsilon', e, k_p)$ denotes education choice.

2.2.2 Production

Production is a standard static representative firm's problem. In each region, a representative firm hires workers in both college and non-college in order to maximize output with the following CES technology. The time index t is omitted to ease the notation.

$$Y^k = \Theta^k \Big[A_c^k (L_c^{kD})^{\frac{\sigma-1}{\sigma}} + A_n^k (L_n^{kD})^{\frac{\sigma-1}{\sigma}} \Big]^{\frac{\sigma}{\sigma-1}}, \tag{11}$$

where Θ^k is region k's output TFP, σ is the elasticity of substitution, and $A_i^k(i = c, n)$ and the CES demand shifters ($A_c^k + A_n^k = 1$). The education-specific effective labors are not perfectly substitutable, but workers are perfectly substitutable within each education cell. $L_i^k(i = c, n)$ is the effective labor demanded by the final good producer. The final good is non-tradeable. The labor demand schedule is standard.

$$L_o^{kD} = (\Theta^k)^{\sigma-1}\Big(\frac{w_o^k}{A_o^k P^k}\Big)^{-\sigma} Y^k \tag{12}$$

Given the market wages of each occupation (w_c^k, w_n^k), the competitive producer's cost minimization yields the price of the final good:

$$P^k = \frac{1}{\Theta^k}\Big[(A_c^k)^\sigma (w_c)^{1-\sigma} + (A_n^k)^\sigma (w_n)^{1-\sigma}\Big]^{\frac{1}{1-\sigma}} \tag{13}$$

2.2.3 Equilibrium

I close the model with the following market clearing conditions.

Measure Define $\mu_t^q(k; \mathrm{s})$ to be the measure of type s agents age q in time t. Let $\mathcal{S}_e$ and $\mathcal{S}_\varepsilon$ be the space of investment and innate ability, respectively. Define the space space $\mathcal{S} = \mathcal{S}_e \times \mathcal{S}_\varepsilon$ and $\mathcal{B}$ the $\sigma-$ field on $\mathcal{S}$. Each age group has the uniform size $\frac{1}{q_R}$. The measure of workers is normalized: $\sum_{q=1}^{q_R} \sum_{k_B,k,o} \int_{\mathcal{B}} d\mu_t^q(k; \mathrm{s}) = 1$ for any t. The measure of children in region k before drawing abilities and making any choice is identical to the measure of parents: $\mu_t^0(k) = \sum_{k_B} \int_{\mathcal{B}} d\mu_t^{q_P}(k; \mathrm{s})$.

Local Goods market clearing (spot) The total expenditure on the final good and equals the total consumption and skill investment.

$$Y_t^k = C_t^k + E_t^k, \tag{14}$$

where

$$C_t^k = \sum_{q \geq 1, k_B} [\int_{\mathcal{B}} c^q(\cdot) d\mu_t^q], \qquad E_t^k = \sum_{k_B} \int_{S_{\varepsilon'}} \int_{\mathcal{B}} e_t' d\mu_t^{q_P}$$

Note that E is integrated over the measure of parents, because they choose e'.

Local Labor Market Clearing (spot) This should be straightforward: no excess demand in the labor market.

$$L_{ot}^{kS} = L_{ot}^{kD} \qquad \forall_{o,k,t} \tag{15}$$

where

$$L_{ot}^{kS} = \sum_{q \geq 1} L_{ot}^{kS}(q) \tag{16}$$

$$L_{ot}^{kS}(q) = \sum_{k_B} \Big[\int_{\mathcal{B}} (h^{k_B} e^{\eta}) \varepsilon d\mu_t^q \Big] \tag{17}$$

Notice that the summation in (17) is with respect to region of birth k_B given the region of residence k.

Law of motion: global evolution of the labor force We need an equation that summarizes the aggregate labor market evolution.

$$\mu_{t+1}^{q+1}(k', \mathbf{s}) = \sum_k m_t^q(k, k'; \mathbf{s}) \mu_t^q(k, \mathbf{s}) \qquad \forall t, \quad 0 \leq q \leq q_R - 1 \tag{18}$$

That is, for $q \geq 1$, labor supply in region k' at time $t+1$ is a sum of effective labor of staying workers and in-migrants from all regions.

Equilibrium At every period, endogenous outcomes are determined by the distributions of labor (so human capital). The distribution is determined by the dynamic problem of workers and the constant fundamentals of the economy, which I denote by Γ. It consists of the total factor productivity Θ^k, the skill productivities h^k, amenities a^k, CES demand shifters A_o^k, and moving costs $\tau_o^{k,k'}$. The structural parameters in the model are given by the risk aversion ρ, the migration elasticity ν, the dispersion of the innate ability distribution σ_ε, the elasticity of skill production η, college fixed cost χ, college idiosyncratic cost dispersion σ_z, and the CES substitution parameter σ. I define a stationary equilibrium as follows. With the paper's analysis centered on steady states, time subscripts are omitted.

> **Definition 1.** Given the fundamentals Γ, a stationary equilibrium of the model consists of prices $\{w_o^k, P^k\}$, value functions and policy functions for workers $\{V^q(k;\mathbf{s})\}, c(k;\mathbf{s}), e'(k;\mathbf{s}), o(k,e,\varepsilon), m^q(k;\mathbf{s})$, policies for firms $\{L_o^{kD}\}$, and a stationary measure $\mu^q(k;\mathbf{s})$ that solves the dynamic problem of individuals (4), (6), (7), (8), (9), (10), and the corresponding laws of motion for labor (18), sub-static subproblem for production (11), (12) (13), and market clearing conditions (14), (15).

In English, a stationary equilibrium describes a situation in which no aggregate variables change over time. Individuals may move from one market to another, but the net flows are zero. Appendix D.2 is a complete description of my solution procedure for the equilibrium.

2.3 Qualitative Predictions

This subsection describes the qualitative predictions of the model compare it with the existing literature. In order to unpack the intuition transparently, I consider a simplified partial equilibrium model. I specifically consider workers live through three periods ($q_R = 3$), after one period of education as a child. I shut down college choice.

2.3.1 Skill Investment in the OLG setup

The OLG structure in the model follows a common practice of the literature in that a working parent makes an investment decision. The implementation is somewhat non-standard in that the continuation value of children is considered in the middle of the parents' lifecycle. Therefore, it is useful to clarify skill investment incentives and under which conditions those are operative.

I focus on the skill investment choice e'. There is no migration for now, thus the worker's state vector s $\equiv (e, \varepsilon)$. $V^q(e, \varepsilon)$ is a value of a worker with age q, skill investment e, and ability ε. Wage w is exogenous and the final good price is normalized to 1. The worker's problem then can be written as the following finite-horizon dynamic programming problem.

$$
\begin{aligned}
&V^3(e,\varepsilon) = u(I) \\
&V^2(\varepsilon'; e, \varepsilon) = \max_{c,e'} u(I - e') + \alpha V^0(e', \varepsilon') + \beta V^3(e, \varepsilon) \quad \text{s.t.} \quad c + e' = w(he^{\eta}\varepsilon) \\
&V^1(e,\varepsilon) = u(I) + \mathbb{E}_{\varepsilon'} \beta V^2(\varepsilon'; e, \varepsilon) \\
&V^0(e,\varepsilon) = \beta V^1(e, \varepsilon).
\end{aligned} \tag{19}
$$

Differentiating (19) with respect to e', we have the FOC that governs the investment choice, which is the heart of skill formation:

$$
u_c(I(e,\varepsilon) - e') = \alpha V^0_{e'}(e', \varepsilon') \tag{20}
$$

The LHS represents the parental side of the choice by the marginal cost of skill investment due to the foregone consumption of parents. The RHS, the child's side of the choice, the marginal value of the skill investment, is fundamentally a function of the discounted sum of the marginal utility of consumption of children throughout their lifetime. In other words, (20) is the Euler equation that governs the consumption-skill investment decision.[8]

Parental income, which is a function of (e, ε), and the child's ability ε' are two main factors that determine the optimal e'. A higher parental income, from a higher parental H, shifts it up ("parental income effect"). A higher ability ε' implies a higher HC given e', and so does higher consumption, thus shifting it up ("child ability effect"). Those recast standard lessons of an intertemporal consumption choice. Consider the child's ability effect first. A higher ε' implies a higher income for the child and a rise in the child's consumption in all periods. (income effect) It changes the relative prices of the parent's own consumption in the current period and the child's lifetime consumption. (substitution effect) In other words, the child ability effect is a version of the (intertemporal) price effect. The parental income effect, on the other hand, changes a degree of the intertemporal substitution effect by changing the marginal utility of parental consumption, which eventually changes the relative price of consumption and investment, given ε'. Since both sides are a function of marginal utility, the curvature of $u(c)$ has substantial implications on the e' choice. In a nutshell, under the CRRA setup, the model needs $\rho \in (0, 1)$ to generate reasonable qualitative predictions. It makes the substitution effect dominates the income effect. Otherwise, the RHS decreases in ε', making a weird prediction that parents spend less for smarter children.[9]

Comparison against reference papers The HC production function closely follows Hsieh et al. (2019). Human capital investment is increasing in innate ability. That is, the HC production function is not supposed to be sensitive to

[8] This equation cannot be rearranged for a closed-form solution of e' as the RHS involves V^0. Still, it can be shown that the FOC has a unique solution under many different setups as long as value function is concave.

[9] See Appendix A.1 for derivations and risk aversion choice.

utility function choice to this extent. The sensitivity to the utility form comes from the OLG setup. In Hsieh et al. (2019), the key problem is to maximize the net income after skill investment. The HC production function governs the marginal benefit of investment, and the marginal cost is nothing but the unit price of investment. Utility functions thus play no salient roles there. In my model, both marginal benefit and cost are calculated through the utility function given the OLG structure.

2.3.2 Skill Investment in Spatial Economy

I now introduce geography in the model. There are two regions, A and B, and workers can move between those. Migration is driven by idiosyncratic taste shock and moving cost τ. Wages are exogenous, but differ across regions. I consider $w^A < w^B$ without loss of generality. In order to focus on the impact of geography on educational choices, I let HC production efficiency h be identical between the regions.

In words, allowing migration lets workers have a chance to earn from regions other than they are born and educated. It changes the expected returns to education compared to the no migration case. The option to access more lucrative markets encourages low-wage region parents to spend more on their children's education. It is the "American Dream" part of the migration. On the other hand, migration is a *downgrade risk* to the high-wage region parents. It thus discourages them and the optimal human capital choice becomes lower than the case where children stay in the high-region forever. A formal argument starts by rewriting the worker's problem and deriving a new Euler equation. k is a region index.

$$
\begin{aligned}
V^3(k,e,\varepsilon) &= u(I^k) \\
V^2(k,\varepsilon';e,\varepsilon) &= \max_{c,e'} u(I^k - e') + \alpha V^0(k,e',\varepsilon') \\
&\qquad + \nu \log\Big[\sum_{k'} \exp\Big(\beta V^3(k',e,\varepsilon) - \tau \mathbb{D}_{k'\neq k}\Big)^{1/\nu}\Big] \qquad (21) \\
&\text{s.t.} \qquad c + e' = w^k(he^\eta \varepsilon) \\
V^1(k,e,\varepsilon) &= u(I^k) + \nu \log\Big[\sum_{k'} \exp\Big(\beta \mathbb{E}_{\varepsilon'} V^2(k',\varepsilon';e,\varepsilon) - \tau \mathbb{D}_{k'\neq k}\Big)^{1/\nu}\Big] \\
V^0(k,e,\varepsilon) &= \nu \log\Big[\sum_{k'} \exp\Big(\beta V^1(k';e,\varepsilon) - \tau \mathbb{D}_{k'\neq k}\Big)^{1/\nu}\Big] \qquad (22)
\end{aligned}
$$

Differentiating Eq. (21), we obtain a new Euler equation that echoes Eq. (20):

$$
u_c(I^k(e,\varepsilon) - e') = \alpha V^0_{e'}(k;e',\varepsilon') \qquad (23)
$$

Equation (23) clarifies that geography works only through children's value V^0 since the future value of parents does not affect the investment decision. The LHS, the marginal cost, is the same as the no-migration Euler equation. The RHS, the marginal benefit, is now a function of all possible wage flows throughout the child's life. To make this point clear, I differentiate Eq. (22) with respect to e, for a child in region A:

$$
\begin{aligned}
V^0_e(A,e,\varepsilon) &= \frac{\exp\Big(\beta V^1(A,e,\varepsilon)\Big)^{1/\nu} \beta V^1_e(A,e,\varepsilon) + \exp\Big(\beta V^1(B,e,\varepsilon) - \tau\Big)^{1/\nu} \beta V^1_e(B,e,\varepsilon)}{\sum_{k'} \exp\Big(\beta V^1(k';e,\varepsilon) - \tau \mathbb{D}_{k'\neq k}\Big)^{1/\nu}} \\
&= m^1(A,A)\beta V^1_e(A,e,\varepsilon) + m^1(A,B)\beta V^2_e(B,e,\varepsilon) \qquad (24)
\end{aligned}
$$

where $m^{k,k'}$ is a probability of migration from k to k'. In the world of no freedom of movement, $m^{AB} = 0$ and $V^1_e(A,e,\varepsilon)$ is a function of human capital and w^A. In the migration world, $m^{AB} > 0$, and now the marginal value of investment becomes a function of both w^A and w^B. Since the marginal value of investment is higher in B, thanks to $w^A < w^B$, A-born children now enjoy a higher expected marginal value of investment. ("American Dream") It shifts up the RHS of the Euler eqn thus inducing higher investment, ceteris paribus. The story goes in the opposite direction to B-born children. Their expected marginal value of investment decreases, compare to the no migration case.

In other words, migration "averages" the marginal value of investment across region and make regional optimal choices converge to some extent, but not completely. This result also has an important implication for the TFP-induced amplification mechanism. While amplification is still operative in the migration world, its degree would be weaker compared to the no-migration world, because the increase of skill investment is smaller due to the discouragement. It is also notable that the drop in the high-wage region and rise in the low-wage region can be asymmetric, depending on utility form: the rise is greater than the drop under quadratic utility. It is related to the HC gradient of migration propensity, which will be discussed in Section 3. In short, the rise in the low-wage region is large if the gradient is positive and steep. Going back to Equation (24), the two components of the second term of the RHS are increasing in $H\varepsilon$, so skill investment is the one stone that kills two birds.

Discussion: Does migration discourage skill investment in high-wage regions? It may sound unrealistic that migration discourages skill investment in high-wage regions. It could be seen as a result of taste-shock-driven migration. However, (i) if migration happens in every period, (ii) a probability of migration from a high- to low-wage region is non-zero, it is a simple consequence of the tradeoff between parental consumption and children's human capital. Consider a parent in B, the high-wage region. She has to decide how much to spend on her child. She knows that the child has to leave B tomorrow in 100% probability. Then, it is optimal to choose as if she is a parent in A and spend more on her own consumption. In other words, it is about a choice facing a probability of getting into an undesired state.

The "discouragement" happens even if migration is purely selection-based. Suppose that there is an exogenous threshold of skill for B-educated children to stay in B 100%. In this case, again, parents with low-ability children will choose as if they are in A because they know that the child has to leave tomorrow. It means more heterogeneity in education choice across abilities given the region of birth and shows there always exists discouraged parents unless one can permanently live in the high-wage region. Thus, the discouragement is a feature that stems the nature of migration, not from the current modeling strategy.

2.3.3 Migration Propensity and Human Capital

In this subsection, I describe individual migration patterns by one's human capital, such as selection and sorting. It also includes a tradeoff between wage and education quality that all workers to be parents face. I start with the two-region partial equilibrium setup in the previous section first and consider an extra region if needed. Here, all selection and sorting through skill occur within a given education level. A difference in migration propensity across education levels is generated by education-specific moving costs.

Selection I first study selection. I define there is a positive selection if high-skill workers are more likely to move to high-wage regions. It is useful to consider age $q_R - 1$ worker s who is in moves from A to B and will retire tomorrow. The skill gradient of migration propensity $\frac{\partial m^{q=2}(k,k';\mathbf{s})}{\partial(H\varepsilon)}$ is as follow:

$$\frac{\partial m^{q_R-1}(A,B;\mathbf{s})}{\partial(H\varepsilon)} = m^{q_R}(A,A;\mathbf{s})m^{q_R}(A,B;\mathbf{s})\frac{\beta}{\nu} \times \underbrace{\left[\frac{\partial V^{q_R}(B;\mathbf{s})}{\partial(H\varepsilon)} - \frac{\partial V^{q_R}(A;\mathbf{s})}{\partial(H\varepsilon)}\right]}_{\text{net gain of migration from marginal value of HC}}$$

It is intuitive that the worker is more likely to move to B if the net gain is positive. Since $V^{q_R}(k';\mathbf{s}) = u(c^k)$ where $c^k = w^{k'}H\varepsilon$, I can rewrite the net gain term as follows:

$$\frac{\partial V^{q_R}(B;\mathbf{s})}{\partial(H\varepsilon)} - \frac{\partial V^{q_R}(A;\mathbf{s})}{\partial(H\varepsilon)} = w_o^B u'(c^B) - w_o^A u'(c^A) \tag{25}$$

which shows that the marginal value of skill is a product of its unit price, wage, and marginal utility of consumption. It is clear that utility form specification again plays a pivotal role here. Given $w^A < w^B$, the direction of the net gain is governed by the degree of diminishing marginal utility of consumption. The net gain is positive only if the wage gain from migration is greater than the decrement of marginal utility.

I have shown that high-skill workers are more likely to move into the high-wage region, namely B. The flip side of the coin is, they are less like to move into the low-wage region, namely A. It is a natural question to ask whether they are more likely to move compared, in general. It can be analyzed using the skill gradient of the probability of leaving current region k, $1 - m(k,k)$. In short, compared to low-skill ones, high-skill workers are less likely to leave k if k offers a high marginal value of skill but more likely to leave if there are better outside options. It is straightforward to calculate the gradient:

$$\frac{\partial[1 - m^{q_R}(k,k)]}{\partial(H\varepsilon)} = \underbrace{-m^{q_R}(k,k)(1 - m^{q_R}(k,k))}_{(-)} \\ \times \underbrace{\Big[\frac{\partial V^{q_R}(k;\mathbf{s})}{\partial(H\varepsilon)} - \Big(\frac{1}{1 - m^{q_R}(k,k)} \sum_{k' \neq k} m^{q_R}(k,k') \frac{\partial V^{q_R}(k;\mathbf{s})}{\partial(H\varepsilon)}\Big)\Big]}_{(*)} \tag{26}$$

Note that the last term of the RHS is the expected marginal value of skill investment because $1 - m^{q_R}(k,k) = \sum_{k' \neq k} m(k,k')$. Thus, $(*)$ is the difference between the marginal value of investment in the current region and the expected marginal value of skill in other regions, which I called the "outside option." If the current region offers a higher marginal value of skill, $(*)$ is positive, and the whole RHS is negative, meaning that the probability of leaving k is decreasing in skill. The direction goes opposite if the outside option is better than staying.

In sum, higher-skill workers are more responsive to marginal value differential: they are more likely to move to seek higher marginal value, which is coming from higher wages. It is how the model features selection.

Sorting Now I consider sorting across destinations. The same intuition applies. Consider a third region C and the first derivative of the odds ratio between migration propensities from A to B and A to C, $m^{q_R}(A,B;\mathrm{s})$ and $m^{q_R}(A,C;\mathrm{s})$, with respect to skill:

$$\frac{\partial}{\partial(H\varepsilon)} \log\Big[\frac{m^{q_R}(A,B;\mathbf{s})}{m^{q_R}(A,C;\mathbf{s})}\Big] = \frac{\beta}{\nu}\Big[\frac{\partial V^{q_R}(B;\mathbf{s})}{\partial(H\varepsilon)} - \frac{\partial V^{q_R}(C;\mathbf{s})}{\partial(H\varepsilon)}\Big] \tag{27}$$

Here, the direction is again determined by the relative marginal values of human capital. If $w^C < w^B$, the bracket term in the RHS is positive, and high-skill workers are more likely to sort into B, compared to C.

Tradeoff for Parents: Wage versus Education Quality I have been considering the case of workers who will retire tomorrow. Workers who are going to have a child tomorrow face another tradeoff: high wage does not necessarily come with high education quality. If a region has high wages and high education quality, they reinforce each other and increase the migration propensity to the region. But if a region has high wages and low skill production efficiency, or vice versa, it depends on the relative sizes. It would be best illustrated by the total derivative of migration propensity from k to k' of workers to be parents tomorrow.

$$dm^{q_P-1}(k,k') = \frac{\partial m^{q_P-1}(\cdot)}{\partial w^{k'}}dw^{k'} + \frac{\partial m^{q_P-1}(\cdot)}{\partial h^{k'}}dh^{k'} \tag{28}$$

where both partial derivatives in the RHS are positive. (See Appendix A.1 for derivation.) Equation (28) is a mathematical representation of the intuition that I described above. If $dw^{k'} > 0$ and $dh^{k'} > 0$, the total differential is definitely positive. But once their signs differ, the whole direction depends on the size. It is a distinctive incentive of migration for parents.

2.3.4 General Equilibrium Mechanisms

I explain the general equilibrium mechanisms in the model and relate those to the individual-level behaviors stated in previous sections.

Characterizing Equilibrium The equilibrium is characterized by regional fundamentals (Θ^k, h^k) and moving costs. The moving cost is a key friction to feature persistent geographic wage differentials in the steady state. (Cameron, Chaudhuri, and McLauren 2007) Diminishing returns to scale is working as a congestion force here that prevents every worker moving into the highest-wage region in the presence of persistent wage differentials. It also implies that a region must have a high TFP in order to host a high HC stock, by paying them high wages.

Amplification 1: TFP-induced A higher Θ translates into a higher w, inducing a higher e' that further increases the gross output Y and per capita output $y = Y/N$, all others equal. This positive feedback is the first amplification mechanism the model features. In the presence of migration, a higher w induces a higher HC inflow, which is likely to be a result of the inflow of high-skill workers, from the selection at the individual level. It further boosts the gross output and gross skill investment. However, increasing regional HC stock (from both amplification and migration) compresses the regional wage, which while gross output and investment are increasing, the directions of *per capita* output and investment are ambiguous.

Amplification 2: HC-induced Without migration, h works as a part of TFP in the sense that the aggregate labor supply L is a linear function of regional h. Thus, a higher h induces higher investment, which yields a higher gross output thus higher investment. The positive feedback is weaker than the TFP-induced amplification since the marginal product of labor declines and there is no initial boost in regional wage, compared to the TFP-induced amplification. In other words, the total factor productivity is a dominant force because high h-educated workers are likely to move into high Θ regions that offer higher wages. Moreover, skill production efficiency h only affects migratino decision of workers who will be a parent tomorrow.

Amplification 3: Migration-induced Net inflows of human capital is another source of positive feedback cycle of gross regional output growth from higher HC stock, additional skill investment, and additional output growth. However, the direction is ambiguous in per capita term, because the inflow lowers the marginal production of human capital. In terms of migration, a higher h induces a higher inflow of parents, which also pushes down the regional wage. So the same intuition applies. It is useful to consider Equation (28) in this perspective: $dh^{k'} > 0$ decrease $dw^{k'}$, considering the GE effect, thereby quantitatively limiting the region's advantage of having superior education quality.

Amplification 4: Heterogeneity-induced College and non-college workers are an imperfect substitute. Therefore, a higher stock of college-level HC boosts non-college wages, induces in-migration and generates another cycle of positive feedback.

2.4 Leveraging Model-inferred Regional Human Capital

2.4.1 Adjusted Net Inflow: Gains from Human Capital Migration

The model quantifies regional human capital stock and flows. I leverage the model-inferred quantities to study the regional human capital gains (losses) from migration and the contribution of in-migrants to regional economy. I introduce several additional notations: population by current residence N^k, output per capita $y^k \equiv Y^k/N^k$, population and human capital by region of birth $\tilde{L}^k$ and $\tilde{N}^k$, respectively. Moreover, human capital carried by workers from s, who are now in d is denoted by $L^d(s)$. I omit the education subscript o for now to ease the notation.

Suppose that there are two regions A and B. Rewriting the local labor market clearing condition, we have A's human capital stock as a sum of A-born, in-migrants, and out-migrants, which is analogous to the GNP, import, and export in the national accounting identity:

$$\overbrace{L^A}^{\text{Current HC stock}} = \overbrace{\widetilde{L}(A)}^{\text{A-born workers}} + \overbrace{L^A(B)}^{\text{in-migrants}} - \overbrace{L^B(A)}^{\text{out-migrants}} \tag{29}$$

The model-inferred L quantifies the net human capital inflow $L^A(B) - L^B(A)$. It is different from the net population inflow $N^A(B) - N^B(A)$. Region A may suffer from brain drain under the zero net flow of population ($N^A(B) = N^B(A)$) if the out-migrants possess higher human capital ($L^A(B) < L^B(A)$). In general, for a given observed migration flow measured in headcounts, there can be infinitely many possible combinations of embodied human capital in region-born, inflow, and outflow. A high HC stock of region k could be from either its own high HC efficiency or net positive inflow. The population flow, which is observable, is not a perfect measure to deterimne the human capital flows, while it helps the model pin down the per capita HC by source and destination.

Thus, I propose the gross and per capita HC accounts in spatial economy as follows.

$$L^k = \widetilde{L}(k) + \sum_{s \neq k} L^k(s) - \sum_{d \neq k} L^d(k) \tag{30}$$

$$\frac{L^k}{N^k} = \frac{\widetilde{N}^k}{N^k}\Big[\frac{\widetilde{L}^k}{\widetilde{N}^k} + \sum_{s \neq k} \frac{N^k(s)}{\widetilde{N}^k}\frac{L^k(s)}{N^k(s)} - \sum_{d \neq k} \frac{N^d(k)}{\widetilde{N}^k}\frac{L^d(k)}{N^d(k)}\Big] \tag{31}$$

Equation (30) is a multi-region version of Equation (29). It is about the total volume of human capital flows. It is related to the regional output. Equation (31), adjusts the volume using the size of $k-$born population.

The model plays a crucial role in characterizing the structural HC flows, which needs both N and L. The former is observable from the real-world data. L has to be measured through the model, especially to be consistent with the endogenous components. It is the second and third terms in the bracket in (31). I close this subsection by defining the net human capital inflow in terms of per capita measure, Adjusted Net Inflow (ANI) for future use. We now have the education subscript o back.

$$\begin{aligned} ANI_o^k &\equiv \sum_{s\neq k} \frac{N_o^k(s)}{\widetilde{N}_o^k} \frac{L_o^k(s)}{N_o^k(s)} - \sum_{d\neq k} \frac{N_o^d(k)}{\widetilde{N}_o^k} \frac{L_o^d(k)}{N_o^d(k)} \\ &= \frac{1}{\widetilde{N}_o^k} NGI_o^k \end{aligned} \tag{32}$$

The Human Capital Flow and Regional Education The model-inferred L shows a region's working population does not reflect its h. Recall that migration shuffles workers across spaces, and so does h^k. Let's denote a j-educated HC stock in k by $L^k(j) = h^j \hat{L}^k(j)$. $\hat{L}$ is a function of investment and ability, which would be used by an econometrician to measure the region's human capital stock.[10] It lets me write

$$\begin{aligned} L^k &= \sum_j L^k(j) = \sum_j h^j \hat{L}^j \\ &= \underbrace{\Big[\sum_j \Big(\frac{\hat{L}^j}{\sum_j \hat{L}^j} \Big) h^j \Big]}_{\equiv \hat{h}^k} \sum_j \hat{L}^j \end{aligned}$$

where $\hat{h}^k$ is the "migration-averaged" HC efficiency in that it is an weighted mean of h^j by total efficiency units by source j. The second summation term of the last equality is a function of education investment and ability, which are the variables employed by the econometrician. Clearly, $h^k \neq \hat{h}^k$. For example, if h^k is the highest (lowest) among regions, $\hat{h}^k$ underestimates

[10]Consider an econometrician who predicts regional HC stock using individual expenditure, controlling unobserved heterogeneity ("ability") through a framework such as a dynamic factor model.

(overestimates) the region's education quality. It shows that $\hat{h}^k$ cannot be taken as an education quality of region k. My structural framework lets us to back out h^k consistently, going beyond data-driven accounting exercises which accounts for population flows.

2.4.2 Measuring In-Migrants' Contributions in the Economy

My model also quantify the in-migrant's contribution in gross output. It is seemingly straightforward: under perfect competition and perfect substitutability across origin given education, income shares deliver the contribution of $L_o^k(k_B)$ in the neoclassical sense.

$$Y^k = \sum_o w_o^k [\underbrace{L_o^k(k)}_{\text{k-born k-residents}} + \underbrace{L_o^k(-k)}_{\text{all migrants}}] \tag{33}$$

where $-k$ denotes any *s*ource regions other than k, i.e. $L_o^k(-k) \equiv \sum_{s \neq k} L_o^k(s)$. The income share of all migrants with education o is then calculated by $\varpi_o^k(-k) \equiv w_o^k L_o^k(-k)/Y^k$. It however does not identify the *value-added* contribution of migrants. Suppose that all economies are symmetric and $L_o^k(s)$, the migrants' human capital, is identical across s. As a result, no regional economies obtain additional output boosts from migration, no matter how high (or low) $\varpi_o^k(-k)$ is. Any changes in the migrant income shares $\varpi_o^k(-k)$ are mechanical.

It is now clear that *gross* and *value-added* contribution of migrants should be distinguished as in goods and services trade.[11] The notion of value added contribution again calls for the per capita representation. After some algebra, Equation (33) can be rewritten in per capita terms, as a weighted sum across

[11]The notion I have in mind is the double-counting problem and measurement of value-added content of trade. (HIY 2001, etc) Here, I frame the "purely shuffled" HC as the "double counted" domestic content of exports.

occupation and origin:

$$y^k = \sum_o w_o^k \overbrace{\frac{N_o^k}{N^k}}^{\text{Emp shares by Educ}} \Big[\underbrace{\frac{N_o^k(k)}{N_o^k}\frac{L_o^k(k)}{N_o^k(k)}}_{\text{k-born k-residents}} + \underbrace{\frac{N_o^k(-k)}{N_o^k}\frac{L_o^k(-k)}{N_o^k(-k)}}_{\text{all migrants}} \Big] \tag{34}$$

where

$$\frac{N_o^k(-k)}{N_o^{-k}}\frac{L_o^k(-k)}{N_o^k(-k)} \equiv \sum_{s \neq k} \frac{N_o^k(s)}{N_o^k}\frac{L_o^k(s)}{N_o^k(s)}$$

The per capita income is decomposed into the per capita HC by occupation and origin, weighted by occupational employment shares and within-occupation origin composition. I define the Relative Human Capital Per Capita (RHCPC) of migrants in k as a per capita HC ratio of migrants and natives:

$$RHCPC_o^k = \frac{L_o^k(-k)/N_o^k(-k)}{L_o^k(k)/N_o^k(k)} - 1 \tag{35}$$

Simply put, $RHCPC_o^k$ is the excess per capita human capital of migrants. It is positive (negative) if the migrants possess a higher (lower) per capita HC compared to the native residents. $RHCPC_o^k = 0$ if the migrants and natives have the same per capita HC. The underlying thought experiment is: which level of per capita output would k have attained if the economywide per capita HC were the same as the native residents, given w? Denoting by $\bar{y}^k$ the counterfactual per capita output, the comparison is between y and $\bar{y}^k$.

It is worthwhile to clarify a conceptual difference of ANI and $RHCPC$. Clearly, ANI is an absolute measure related to gross output Y, while $RHCPC$ is a relative measure related to per capita output y. The more important difference is the underlying logic. The former compares the in-migrants against the out-migrants while the latter does against the stayers. Suppose that the latter follows the former. The underlying thought experiment compares the current per capita output against a counterfactual per capita output before HC reallocation, i.e. y and $\tilde{y}$.[12] As $\tilde{y}$ is calculated as if there is no labor reallocation between economies, this alternative comparison is about the additional gain from featuring endogenous migration on top of endogenous human capital accumulation. It is a relevant question for quantitative evaluation of each model component (comparison between equilibrium outcomes), but not for measurement of value added contribution of migrants given equilibrium.

2.5 Development Accounting

It is straightforward to derive the following accounting equation from Equation (11):

$$1 = \frac{Cov[\ln y, \ln g(L_c, L_n)]}{Var[\ln y]} + \frac{Cov[\ln y, \ln \Theta)]}{Var[\ln y]} \tag{36}$$

where g is the CES aggregator without the TFP . Equation (36) is a standard development accounting equation, which decomposes cross-region per capita income variations into the TFP and the factor variation. Using the model-inferred L seamlessly accounts for the amplification effects via endogenous skilll formation and migration in the decomposition.

3. Taking the Model to the Data

In this section, I describe how I take the model to the data. I begin by briefly describing the data sources. I then detail how I identify the model's structural parameters with a combination of estimation and calibration strategies. Finally, I show the quantification results at U.S. census regions and U.S. states, respectively.

[12]Technically, the counterfactual wage $\tilde{w}$ would be different from w and $\tilde{y}$ should be adjusted accordingly. I deliberately ignore it as the key is the choice of counterfactual here.

3.1 Data

This subsection summarizes the data sources and measurements used to take the model to the data. All further data descriptions are in Appendix C.1. All monetary values are expressed in 2015 U.S. Dollars and normalized with respect to the 2000 median wage, \$42,119.

Region Choice, Regional Output, and Working Population I start with 47 U.S. states. I exclude Alaska, Delaware, Wyoming, and the District of Columbia to rule out potential confounding following Hanushek et al. (2017). I use the 1990 GDP by State data of the Bureau of Economic Analysis. As the model has no capital, I trim capital quantitatively, assuming the Cobb-Douglas with labor and capital and the labor's share θ = 0.604 (Hsieh et al. 2019). Specifically, suppose $Y = L^{\theta}K^{(1-\theta)}$. I use $\hat{Y} = (Y/K^{1-\theta})^{1/\theta}$ for calibration. I take the state-level capital data from El-Shagi and Yamarik (2019). I use total employment of the BEA data for the regional population, given the model assumes all adults are employed.

Wages and College Shares I use the 1990 U.S. Census to calculate the average wages by education and state and college-educated population shares by state. I define a worker as college-educated if the reported education status is more than or equal "some college".

Migration Flows I construct 5-year migration flows across states for each education type using the 1985-2010 Current Population Survey Annual Social and Economic Supplement (CPS-ASEC) and 1980-2000 Decennial Census. See Appendix C.2 for details.

3.2 Model Parameters

This subsection discusses the estimation and calibration strategies of the model parameters.

Table 1: Model Parameters

Parameter	Description	Value	Source
A. Estimated Parameters			
$1/\nu$	Migration elasticity	1.62	PPML estimation, Table 2
$\tau_o^{k,k'}$	Moving costs		PPML estimation, Table 3
B. Internally Calibrated Parameters			
σ_ε	Innate ability dispersion		
χ	College Attendance Fixed Cost		
σ_z	Idiosyncratic College Cost Dispersion		
a^k	Amenity	ZZZ (mean)	
h^k	Human Capital Production Efficiency	ZZZ (mean)	
Θ^k	Total Factor Productivity	ZZZ (mean)	
A_o^k	CES demand shifter	ZZZ (mean)	
C. Externally Set Parameters			
α	Altruism	0.5	Del Boca et al. (2019)
β	Subjective Discount Factor	0.9	Equiv. to 0.98 (annual)
ρ	Risk Aversion	0.9	Chetty (2006)
σ	Elasticity of Substitution	1.5	Cantore et al. (2017)
η	Elasticity of Human Capital Investment	0.103	Hsieh et al. (2019)

Mean moving costs are by education (non-college, college).

3.2.1 Estimated Parameters

The migration elasticity $1/\nu$ and the moving cost $\tau_o^{k,k'}$ are the key parameters that govern the migration layer of the model. I use the CPS-ASEC (Current Population Survey-Annual Social and Economic Supplement) data to estimate the parameters. I estimate it by using a two-step Poisson Pseudo-Maximum Likelihood (PPML) regression, following Artuc (2013) and Artuç and McLaren (2015).[13] The first step exploits the variations in bilateral gross

[13] See Appendix D.1 for all the details and derivations.

migration flows to identify the region-specific continuation values and the moving costs. The first step regression equation is derived from Equation (5):

$$Z_t^q(k,l;\mathbf{s}) = \exp\left[dest_t^q(l;\mathbf{s}) + orig_t^q(k;\mathbf{s}) - \frac{\tau_o^{k,l}}{\nu}\right] + \zeta_t^q(k,l;\mathbf{s}) \tag{37}$$

where $Z_t^q(k,l;\mathbf{s})$ is the headcount of individuals moving from k to l, $dest_t^q(l;\mathbf{s})$ is a destination fixed effect, $orig_t^q(k;\mathbf{s})$ is an origin fixed effect, $\tau^{k,l}/\nu$ is the moving cost normalized by the migration elasticity, and $\zeta_t^q(k,l;\mathbf{s})$ is an error term. The second step uses the first stage results and the wage variations across regions and education to identify the migration elasticity, estimating the following equation derived from (4):

$$\kappa_t^q(k,\mathbf{s}) = D_t^q(\mathbf{s}) + \frac{\beta}{\nu}u\Big(\frac{I_{t+1}^k(\mathbf{s})}{P_{t+1}^k}\Big) + \zeta_t^q(k,\mathbf{s}) \tag{38}$$

where $\kappa_t^q(k,\mathbf{s})$ is a function of $N_t^q(k,\mathbf{s})$ and the fixed effect estimates, $D_t^q(k,\mathbf{s})$ is time dummy and $\zeta_t^q(k,\mathbf{s})$ is the error term. As in the previous studies, the cross-sectional migration flows convey information on expected future values that are a function of future wages and option values. Moreover, future migration flows are sufficient statistics for the option values of the regional labor market.

Setting $\beta = 0.9$, I obtain the preferred $1/\nu$ estimate 0.617, which implies $\nu = 1.62$. (Table 2) It is slightly lower but comparable to the annual estimates CDP's 2.02 for the U.S. commuting zones and Caliendo et al. (2021)'s 2.3 for European countries. The mean of the preferred $\tau^{k,k'}$ estimates are 7.54 and 6.56 for non-college and college, respectively. (Table 3)

3.2.2 Calibrated Parameters

I calibrate the region-specific fundamentals $\{h^k, \Theta^k, A_o^k, a^k\}$ (188 parameters) and the college choice parameters $\{\chi, \upsilon\}$ (2 parameters), given the estimated

Table 2: The Migration Elasticity

	(1) All	(2) Non College	(3) College
$1/\nu$	0.617***	0.546***	0.643***
	(0.142)	(0.188)	(0.141)
N	1880	940	940

The difference of college and non-college estimates are statistically insignificant.

Table 3: The Estimated Moving Costs

	Mean	SD	Min	Max
Non-College	7.234	1.744	3.081	12.139
College	6.450	1.304	3.307	10.712

See Appendix C.2 for the sample definitions. Non-college and College stand for "High School or less" and "Some College or More", respectively.

migration-related parameters. My data targets are regional per capita GDP y^k, population N^k, average wage by education and region w_o^k, and the national population share of college graduates. All others equal, (y, w) identifies the TFP Θ^k via the CES aggregator and human capital demands. Higher (y^k, w_o^k) implies higher Θ^k. Amenities a^k, which clearly makes regions more attractive, adjust to target N^k. The human capital production efficiency h^k governs the educational investment and thus human capital supply. It adjusts to clear all regional labor markets, where per capita labor demand L_o^k/N^k is characterized given the data moments and Θ^k. The CES demand shifter A_o^k adjusts to match the *relative* labor demand L_c^k/L_n^k to the relative supply. A higher per capita human capital carried by college workers implies a higher A_c^k/A_n^k. Lastly, χ and υ jointly match the national share of college graduates.

A higher χ makes college education expensive, thus decreasing the college share but increasing the per capita HC of college workers, given wage premia w_c^k/w_n^k. The dispersion of idiosyncratic cost governs the college enrollment behavior by skill. A higher dispersion implies that any high (low) human capital child goes (does not go) to college, thus equalizing the per capita HC of college and non-college workers. Thus, χ and υ jointly pin down the college population share at the national level and the implied per capita HC. Lastly, the standard deviation of labor income pins down the innate ability dispersion σ_ε. Appendices D.2 and provide the full detail of the calibration and model solution method.

3.3 Quantification: U.S. States

I calibrate the model to the 2000 U.S. economy. I include 47 U.S. states, excluding Alaska, Delaware, Wyoming, and the District of Columbia, following Hanushek et al. (2017).

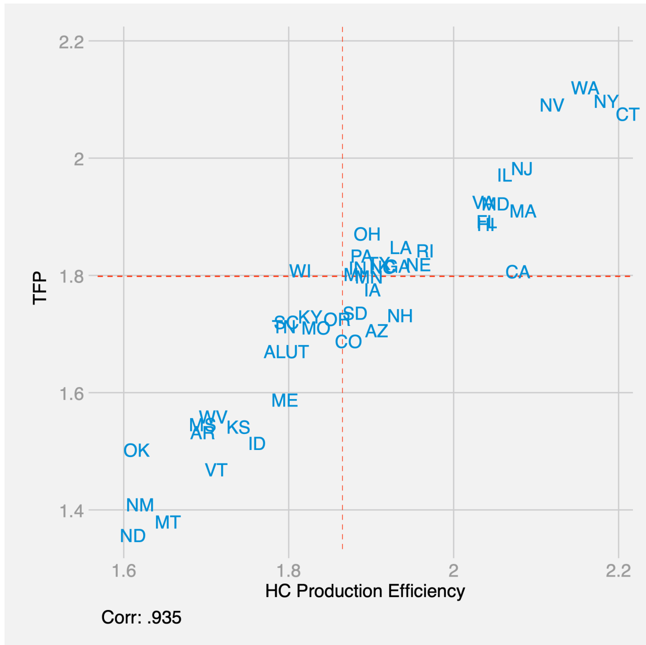
3.3.1 Model Fit

The model exactly matches the state-level moments: per capita GDP, wages by education, and population shares. (Appendix Table E.1) It also closely matches the national population share of college graduates. Moreover, while not being directly targeted, migration rates by education are close to the data counterpart. (Table 4)

Table 4: Model Fit: College and Migration Rates

	Model	Data
College-graduate population share	0.668	0.614
Migration Rate (Non-College)	0.085	0.071
Migration Rate (College)	0.135	0.122

Figure 1: Model-Inferred TFP and HC Production Efficiency



3.3.2 Model-Inferred TFP and Human Capital Production Efficiency

The model-inferred TFP and h are highly correlated. (Figure 1)

Validating the Model-Inferred Parameters The main goal of the calibration is to back out the human capital production efficiency, h^k and the final good TFP, Θ^k. I first examine the validity of model-inferred h using public education expenditure per pupil. education-related variables as a proxy of education quality. I consider a simple OLS regression of h on public education expenditure per pupil:

$$h^k = b_0 + b_1 \text{Expenditure}^k + \varepsilon^k,$$

where Expenditure can be either level (in terms of 1,000 USD) and log variables. Using the 1990 and 2000 expenditure data, Table 5 reports that h is positively correlated with expenditure per pupil. A $1,000 USD increase in per pupil expenditure leads to 0.031-0.038 unit increases in h. In terms of percentage, a 1% increase in per pupil expenditure leads to 0.387-0.378 unit increase in h. Combining this to the interpretation of h, a unit converted to efficiency units of labor when $\$1$ spent on human capital investment of households, it means a 1% increase in *public* expenditure per pupil is associated with nearly 40% higher human capital supplied in the state.

Table 5: h and Public Education Expenditure per Pupil

	Level		Log	
	(1) 1990	(2) 2000	(3) 1990	(4) 2000
Expenditure Per Pupil	0.031***	0.038***	0.387***	0.378***
	(0.006)	(0.010)	(0.072)	(0.101)
R-squared	0.329	0.224	0.331	0.214

Robust standard errors in parentheses. $*p < 0.1, **p < 0.05, ***p < 0.01$.

In addition, the model-inferred TFP is positively correlated with the existing estimates from the literature. There are positive rank correlations between the calibrated Θ and the 1980-2014 estimates taken Herkenhoff et al. (2018) . It is notable that the stronger correlations between the model and the 1990 and 2000 estimates, 0.703 and 0.743 respectively, given that the model is calibrated to the 2000 U.S. economy.

Table 6: Model Validation: TFP

	1980	1990	2000	2014
Rank Corr.	0.280	0.703	0.743	0.580
	(0.060)	(0.000)	(0.000)	(0.000)

Figure 2: Human Capital Flows: ANI by Education

(a) High School or Less (b) Some College or More

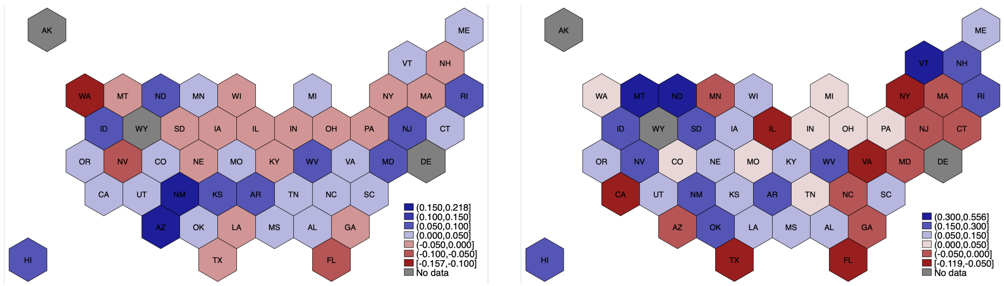
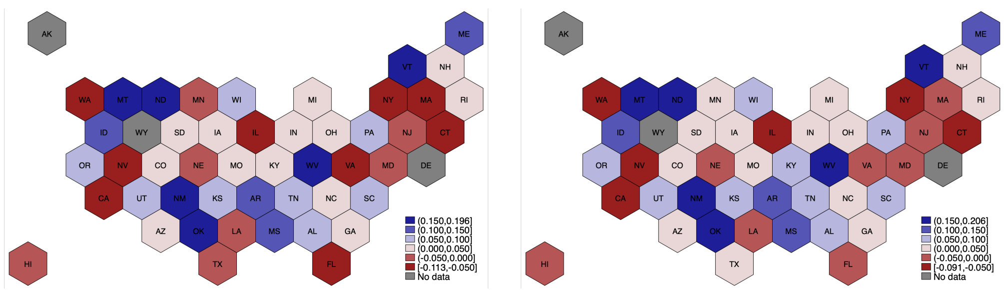


Figure 3: Human Capital Flows: RHCPC by Education

(a) High School or Less (b) Some College or More



3.3.3 Model-Inferred Human Capital Flows

I now use the human capital flows quantified via the lens of the model to disentangle the migration flows in terms of headcount and the implied human capital. Figure 2 visualizes the ANI by states and education. Similarly, Figure 3 visualizes the $RHCPC$ by states and education.[14]

3.3.4 Development Accounting

Finally, I use the model-inferred regional human capital stock L^k to quantify the share of per capita GDP variations explained by human capital. Table 7 re-

[14] For levels, See Figures E.1-E.2 in the appendix.

ports a decomposition result from the accounting equation (36). The human capital variations across states explains 46.6% of per capita GDP variations. It is more than twice as large as the literature, such as 22.8% of Hanushek et al. (2017). The difference comes from the way to measure regional human capital stock. Hanushek et al. (2017) is an accounting exercise with a data-driven measure of human capital. They build an index based on observed educational outcomes such as years of schooling and test scores. They account for migration by weight-averaging the education outcomes of state of birth for each state's working population. My result is inferred from the general equilibrium framework that systematically accounting for migration and its impact on skill formation. Second, Hanushek et al. (2017) has no amplification mechanisms. My model incorporates multiple amplification channels of skill formation, skill heterogeneity, and migration discussed in Section 2.3.4. Those channels can generate large variations of per capita output from tiny variations of TFP and skill production efficiency.

Table 7: Development Accounting: U.S. States

	TFP	Human Capital
explained share of per capita GDP variations	0.534	0.466

4. Counterfactual Exercise

In this section, I implement counterfactual analysis in several dimensions. First, I shut down migration to quantify its role in the regional and national economies. Second, I consider a placed-based education reform. I boost the human capital production efficiency h of a region. It could be interpreted as increasing public educational spending on K-12.

4.1 A House Divided: No Migration Economy

To implement the no migration exercise, I assign $\tau_o^{k,k'} = 3,000$ and solve for market-clearing w_o^k vector that clears all labor markets, given the other region fundamentals $\{\Theta, A, a\}$. The resulting human capital distribution across regions let me calculate the counterfactual per capita GDP of all regional economies and the U.S. economy.

The national output drops by 6.9%. It is amplified through the channels that had boosted the national economy. For example, since skill formation is endogenous, lower income leads to lower skill investment, which generates lower labor supply, and so on. On the other hand, the state-level responses are heterogeneous. Figure 4 shows the changes in per capita output from free mobility to no migration economy. The states with low h, such as Oklahoma, North Dakota, Montana, etc suffers from substantial losses up to 7%. Clearly, those have been benefiting from human capital carried by in-migrants. The states with high h benefits from isolation, as they fully utilize human capital produced within the border. Figures 5 and 6 illustrate the negative correlation between the changes in per capita output and ANI and $RHCPC$ by education in baseline economy. It is notable that net population inflow shows no or weak correlation.

4.2 Evaluation of Race to the Top Grant Program

Race to the Top (RTT) is a $4.1 billion competitive grant program of the U.S. Department of Education, implemented by the Obama administration. The grant amounts vary across states from $18 millions (Colorado) to $700 millions (New York). I use the calibrated model to evaluate the direct and general equilibrium impact of the grant on the state and national economies. Moreover, I examine if an alternative grant allocation scheme could outperform the original.

Figure 4: Changes in per capita Output

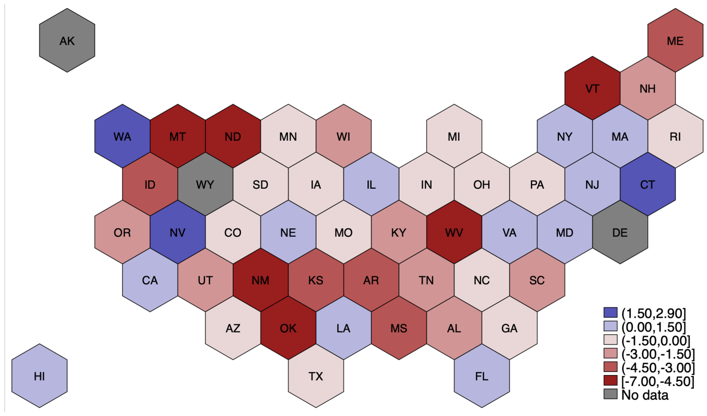
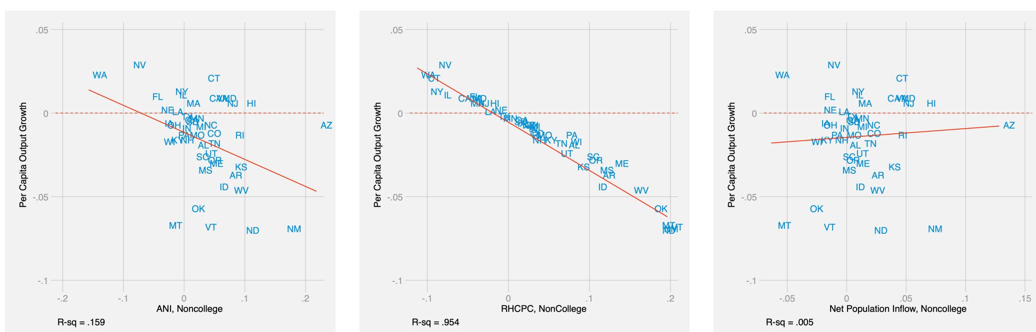


Figure 5: Lost Human Capital Flows and Changes in per capita Output (non-College)

(a) ANI (b) RHCPC (c) Net Population Inflow



4.2.1 RTT in the Calibrated Economy

Simulation Procedure I first convert the RTT grants to per pupil expenditure. It is interpreted as an exogenous subsidy. I take the regression results in Table 5, 0.0381, to further translate the per pupil values to changes in $h \equiv \Delta h$. Finally, I add Δh to the basline h to obtain h_{sim}.

Table 8 is a summary of RTT grant-winning states, funding amounts, and the corresponding Δh.

Figure 6: Lost Human Capital Flows and Changes in per capita Output (College)

(a) ANI (b) RHCPC (c) Net Population Inflow

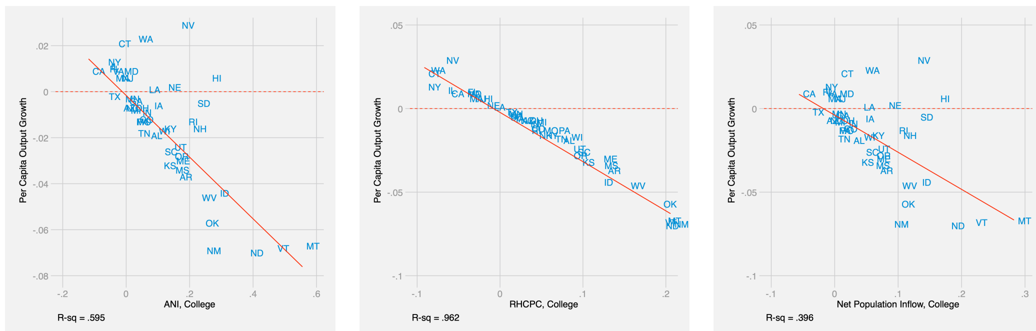


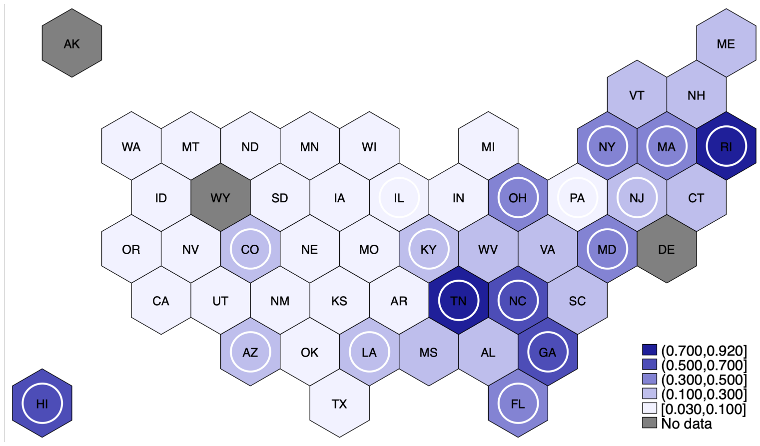
Table 8: Race to the Top

	Grant ($ Mill.)	$ per pupil	% to annual spending	Δh
Arizona	25	28.5	0.38	0.001
Colorado	18	24.8	0.28	0.001
Florida	700	287.5	3.46	0.011
Georgia	400	276.8	2.97	0.011
Hawaii	75	406.8	4.58	0.015
Illinois	43	21.0	0.20	0.001
Kentucky	17	25.5	0.31	0.001
Louisiana	17	22.9	0.28	0.001
Maryland	250	293.1	2.63	0.011
Massachusetts	250	256.4	2.00	0.010
New Jersey	38	28.9	0.19	0.001
New York	700	242.9	1.68	0.009
North Carolina	400	309.2	3.62	0.012
Ohio	400	218.0	2.14	0.008
Pennsylvania	41	22.6	0.20	0.001
Rhode Island	75	476.7	3.80	0.018
Tennessee	500	550.0	7.18	0.021
Delaware	100	872.0	7.23	0.033
District of Columbia	75	1088.1	6.71	0.041

My quantification sample excludes Delaware and the D.C.

Simulated RTT The U.S. output gain is 0.2%. Figure 7 illustrates the regional changes. Output per capita increases in every states. The grant-winning states enjoy higher output per capita, depending on the size of the per-pupil grant.

Figure 7: Race to the Top and Output per Capita



The spillover effect is concentrated to the adjacent states. As a result, the eastern states are more likely to enjoy output gains.

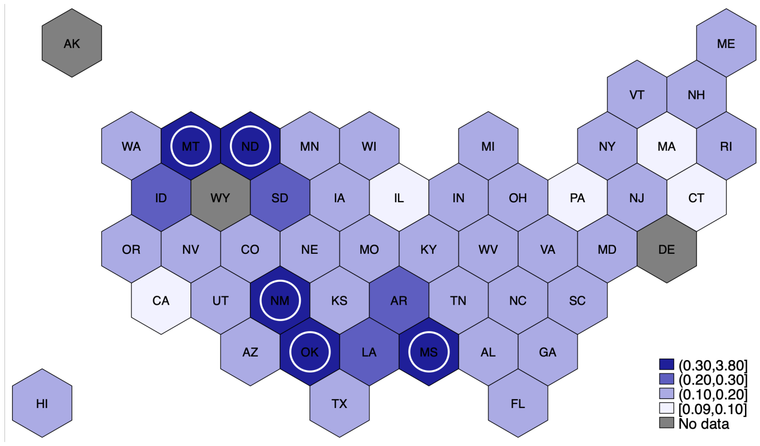
4.2.2 Alternative Grant Allocation

In this subsection, I examine the RTT grant allocation in terms of national output gain. The grant clearly boosts the recipient states' per capita output. It does not imply that the RTT allocation maximizes the national output gain. I experiment with two alternative allocation schemes. First, I allocate the grants to the five states at the bottom of h distribution. Allocating the total RTT budget to those states increases their per pupil spending by \$2,315.6, thus $\Delta h = 0.088$. Second, I allocate the grant to the top five h states. It increases their spending by \$647.0 and $\Delta h = 0.025$. Table 9 summarizes the alternative allocation schemes.

Table 9: Alternative Allocation Schemes

	States	\$ Per Pupil	Δh
Bottom 5	ND, OK, NM, MT, MS	2315.6	0.088
Top 5	NJ, NV, WA, NY, CT	647.0	0.025

Figure 8: Changes in output per capita: Subsidizing the Bottom 5



Subsidizing Bottom 5 Figure 8 shows the output per capita change with respect to the baseline. The grant again benefits the five grant-winning states and the adjacent ones. The five states' output gains vary from 3.25% to 3.8%. The per capita output ratio between New York, the richest state, and Oklahoma is 0.58, which is slightly higher than 0.56 in the baseline. That implies that federal education grant can be a policy device to reduce income equality across states. It comes with a 0.13% U.S. GDP gain.

Subsidizing Top 5 Figure **??** illustrates the output per capita change after subsidizing Top 5, compared to the baseline. The five states' output gains are from 0.8% to 1.1%, which is lower than the Bottom 5 case. The resulting NY-OK output ratio, y_{OK}/y_{NY} stays the same to 0.56, thus boosting the top 5 does not aggravates the regional income inequality. The U.S. GDP, however, increases by 0.27%, which is as twice larger than the Bottom 5 scheme. There are two main factors driving the difference. First, the Top 5 states have superior Θ, thus stronger amplification and spillover, compared to the Bottom 5. Second, compared to the RTT and Bottom 5 case, there are now two separate "clusters" in the East and West coasts. It implies more states benefit from spillover, and it goes into the higher national per capita output gain. In sum, this result illustrates the range of possibility of federalism in education, in the world where workers move around carrying human capital.

5. Conclusion

This paper proposes a unified framework of quantifying human capital stock and flows across spaces. Calibrating the model to the U.S. states, I find that skill production efficiency varies across states up to 20%, which results in cross-state income disparities. Migration-driven reallocation particularly benefits states with lower skill production efficiencies, such as North Dakota and Oklahoma, boosting their per capita output up to 7%. Overall, the national output gain from the free mobility of human capital is 6.9%. Moreover, the model suggests that variations in human capital account for 46.6% of the state variation per capita output. I use the calibrated model to analyze the impact of Obama Administration's Race to the Top grant. The $4.1 billion program yields a 0.2% higher U.S. GDP, mostly benefiting the grant-winning and adjacent states. Alternative grant allocation schemes could yield varied output gains at both regional and national levels. While these findings underscore the significance of education systems in regional economic development, allocating grants to states with lower skill production efficiency may not yield the maximum output gain at the national level.

Moving forward, this work paves the way for a series of intriguing questions to investigate further. In particular, understanding how to tailor federal grant distributions to amplify economic growth invites a closer examination of the interplay between federal education policies and regional economic development. Extending the model to other large economies, such as the European Union, offers a fertile ground for further study. Here, the nuances of migration policies and educational investments within the EU—akin to international dynamics due to the union's distinct structure—pose fascinating scenarios for analysis. Examining the Erasmus Programme, for instance, could yield insights into how education subsidies influence labor mobility and economic integration among EU member states. Future research in these areas holds not only academic value but is also essential for policymakers navigating the landscape of regional and economic policies within large, integrated economies.

References

Amior, Michael, “Education and Geographical Mobility: The Role of the Job Surplus,” Technical Report 1616, CEP May 2019.

Artuc, Erhan, “PPML Estimation of Dynamic Discrete Choice Models with Aggregate Shocks,” June 2013.

Artuç, Erhan and John McLaren, “Trade policy and wage inequality: A structural analysis with occupational and sectoral mobility,” *Journal of international economics*, November 2015, *97* (2), 278–294.

— , Shubham Chaudhuri, and John McLaren, “Trade Shocks and Labor Adjustment: A Structural Empirical Approach,” *The American economic review*, June 2010, *100* (3), 1008–1045.

Bayer, Patrick and Christopher Timmins, “Estimating Equilibrium Models of Sorting Across Locations,” *The Economic Journal*, April 2007, *117* (518), 353–374.

Berry, Christopher R and Edward L Glaeser, “The divergence of human capital levels across cities,” *Papers in regional science: the journal of the Regional Science Association International*, August 2005, *84* (3), 407–444.

Bertoli, Simone and Hillel Rapoport, “Heaven’s Swing Door: Endogenous Skills, Migration Networks, and the Effectiveness of Quality-Selective Immigration Policies,” *The Scandinavian journal of economics*, April 2015, *117* (2), 565–591.

Borjas, G J, S G Bronars, and S J Trejo, “Self-selection and internal migration in the United States,” *Journal of urban economics*, September 1992, *32* (2), 159–185.

Bryan, Gharad and Melanie Morten, “The Aggregate Productivity Effects of Internal Migration: Evidence from Indonesia,” *The journal of political economy*, 2019, *127* (5), 2229–2268.

Caliendo, Lorenzo, Luca David Opromolla, Fernando Parro, and Alessandro Sforza, "Goods and Factor Market Integration: A Quantitative Assessment of the EU Enlargement," *The journal of political economy*, December 2021, *129* (12), 3491–3545.

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

Chetty, Raj, Nathaniel Hendren, and Lawrence F Katz, "The Effects of Exposure to Better Neighborhoods on Children: New Evidence from the Moving to Opportunity Experiment," *The American economic review*, April 2016, *106* (4), 855–902.

Crews, Levi Garrett, "A Dynamic Spatial Knowledge Economy." PhD dissertation, The University of Chicago June 2023.

Cubas, German, B Ravikumar, and Gustavo Ventura, "Talent, labor quality, and economic development," *Review of economic dynamics*, July 2016, *21*, 160–181.

Dahl, Gordon B, "Mobility and the Return to Education: Testing a Roy Model with Multiple Markets," *Econometrica: journal of the Econometric Society*, November 2002, *70* (6), 2367–2420.

Delogu, Marco, Frédéric Docquier, and Joël Machado, "Globalizing labor and the world economy: the role of human capital," *Journal of Economic Growth*, June 2018, *23* (2), 223–258.

Desmet, Klaus, Dávid Krisztián Nagy, and Esteban Rossi-Hansberg, "The Geography of Development," *The journal of political economy*, June 2018, *126* (3), 903–983.

Diamond, Rebecca, "The Determinants and Welfare Implications of US Workers' Diverging Location Choices by Skill: 1980-2000," *The American economic review*, March 2016, *106* (3), 479–524.

Eckert, Fabian and Tatjana Kleineberg, "Saving the American dream? Education policies in spatial general equilibrium," *SSRN Electronic Journal*, 2021.

El-Shagi, Makram and Steven Yamarik, “State-level capital and investment: Refinements and update,” *Growth and change*, December 2019, *50* (4), 1411–1422.

Erosa, Andrés, Tatyana Koreshkova, and Diego Restuccia, “How Important Is Human Capital? A Quantitative Theory Assessment of World Income Inequality,” *The Review of economic studies*, October 2010, *77* (4), 1421–1449.

Fajgelbaum, Pablo D and Cecile Gaubert, “Optimal Spatial Policies, Geography, and Sorting,” *The quarterly journal of economics*, January 2020, *135* (2), 959–1036.

Ferriere, Axelle, Gaston Navarro, and Ricardo Reyes-Heroles, “Escaping the losses from trade: The impact of heterogeneity on skill acquisition,” September 2021.

Fogli, Alessandra and Veronica Guerrieri, “The End of the American Dream? Inequality and Segregation in US Cities,” August 2019.

Ganong, Peter and Daniel Shoag, “Why has regional income convergence in the U.S. declined?,” *Journal of urban economics*, November 2017, *102*, 76–90.

Giannone, Elisa, Qi Li, Nuno Paixão, and Xinle Pang, “Unpacking Moving,” March 2023.

Grogger, Jeffrey and Gordon H Hanson, “Income maximization and the selection and sorting of international migrants,” *Journal of development economics*, May 2011, *95* (1), 42–57.

Hanson, Gordon H and Chen Liu, “Immigration and Occupational Comparative Advantage,” Technical Report w29418, National Bureau of Economic Research November 2021.

Hanushek, Eric A, Jens Ruhose, and Ludger Woessmann, “Knowledge Capital and Aggregate Income Differences: Development Accounting for US States,” *American Economic Journal: Macroeconomics*, October 2017, *9* (4), 184–224.

Hendricks, Lutz, “How Important Is Human Capital for Development? Evidence from Immigrant Earnings,” *The American economic review*, March 2002, *92* (1), 198–219.

— and Todd Schoellman, "Human Capital and Development Accounting: New Evidence from Wage Gains at Migration," *The quarterly journal of economics*, May 2018, *133* (2), 665–700.

Herkenhoff, Kyle F, Lee E Ohanian, and Edward C Prescott, "Tarnishing the golden and empire states: Land-use restrictions and the U.S. economic slowdown," *Journal of monetary economics*, January 2018, *93*, 89–109.

Hsiao, Allen, "Educational Investment in Spatial Equilibrium: Evidence from Indonesia." PhD dissertation, Massachusetts Institute of Technology May 2021.

Hsieh, Chang-Tai and Peter J Klenow, "Development Accounting," *American Economic Journal: Macroeconomics*, January 2010, *2* (1), 207–223.

— , Erik Hurst, Charles I Jones, and Peter J Klenow, "The Allocation of Talent and U.S. Economic Growth," *Econometrica: journal of the Econometric Society*, 2019, *87* (5), 1439–1474.

Jackson, C Kirabo, Rucker C Johnson, and Claudia Persico, "The Effects of School Spending on Educational and Economic Outcomes: Evidence from School Finance Reforms," *The quarterly journal of economics*, February 2016, *131* (1), 157–218.

Kaplan, Greg and Sam Schulhofer-Wohl, "Understanding the Long-Run Decline in Interstate Migration," *International economic review*, February 2017, *58* (1), 57–94.

Kennan, John and James R Walker, "The Effect of Expected Income on Individual Migration Decisions," *Econometrica: journal of the Econometric Society*, 2011, *79* (1), 211–251.

Klein, Paul and Gustavo J Ventura, "TFP Differences and the Aggregate Effects of Labor Mobility in the Long Run," *The B.E. Journal of Macroeconomics*, May 2007, *7* (1).

— and Gustavo Ventura, "Productivity differences and the dynamic effects of labor movements," *Journal of monetary economics*, November 2009, *56* (8), 1059–1073.

Lucas, Robert E, “On the mechanics of economic development,” *Journal of monetary economics*, July 1988, *22* (1), 3–42.

Lyon, Spencer and Michael E Waugh, “Quantifying the losses from international trade,” 2019.

Manuelli, Rodolfo E and Ananth Seshadri, “Human Capital and the Wealth of Nations,” *The American economic review*, September 2014, *104* (9), 2736–2762.

Molloy, Raven, Christopher L Smith, and Abigail K Wozniak, “Declining Migration within the U.S.: The Role of the Labor Market,” April 2014.

Nechyba, Thomas J, “Chapter 22 Income and Peer Quality Sorting in Public and Private Schools,” in E Hanushek and F Welch, eds., *Handbook of the Economics of Education*, Vol. 2, Elsevier, January 2006, pp. 1327–1368.

Rossi, Federico, “The Relative Efficiency of Skilled Labor across Countries: Measurement and Interpretation,” *The American Economic Review*, January 2022, *112* (1), 235–266.

Rossi-Hansberg, Esteban, Pierre-Daniel Sarte, and Felipe Schwartzman, “Cognitive Hubs and Spatial Redistribution,” Technical Report 26267, National Bureau of Economic Research September 2019.

Roy, A D, “Some Thoughts on the Distribution of Earnings,” *Oxford economic papers*, 1951, *3* (2), 135–146.

Schoellman, Todd, “Education Quality and Development Accounting,” *The Review of economic studies*, January 2012, *79* (1), 388–417.

— , “Early Childhood Human Capital and Development,” *American Economic Journal: Macroeconomics*, July 2016, *8* (3), 145–174.

Suzuki, Yuta, “Local Shocks and Regional Dynamics in an Aging Economy,” 2021.

Xiang, Chong and Stephen Yeaple, “Cross Country Income Differences and the Distinction between Hard and Soft Skills in Human Capital,” December 2021.

Young, Alwyn, "Inequality, the Urban-Rural Gap, and Migration," *The quarterly journal of economics*, November 2013, *128* (4), 1727–1785.

Appendix

A. Theory

A.1 Derivation of Model Predictions

This section builds on the simplified setup in Section 2.3. For the following paragraphs, I would abuse the notation to ease the notation to make the point crystal clear. Notice that $V^0_{e'}(\cdot)$ is a discounted sum of weighted sum of $\frac{\partial u}{\partial c'}\frac{\partial c'}{\partial e'}$, where the weight is the migration propensities. Thus it is a linear combination of the derivative. I replace the RHS with the derivative, ignoring the linear combination structure. Moreover, from $c = I = wH\varepsilon$, $\frac{\partial c'}{\partial e'} = \eta wh(e')^{\eta-1}\varepsilon'$, the marginal increase in consumption due to additional investment, does not depend on the utility form.

Log, Linear, CRRA Utility Suppose $u(c) = \frac{1}{1-\rho}c^{1-\rho}$. The FOC becomes

$$\begin{aligned}(whe^{\eta}\varepsilon - e')^{-\rho} &= \alpha\Big[\frac{\partial u}{\partial c'}\frac{\partial c'}{\partial e'}\Big] \\ &= \alpha \times (wh(e')^{\eta}\varepsilon')^{-\rho} \times \frac{\partial c'}{\partial e'} \\ &= \eta(wh\varepsilon')^{1-\rho}(e')^{\eta(1-\rho)-1}\end{aligned}$$

Both sides are a function of ρ.

1. Log utility: let $\rho = 1$. We have

$$\frac{1}{whe^{\eta}\epsilon - e'} = \frac{\alpha\eta}{e'}$$

   The LHS is a decreasing function of (e, ε), the determinant of parental HC. However, the RHS is no longer a function of children's ability ε'

because of the combination of log utility and log-linear HC production function. It mutes the ability effects. It is also notable that the RHS is not even a function of wages, which would have a substantial implication on investment and migration choices in the spatial economy.

2. Linear utility: let $\rho = 0$. In this case, the story goes to the opposite of the log case. The FOC becomes

$$1 = \alpha\eta wh(e')^{\eta-1}\varepsilon'$$

The RHS is an increasing function of ε', so the ability effect is working. On the other hand, The income effect is not operative as the LHS is no longer a function of parental HC.

3. CRRA utility: It features both effects. The LHS is similar to the case of log utility, by definition of CRRA and log utilities, and the parental income effect is just fine. However, the ability effect may go in the wrong direction. If $\rho > 1$, a higher ε' *decreases* the marginal value of educational investment, which implies a reverse ability effect. It requires $\rho \in (0, 1)$ to be in the right direction.

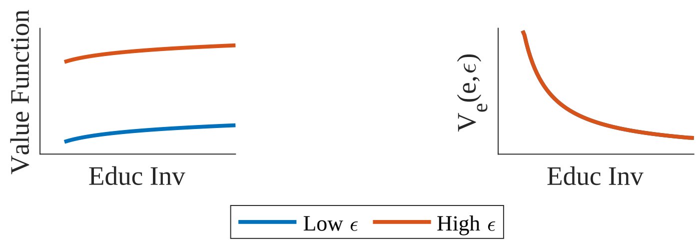
Quadratic Utility In short, both effects are operative here and in the right direction. Suppose $u(c) = -\frac{1}{2}(c - \bar{c})^2$ where $\bar{c}$ is a sufficiently high number. The FOC becomes

$$\begin{aligned}\bar{c} - e' - whe^{\eta}\varepsilon &= \alpha\Big[\frac{\partial u}{\partial c'}\frac{\partial c'}{\partial e'}\Big]\\ &= \alpha(\bar{c} - wh(e')^{\eta}\varepsilon') \times (\eta wh(e')^{\eta-1}\varepsilon')\end{aligned}$$

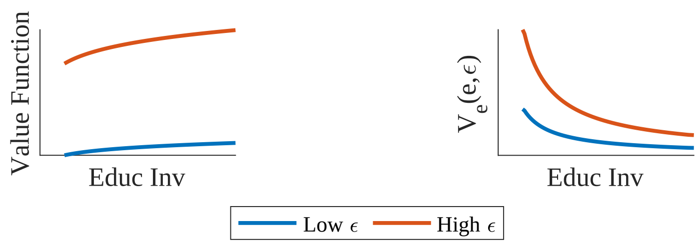
The LHS is decreasing in (e, ε) (income effect). The RHS is increasing in ε' (ability effect). Why? Compared to the linear utility, the LHS has a curvature since the quadratic utility is twice differentiable. Compared to the log utility, the RHS remains to be a function of ε' as the utility function does not convert the log-linear HC into an additively separable object. It is straightforward to show that the RHS is increasing in ε' as long as $\bar{c}$ is sufficiently high.

Figure A.1: Value and Marginal Value of Child by utility form

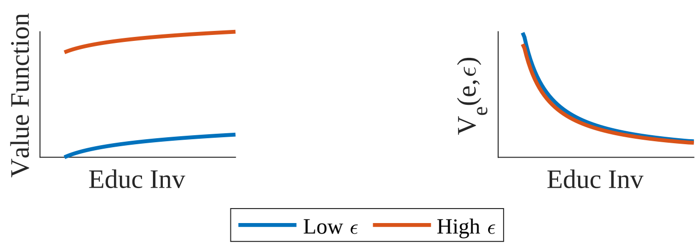
(a) Log utility



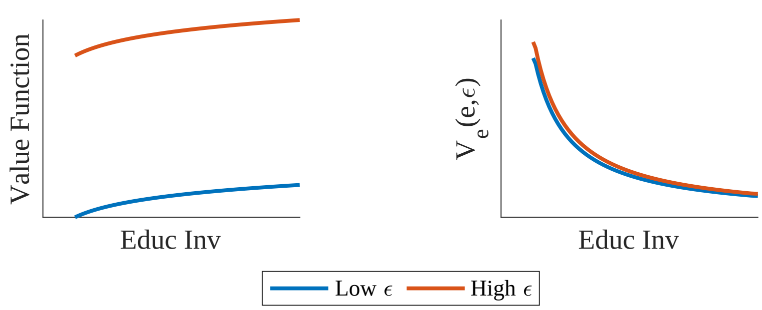
(b) Linear utility



(c) CRRA utility, $\rho = 1.1$



(d) CRRA utility, $\rho = 0.9$



(e) Quadratic utility

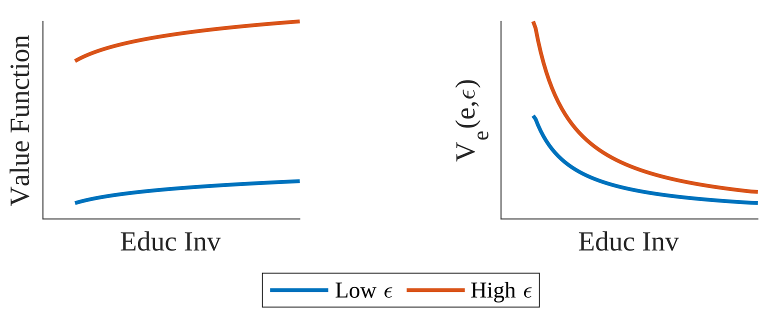
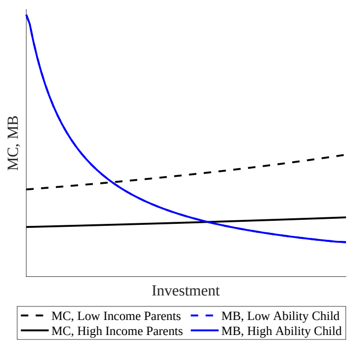
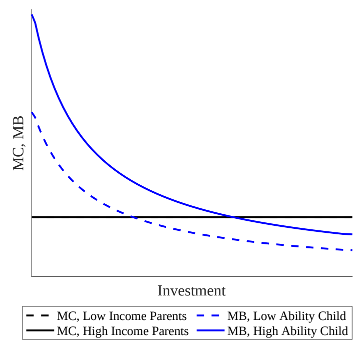


Figure A.2: The Euler Equations by utility form

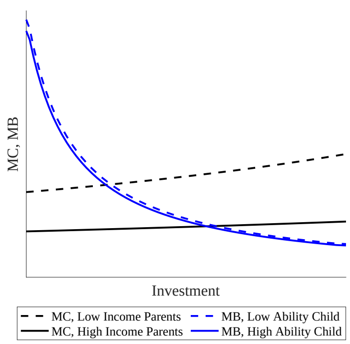
(a) Log utility (No changes in MB)



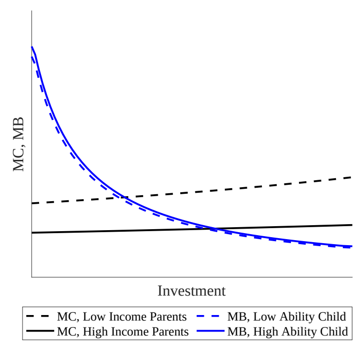
(b) Linear utility (No changes in MC)



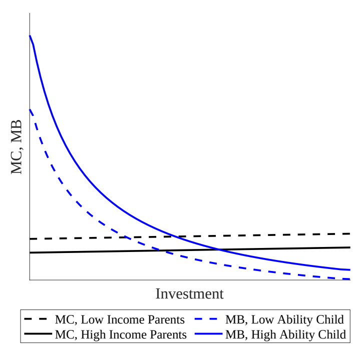
(c) CRRA utility, $\rho = 1.1$



(d) CRRA utility, $\rho = 0.9$



(e) Quadratic utility



B. Utility Forms and Migration Patterns

B.1 Net Marginal Gain of Migration

Recall that $w^A < w^B$. Since the terminal-age worker has no future value, it is straightforward that

$$\frac{\partial V^{q_R}(B;\mathbf{s})}{\partial(H\varepsilon)} - \frac{\partial V^{q_R}(A;\mathbf{s})}{\partial(H\varepsilon)} = w^B u'(w^B H\varepsilon) - w^A u'(w^A H\varepsilon) \tag{39}$$

Under the four utility forms, the RHS can be calculated as follows.

1. Log utility: there is no gain. Again, any wage change in k' serves as a price effect. It is muted under the log utility.

$$\begin{aligned} w^B u'(w^B H\varepsilon) - w^A u'(w^A H\varepsilon) &= \frac{w^B}{w^B H\varepsilon} - \frac{w^A}{w^A H\varepsilon} \\ &= \frac{1}{H\varepsilon} - \frac{1}{H\varepsilon} \\ &= 0 \end{aligned}$$

2. Linear utility: net gain comes from moving to a high-wage region. Its downside is, clearly, the gain is not a function of HC level (e, ε).

$$\begin{aligned} w^B u'(w^B H\varepsilon) - w^A u'(w^A H\varepsilon) &= w^B - w^A \\ &> 0 \end{aligned}$$

3. CRRA utility: net gain comes from moving to a high-wage region only if $rho \in (0,1)$. The gain is a positive but decreasing function of H.

$$
\begin{aligned}
w^B u'(w^B H\varepsilon) - w^A u'(w^A H\varepsilon) &= \frac{w^B}{w^B H\varepsilon} - \frac{w^A}{w^A H\varepsilon} \\
&= \Big[(w^B)^{1-\rho} - (w^A)^{1-\rho}\Big](H\varepsilon)^{-\rho} \\
&\begin{cases} > 0 & \text{if } \rho \in (0,1) \\ < 0 & \text{if } \rho > 1 \end{cases}
\end{aligned}
$$

4. Quadratic utlity:

$$
\begin{aligned}
w^B u'(w^B H\varepsilon) - w^A u'(w^A H\varepsilon) &= \frac{w^B}{w^B H\varepsilon} - \frac{w^A}{w^A H\varepsilon} \\
&= \Big[(w^B)^{1-\rho} - (w^A)^{1-\rho}\Big](H\varepsilon)^{-\rho} \\
&\begin{cases} > 0 & \text{if } \rho \in (0,1) \\ < 0 & \text{if } \rho > 1 \end{cases}
\end{aligned}
$$

B.2 More Gradients

The expectation is taken with respect to the child's ability that will be realized tomorrow.

Wage gradient of migration propensity

$$
\begin{aligned}
\frac{\partial m^{q_P-1}(\cdot)}{\partial w^{k'}} &= m^{q_P-1}(k,k')\frac{\partial \mathbb{E}V^{q_P}}{\partial w^{k'}} \\
&= m^{q_P-1}(k,k')\underbrace{\left[\mathbb{E}\Big(u'(c^{k'})\frac{\partial c^{k'}}{\partial w^{k'}}\Big) + m^{q_P}(k',k')\frac{\partial V^{q_P+1}(k';\cdot)}{\partial w^{k'}}\right]}_{(+)} \qquad (40) \\
&> 0
\end{aligned}
$$

where the expectation is taken across ε'. The bracket term is positive. The first term is trivially positive. The second term is positive because, given $q_R = q_P + 1$,

$$\frac{\partial V^{q_P+1}(k';\cdot)}{\partial w^{k'}} = u'(c^{k'}) \times h e^{\eta} \varepsilon$$

which is positive. This results holds for other age q as long as V is an increasing function of w, which shall be true under any reasonable environments.

h gradient of migration propensity

$$\begin{aligned}\frac{\partial m^{q_P-1}(\cdot)}{\partial h^{k'}} &= m^{q_P-1}(k,k')\frac{\partial \mathbb{E} V^{q_P}}{\partial h^{k'}} \\ &= m^{q_P-1}(k,k')\Big[\mathbb{E}\frac{\partial V^0(k';e',\varepsilon')}{\partial h^{k'}}\Big] \\ &> 0\end{aligned} \tag{41}$$

where the bracket term is the expected marginal value h for children, which is always positive.

B.2.1 The Risk Aversion Choice

Recall the FOC with respect to the investment choice.

$$u_c(I(\cdot) - e') = \alpha V^0_{e'}(\cdot)$$

As both sides are a function of marginal utility, the curvature of $u(c)$ has substantial implications on the model predictions of optimal investment choice. Table B.1 summarizes all results from common utility forms, given the log-linear human capital production function. Figure B.1 visualizes the child ability (first column) and parental income (second and third columns) effects. See Appendix A for the algebra and additional plots.

Table B.1: Utility Forms and Underlying Forces

	Parental income effect	Child ability effect
Logarithmic	Y	N
Linear	N	Y
Quadratic	Y	Y
CRRA	Y	Y only if risk aversion $\in (0, 1)$

The parental income effect is operative as long as the marginal utility of consumption is decreasing. Thus it is working for all but linear, where the marginal utility is always 1. The child ability effect is governed by both HC production and utility function. Recall that HC production is linear in ϵ'.[15] Thus, the effect is operative in the desired way as long as the decreasing rate of the marginal utility of consumption is lower than 1. For log utility, it is $(\varepsilon')^{-1}$, so exactly cancels out. For linear, the marginal utility of consumption is 1 so the marginal increase of HC production, which is again linear in ε' solely governs the child ability effect. For quadratic utility, the substitution effect dominates so the RHS is increasing in ε. Specifically, the marginal utility of consumption and the marginal increase in HC are linear in ε' so the RHS is increasing *quadratically*. Lastly, for CRRA, the RHS becomes a function of $\varepsilon^{1-\rho}$ where ρ is risk aversion. That is, if $\rho \in (0, 1)$, substitution effect dominates income effect but otherwise. So, if $\rho > 1$, RHS is decreasing in ε', making a weird prediction that parents spend less for smarter children.

C. Data

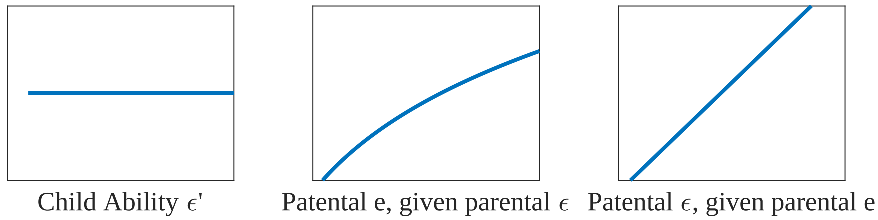
C.1 Data Sources

to be added

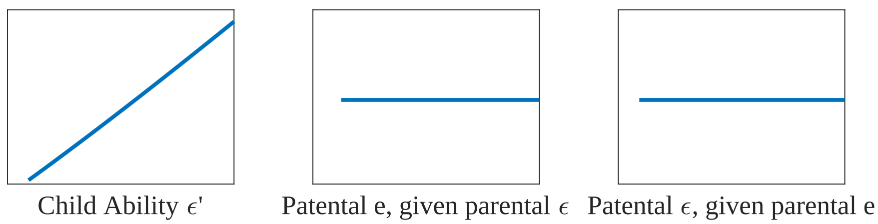
[15]This is linear. The HC technology is log-linear in e, and linear in h and ε.

Figure B.1: Optimal Education Choice by utility form

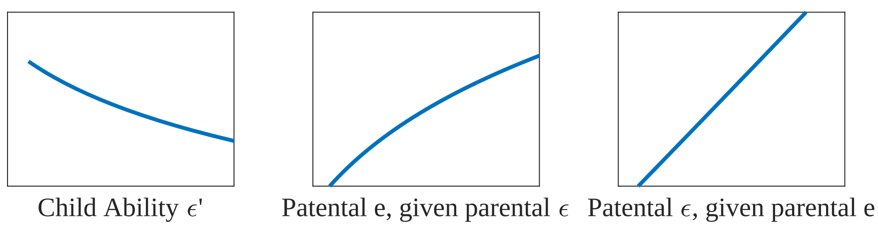
(a) Logarithmic



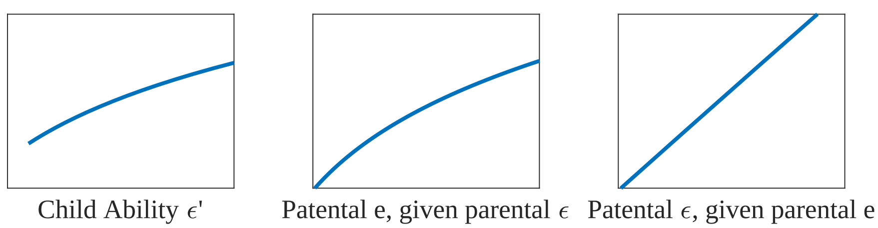
(b) Linear



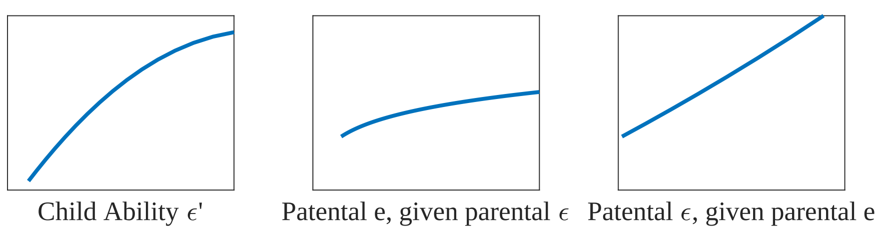
(c) CRRA, $\rho = 1.1$



(d) CRRA, $\rho = 0.9$



(e) Quadratic



C.2 Construction of Migration Flows

Previous studies used two sources of migration flows: (i) Census or equivalent datasets[16] (ii) individual panel datasets.[17] The census-type surveys are more suitable for my goal. Those are designed to cover the entire U.S. and have large sample sizes, which is helpful to construct a migration matrix. However, they do not keep track of the same individuals and the datasets have to be reshaped as a pseudo-panel. The panel datasets' strengths and weaknesses are the opposite.

The two-step PPML estimation needs the flow data from at least two consecutive periods, preferably three or more for lagged instruments. For this reason, using the census-type surveys is not straightforward. One approach with the 5-year flows has to combine two data sources. The other approach with the 1-year flows incurs a "simulation" to translate the observed 1-year flows into 5-year ones.

I start from the 5-year flow approach. The ideal dataset is at least two consecutive series of the 5-year migration flows from a single source. To my knowledge, there is no single dataset that comes with consecutive 5-year migration flows. Two census-type datasets have the 5-year retrospective question ("State of residence 5 years ago"): The decennial census (1980, 1990, 2000) and the March CPS-ASEC (Current Population Survey-Annual Social and Economic Supplement) (1985, 1995, 2005, 2015). For the 1-year question, the literature mostly uses the American Community Survey (2001-2020) and March CPS (1985-2020). (Table C.1) The 1990 and 2000 census are about migration between 1985-1990 and 1995-2000, respectively. The series lacks the information for 1990-1995. Similarly, the ASEC has 1980-85, 1990-95,

[16] Borjas et al. (1992); Dahl (2002); ACM; Molloy et al. (2014); Kaplan and Schulhofer-Wohl (2017); Amior (2019); CDP; Eckert and Kleineberg (2021), etc.

[17] Kennan and Walker (2011) use the NLSY and Kaplan and Schulhofer-Wohl (2017) use the SIPP, for example.

Table C.1: Migration Data Sources

Data Sources by Year								
A. 5-year								
Reference Year	1980	1985	1990	1995	2000	2005	2010	2015
Retrospective	1975	1980	1985	1990	1995	2000	2005	2010
Decennial Census	O		O		O			
ASEC (March CPS)		O		O		O		O
B. 1-year								
ACS				2001 - 2020				
CPS, ASEC				1985 - 2020				

2000-05, and 2010-15. Hence, no single data source is able to construct a consecutive series alone. Alternating the Census and ASEC the only way to have a complete series of observed 5-year flows. In sum, the 5-year migration flows are available, but not from a single source. I name this combined sample the "alternating sample."

Eckert and Kleineberg (2021) use the 1-year flows from the 2006-2010 ACS data in order to construct the 5-year flows. (Appendix B.1) Specifically, they take the following accounting identity seriously:

$$N_{t+1}^{k'} = \mathbf{M}_{t-1,t}^{kk'} N_t^k \tag{42}$$

where N_t^k is the population in origin k before moving and $N_{t+1}^{k'}$ is the population in destination k' after moving. $\mathrm{M}_{t-1,t}^{kk'}$ denotes the the 1-year moving matrix, i.e., the population share that moves from origin m to destination m' between years $t-1$ to t. These objects are observed in the data. They simulate the 1-year moving matrix forward five times to construct 5-year moving flows as:

$$N_t^{k'} = \mathbf{M}_{t-5,t}^{kk'} N_{t-5}^k \approx \Big[\prod_{d=1}^{5} \mathbf{M}_{t,t-d}^{kk'}\Big] N_{t-5}^k \tag{43}$$

I follow their practice to construct the 5-year flows using the CPS-ASEC. As the 1-year migration information is available for more than thirty years and allows me flexibly choosing the initial year of forward iteration. Two samples can be constructed from the 1-year approach. First, the simulated 5-year flows can be used to fill the gaps in the ASEC 5-year information. It constitutes a series of 5-year flows from 1975-2010, where the simulated flows are used for 1985-90, 1995-2000, and 2005-10. Second, a series can be fully simulated from 1985-2010. I call each sample the "simulated 5-year sample." and the "simulated all-year sample," respectively.

All in all, there is no one perfect dataset for migration flows. Table C.2 compares the three constructed samples. Fortunately, the simulated flows are able to predict the observed flows with $R^2 > 0.99$ for any 5-year flows. (Table C.3) The regression coefficients imply the simulated flows are 2-3% smaller than the observed flows while the constant is effectively zero and precisely estimated. I use the simulated all-year sample as a main dataset. The aggregate headcounts are normalized to 100 for each year in order to control for population growth.

Table C.2: Three Samples of 5-year Migration Flows

	Simulated All-year Sample	Simulated 5-year Sample	Alternating Sample
Source	CPS-ASEC	CPS-ASEC	CPS-ASEC and Census
Simulation	Yes, fully	Yes, partially	No
Raw Data Frequency	Annual	Demi-decennial	Demi-decennial
Availablity	1985-2010 (26 periods)	1985-2010 (6 periods)	1985-2000 (4 periods)

Table C.3: Observed and Simulated 5-year Migration Flows

	(1) All ASEC	(2) All Census	(3) 85-90, Census	(4) 90-95, ASEC	(5) 95-00, Census	(6) 00-05, ASEC	(7) 10-15, ASEC
Simulated Flows	1.017***	1.011***	1.026***	1.022***	0.998***	1.016***	1.015***
	(0.008)	(0.006)	(0.012)	(0.018)	(0.014)	(0.018)	(0.004)
Constant	-0.108***	-0.070	-0.160***	-0.136***	0.013	-0.097***	-0.093***
	(0.019)	(0.067)	(0.033)	(0.045)	(0.029)	(0.031)	(0.024)
R-squared	0.999	0.999	0.999	0.998	0.999	0.999	1.000
Obs.	48	32	16	16	16	16	16

Standard errors in parentheses. $*p < 0.1, **p < 0.05, ***p < 0.01$

D. Estimation and Computation

D.1 Moving Costs and the Migration Elasticity: Two-Stage PPML Estimation

I estimate the migration elasticity $1/\nu$ and the moving cost parameter $\tau^{k,l}$ using the two-stage Poisson Pseudo-Maximum Likelihood (PPML) estimation.[18] It derives the regression equations from the household's problem. I derive the estimating equations using Equations (4) and (5), abstracting away from the special periods such as parental, pre-parental, and terminal.

In the first step, I recast (5) as

$$N_t^q(k;\mathbf{s})m_t^q(k,l;\mathbf{s}) = \exp\left[\frac{\beta}{\nu}V_{t+1}^{q+1}(l;\mathbf{s}) - \frac{\beta}{\nu}V_{t+1}^{q+1}(k;\mathbf{s}) + \log N_t^q(k;\mathbf{s}) - \frac{1}{\nu}\Omega_t^q(k,\mathbf{s}) - \frac{\tau^{k,l}}{\nu}\right] \tag{44}$$

where $N_t^q(k;\mathbf{s})m_t^q(k,l;\mathbf{s})$ is the headcount of individuals moving from k to l between t and $t+1$, for each q and s. The option value $\Omega_t^q(k,\mathrm{s})$ is defined as:

$$\Omega_t^q(k,\mathbf{s}) \equiv \sum_{l\in\mathcal{K}} \exp\left[\beta V_{t+1}^{q+1}(l;\mathbf{s}) - \beta V_{t+1}^{q+1}(k;\mathbf{s}) - \tau^{k,l}\right]^{\frac{1}{\nu}}$$

Equation (44) can be rearranged as a PPML regression as follows, which is Equation (37):

$$Z_t^q(k,l;\mathbf{s}) = \exp\left[dest_t^q(l;\mathbf{s}) + orig_t^q(k;\mathbf{s}) - \frac{\tau^{k,l}}{\nu}\right] + \zeta_t^q(k,l;\mathbf{s}) \tag{45}$$

[18] The PPML approach is first employed for estimating the gravity equation in trade literature. Artuc (2013) combined it with dynamic discrete choice framework in the context of sectoral switching. Artuç and McLaren (2015) applied the framework to sectoral and occupational switching in the United States. Recent studies applied it to geographical switching. (e.g. Caliendo et al. (2021), Suzuki (2021) and Eckert and Kleineberg (2021)).

where $Z_t^q(k,l;\mathbf{s})$ is the headcount Nm, $dest_t^q(l;\mathbf{s})$ is a destination fixed effect, $orig_t^q(k;\mathbf{s})$ is an origin fixed effect, $\tau^{k,l}/\nu$ is the moving cost normalized by the migration elasticity, and $\zeta_t^q(k,l;\mathbf{s})$ is an error term. As the fixed effects are identified up to a constant of normalization, I set $dest_t^q(1;\mathbf{s})=0$, or set region 1 as the base category for the destination fixed effect. Specifically, the fixed effects are defined as follows:

$$
\begin{aligned}
dest_t^q(l;\mathbf{s}) &\equiv \frac{\beta}{\nu}\mathbb{E}_t V_{t+1}^{q+1}(l;\mathbf{s}) - \frac{\beta}{\nu}\mathbb{E}_t V_{t+1}^{q+1}(1;\mathbf{s}) \\
orig_t^q(k;\mathbf{s}) &\equiv -\frac{\beta}{\nu}\big[\mathbb{E}_t V_{t+1}^{q+1}(k;\mathbf{s}) - \mathbb{E}_t V_{t+1}^{q+1}(1;\mathbf{s})\big] + \log N_t^q(k;\mathbf{s}) - \frac{1}{\nu}\Omega_t^q(k,\mathbf{s})
\end{aligned}
$$

The normalization with respect to τ is implied by the assumed cost structure $\tau^{k,k}=0\forall k$. Estimating (45) using a PPML is the first step of the procedure.

The second step uses the Bellman equation. Equation (4) can be rearranged as:

$$
\mathbb{E}_t V_{t+1}^{q+1}(k,\mathbf{s}) = u\Big(\frac{I_{t+1}^k(\mathbf{s})}{P_{t+1}^k}\Big) + \beta\mathbb{E}_t V_{t+2}^{q+2}(k,\mathbf{s}) + \Omega_{t+1}^{q+1}(k,\mathbf{s}) \tag{46}
$$

Consolidating the definitions of Ω, $dest$, $orig$ and (46), I obtain the following estimation equation:

$$
\kappa_t^q(k,\mathbf{s}) = D_t^q(\mathbf{s}) + \frac{\beta}{\nu}u\Big(\frac{I_{t+1}^k(\mathbf{s})}{P_{t+1}^k}\Big) + \zeta_t^q(k,\mathbf{s}) \tag{47}
$$

where the dependent variable $\kappa_t^q(k,\mathbf{s})$ is a function of the fixed effect estimates and $N_t^q(k,\mathbf{s})$, $D_t^q(k,\mathbf{s})$ is time dummy and $\zeta_t^q(k,\mathbf{s})$ is the error term. As $\frac{I_{t+1}^k(\mathbf{s})}{P_{t+1}^k}$ is data, I can estimate (47) with the assumption on the form of $u(\cdot)$. The real income is instrumented by the lagged values as in ACM, Artuç and McLaren (2015), and CDP. The literature mostly uses the two-period lagged values, but I use the one-period one, which is effectively five years due to the time specification of the model. The dependent variable and time dummy expressions are as follows.

$$\kappa_t^q(k,\mathbf{s}) \equiv dest_t^q(k,\mathbf{s}) + \beta orig_t^q(k,\mathbf{s}) - \beta \log N_{t+1}^{q+1}(k,\mathbf{s})$$
$$D_t^q(\mathbf{s}) \equiv \frac{\beta}{\nu}\Big(\mathbb{E}_t V_{t+2}^{q+2}(1,\mathbf{s}) - \mathbb{E}_t V_{t+1}^{q+1}(1,\mathbf{s})\Big)$$

There are two issues with applying the PPML estimation procedure for my model. First, in previous studies, dynamic discrete choice framework is translated into the structural regression equations based on two premises: (i) workers live forever; (ii) all workers solve the same problem for every period, facing no intra-period tradeoff.[19] In my model, the agents are finite-lived and the worker's problem changes across periods. The general idea of the PPML procedure is valid while the estimating equations vary by age. Parents face the consumption-educational spending tradeoff, pre-parents take expectation with respect to both idiosyncratic taste shock u and kid's ability draw ε before making relocation decision, and retiring workers hold no option value.

For now, I abstract away from the lifecycle structure and follow the estimation procedure as if the model agents lives forever and solve the same problem for every period. I pool the sample across ages and all state variables other than education. That is, I use gross migration flows and real incomes by region and education for the first and second stages, respectively. (See Appendix C.2 for the three sample description.)

Table D.1 report the summary statistics of $\tau_o^{k,l}/\nu$ estimated in the first stage. The distribution is comparable across samples. Every correlation coefficient between the three sets of estimates is higher than 0.95.

Table D.2 reports the second stage estimates. The all-year sample estimates are moderately higher than the simulated and alternating 5-year

[19]CDP extends their framework to incorporate an elastic labor supply. All workers face the same labor-leisure choice for every period where the optimal leisure choice is up to the time-constant utility parameter with respect to leisure.

Table D.1: The Moving Costs

	Simulated All-year Sample		Simulated 5-year Sample		Alternating Sample	
	$\leq$ HS	$\geq$ SC	$\leq$ HS	$\geq$ SC	$\leq$ HS	$\geq$ SC
Mean	7.539	6.551	7.236	6.440	7.234	6.450
St.Dev	1.473	1.174	1.740	1.287	1.744	1.304
Min	3.745	3.580	3.083	3.323	3.081	3.307
Max	12.017	10.848	11.911	11.003	12.139	10.712

See Appendix C.2 for the sample definitions. HS and SC stand for "High School or less" and "Some College or More", respectively.

estimates and significant at 1%. The two 5-year estimates are all significant but the sizes are different. I take the simulated sample estimate for all workers (Column 1), which is from the largest sample. The difference between HS or Less (Column 2) and Some College or More (Column 3) estimates is statistically insignificant. The Alternating sample estimates are insignificant except the HS or Less sample (Column 7). It could be a result of the small sample size: the identification is from 3 periods, which is likely to be too short, especially compared to the 20 periods of the all-year simulated sample.

Table D.2: The Migration Elasticity

	Simulated Sample, All Years			Simulated Sample, 5 Years			Alternating Sample		
	(1) All	(2) HS or Less	(3) SC or More	(4) All	(5) HS or Less	(6) SC or More	(7) All	(8) HS or Less	(9) SC or More
$1/\nu$	0.617***	0.546***	0.643***	0.385***	0.417**	0.557***	0.246	0.384**	0.243
	(0.142)	(0.188)	(0.141)	(0.137)	(0.171)	(0.161)	(0.153)	(0.185)	(0.183)
N	1880	940	940	470	235	235	282	141	141

See Appendix C.2 for the sample definitions. $*p < 0.1, **p < 0.05, ***p < 0.01$

D.2 Calibration Algorithm

1. Set the convergence tolerance.

2. Take the target moments y^k, w_o^k, N^k, national college share, and national income dispersion from the data. Take the external and PPML-estimated parameters.

3. Guess $h^k, \Theta^k, A_c^k, a^k$. A_n^k is determined by $A_c^k + A_n^k = 1$.

4. Back out L_o^{kD} from y^k, w_o^k, Θ^k, and A_o^k. By doing so, I automatically target y^k and w^k.

5. Solve the model, calculate the model-predicted moments, and regional labor supply L_o^{kS}.

6. Check the distance between data and model predictions for the remaining moments. Also check if all regional markets are cleared.

7. Iterate until converge. I use `lsqnonlin` solver of MATLAB.

D.3 Solution Method of the Household's Problem

This subsection describes a computation strategy for the household's problem. I propose a solution and the corresponding algorithm. I show that the proposed solution delivers a dramatic speed gain and is as accurate as the conventional value function iteration.

D.3.1 Challenge

I start with the computational challenges. The dynamic spatial general equilibrium framework is a high-dimensional problem by nature. The number of states grows with the number of worker types and regions. Even if I strictly limit the number of worker types (inner loop), the model eventually has to deal with the at least 50^2 combinations of 50 U.S. states by each type, to account for migration choice. Since I cannot reduce the number of states, I have to accelerate the inner loop calculation.

The model's inner loop is non-standard. It is a finite-horizon problem, which is typically solved quickly by backward induction demanding no iteration at all. However, the model's parents have to take care of their children in the middle of their life. The initial-period value V^0 thus appears in the parents' Bellman equation. The structure demands an iterative approach so does computation time. A speed of the conventional value function iteration (VFI) effectively prevents the general equilibrium analysis. Parallelization was insufficient to overcome the design limit.

D.3.2 Solution Background: a Brief "Anatomy" of the Model Structure

The idea exploits the finite-horizon OLG structure. To set the stage, I consider the following finite-horizon dynamic programming problem. It is the basic OLG model with no geography and occupation. The age index q is written in general, so q_R is the retirement age, q_P is the parental age. All notations are the same as before.

$$
\begin{aligned}
&V^{q_R}(e,\varepsilon) = u(I) \\
&V^{q_P}(e,\varepsilon,\varepsilon') = \max_{e'} u(I-e') + \alpha V^0(e',\varepsilon') + \beta V^{q_P+1}(e,\varepsilon) \qquad \text{subject to} \qquad c + e' = w(he^{\eta}\varepsilon) \\
&V^{q_P-1}(e,\varepsilon) = u(I) + \mathbb{E}_{\varepsilon'}\beta V^{q_P}(e,\varepsilon;\varepsilon') \\
&V^{q}(e,\varepsilon) = u(I) + \beta V^{q+1}(e,\varepsilon) \qquad \text{for} \quad q \notin \{0, q_P-1, q_P, q_R\} \\
&V^{0}(e,\varepsilon) = \beta V^{1}(e,\varepsilon)
\end{aligned}
\tag{48}
$$

It is useful to have the first order condition and the envelope condition from Equation (48).

$$u_c(I - e') = \alpha V^0_{e'}(e', \varepsilon') \qquad \text{[FOC]} \quad (49)$$

$$V^{q_P}_e(e, \varepsilon, \varepsilon') = u_c(I - e')\Delta + \beta V^{q_P+1}_e(e, \varepsilon) \qquad \text{[EC]} \quad (50)$$

where $\Delta \equiv \frac{\partial I}{\partial e} = \eta w h e^{\eta-1}\varepsilon$. It is worth mentioning that V^q is accurate and does not demand any iteration for $q > q_P$, given the model structure. Moreover, the finite-horizon structure lets me backwardly construct a sequence of V^q_e from the terminal-period marginal value $V^{q_R}_e$, given *any* policy function e':

$$\begin{aligned}
&V^{q_R}_e(e, \varepsilon) = u_c(I)\Delta \\
&V^{q_P}_e(e, \varepsilon, \varepsilon') = u_c(I - e')\Delta + \beta V^{q_P+1}_e(e, \varepsilon) && (51) \\
&V^{q_P-1}_e(e, \varepsilon) = u_c(I)\Delta + \mathbb{E}_{\varepsilon'}\beta V^{q_P}_e(e, \varepsilon; \varepsilon') \\
&V^q_e(e, \varepsilon) = u_c(I)\Delta + \beta V^{q+1}_e(e, \varepsilon) \qquad \text{for} \quad q \notin \{0, q_P - 1, q_P, q_R\} \\
&V^0_e(e, \varepsilon) = \beta V^1_e(e, \varepsilon) && (52)
\end{aligned}$$

Recognize that (51) is the envelope condition (50). That is, the EC holds automatically by construction. It is notable that the EC holds for any choice of e', implying that the FOC and EC are isolated from each other.[20] This result is not surprising and stems from the model structure. Fundamentally, the education investment decision is an endogenous process of worker type that would be fixed over the rest of life, making e operates as an exogenous state variable. The policy function e' is for their children and thus has nothing to do with the EC.

Now the question narrows down to solving the FOC. The LHS $u_c(I - e')$ is readily and inexpensively calculable. The RHS $\alpha V^0_{e'}(e', \varepsilon')$ however involves a derivative of an unknown function V^0. It is seemingly not directly calculable

[20]This is not true in most classes of dynamic programming problems, of course.

and pushes me to use the conventional value function iteration. However, it can be obtained from Equation (52), for *any* e', exploiting the model structure. I simply interpolate the backwardly constructed V_e^0 with the given e' to evaluate $V_{e'}^0(e', \varepsilon')$. Neither backward induction nor interpolation is computationally burdensome.[21] Thus, I need one and only one numerical root-finding step: solving the FOC.

D.3.3 Algorithm

The following algorithm constructs the first derivative of the value function by backward induction, which automatically satisfies the envelope condition. It finds the optimal investment from the FOC. It again constructs value function backwardly. Specifically, the following iteration procedure is my solution algorithm, followed by the conventional value function iteration:

Algorithm 1: Educated Value Function Iteration (name tentative)

(i) Guess e'.

(ii) Find $V_e^0(e, \varepsilon)$ by backward induction, given e'.

(iii) Check if e' satisfies $u_c(I - e') = \alpha V_{e'}^0(e', \varepsilon')$.

Iterate on (i)-(iii) until e' solves the FOC. Given the solved policy function, backwardly construct V.

Algorithm 2: Conventional VFI

(i) Guess V^0.

(ii) Solve for e' that maximizes the parent's Bellman equation (48).

[21]This depends on the interpolation scheme. I briefly discuss this in Appendix D.3.4

(iii) Construct $\hat{V}$ with the solved e' by backward induction and update $V^0 = \hat{V}^0$.

Iterate on (i)-(iii) until convergence $\hat{V} = V$.

While both algorithms solve the same number of maximization problems, Algorithm 1 exploits more information from the model. It leverages the fact that the EC holds regardless of e'. As a result, it does not need to repeatedly evaluate V and (i)-(iii) can be jointly implemented as a FOC rootfinding. It computes V only once at the very end, when both V_e and e' are obtained. On the contrary, Algorithm 2 evaluates V in each iteration and converges to the true solution by gradually updating V^0.

Both require some interpolation of V_e^0 (Algorithm 1) or V^0 (Algorithm 2) with respect to e. The interpolation scheme is described in the following subsection.

D.3.4 Interpolation

I employ the shape-preserving rational function spline Hermite interpolation (Cai Judd 2012). I do not describe the detail. It is inexpensive and preserves a shape of concave, monotonically increasing function. The approximated function is C^1 globally, and C^∞ on each interval. It is applicable only if value function is known to be concave and monotonically increasing. My framework meets the restriction. Popular options such as linear and Chebyshev are not suitable for my goal. I need a curvature of V_e, so linear is not a way to go. Chebyshev is not shape-preserving. it is successful in a number of applications in spite of the drawback. But mine did not work with it.

D.3.5 Numerical Example

I solve a one-region version of the model using Algorithms 1 and 2. The results are comparable, and Algorithm 1 is much faster than Algorithm 2. Both algorithms converge to the (numerically) same fixed point and Algorithm 1 can be used without concern.

Setup

- The terminal age $q_R = 4$.
- CRRA utility with risk aversion $\rho = 0.9$.
- Returns to educational investment $\eta = 0.1$
- Stochastic discount factor $\beta = 0.9$; Altruism parameter $\alpha = 0.5$
- Regional HC production efficiency $h = 1$; Exogenous wage $w = 1$
- Discretized grid of e: $[0.06, 0.4]$. unevenly distributed 50 points; clustered on the left of the interval
- Ability ε follows lognormal distribution. 16 points.

The state space for any $q \neq q_P$ is $50 \times 16 = 800$. For q_P, it is $50 \times 16^2 = 12,800$. Thus, I end up with $800 \times 3 + 12,800 = 15,200$ states. Convergence threshold is 10^{-10} for both.

Results Table D.3 reports the speed of each algorithm. Algorithm 1 takes less than a second, which is almost 300 times faster than Algorithm 2 with an educated guess, which is V^0 constructed from e'_0. The slowest version is Algorithm 2 from a blind guess, $V^0 = 0$. Algorithm 1 is as 400 times faster than it. Since Algorithm 2 converges to the same fixed point by the contraction mapping theorem, I use the results with the educated guess for the following comparison.

Figures D.1 and $D.2$ compare V^0 and e' from each algorithm. In each subplot, a higher curve is from a higher-ability parent. For e', which has an additional dimension ε', I fixed it across all curves. The differences between the algorithms are tiny.

In sum, Algorithm 1 is fast and reliable.

Table D.3: Speed of each algorithm

	Algo 1	Algo 2 (blind guess)
CPU Time (Absolute)	0.7	277.4
CPU Time (Relative)	1	396.3

Figure D.1: Model Solution: Value Function

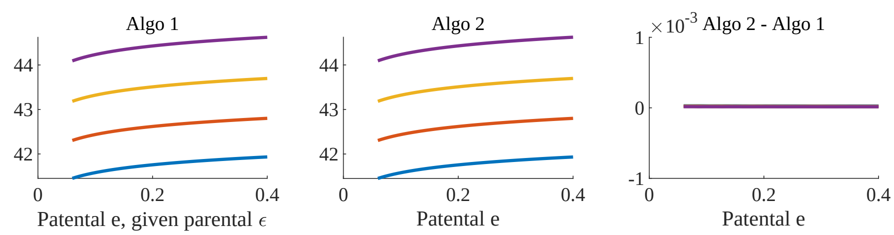
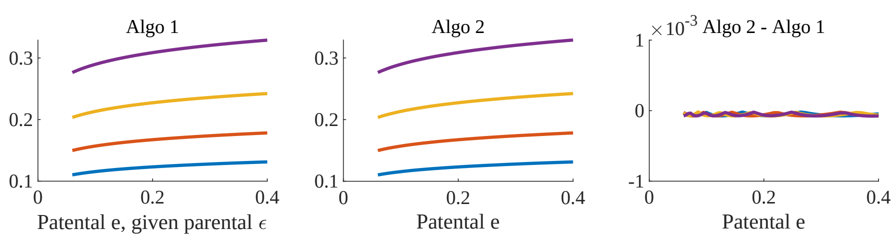


Figure D.2: Model Solution: Policy Function



D.4 Extension to the Spatial Economy

Algorithm 1 completely separates the value function and its derivative. The separation does not hold once the migration is allowed, because the derivative now involves the levels of the next period's value function due to the migration propensity. I propose a workaround based on fixed-point iteration. Although the algorithm does not work as is, a simple modification revives it.

To state the problem, I write the V_e sequence in spatial economy:

$$
\begin{aligned}
&V_e^{q_R}(k;\mathbf{s}) = u_c(I)\Delta \\
&V_e^{q_P}(k,\varepsilon';\mathbf{s}) = u_c(I-e')\Delta + \mathbb{E}\big[\beta V_e^{q_P+1}(k',\mathbf{s})\big] \\
&V_e^{q_P-1}(k,\mathbf{s}) = u_c(I)\Delta + \mathbb{E}\big[\mathbb{E}_{\varepsilon'}\beta V_e^{q_P}(k',\varepsilon';\mathbf{s})\big] \\
&V_e^{q}(k,\mathbf{s}) = u_c(I)\Delta + \mathbb{E}\big[\beta V_e^{q+1}(k';\mathbf{s})\big] \qquad \text{for} \quad q \notin \{0, q_P-1, q_P, q_R\} \\
&V_e^{0}(k,\mathbf{s}) = \mathbb{E}\big[\beta V_e^{1}(k';\mathbf{s})\big]
\end{aligned}
$$

where s $\equiv (k_B, e, \varepsilon)$ is the type-defining state vector, the expectation is taken with respect to migration, i.e., $\mathbb{E}\big[\beta V_e^{q+1}(k',\mathbf{s})\big] = \sum_{k'} m^q(k,k';\mathbf{s})\beta V_e^{q+1}(k';\mathbf{s})$, and $m^q(k,k';\mathrm{s})$ is the migration propensity. It is clear that m^q is a function of V^{q+1} (the formula is omitted), which implies we need V^{q_P}, thus V^0, to obtain V_e^q for any $q < q_P$. Therefore, the proposed strategy does not work as is.

It is useful to clarify two key properties of the finite-horizon problem. First, we can construct V from *any* guess of e' as in conventional policy iteration, without knowing V_e. Second, the V_e sequence can be calculated for any m^q, so is for any V. Such evaluation of V_e is possible because there is no feedback from V_e to V once e' is given. It is another implication of the separability that my modification leverages. If V is correct, the V_e sequence is correct as well, and we can move on to the FOC evaluation.[22]

In short, I guess e' then construct the corresponding V prior to V_e. We can use the constructed V to obtain m and V_e. The FOC can be evaluated using those. The last building block is to calculate V consistent with the guess of e'. It is manageable via fixed-point iteration. Given any policy e', the

[22]Feeding an incorrect V is not a concern. If the given V were incorrect, the generated sequence would be simply incorrect. The error stems from using m^q inconsistent with the true value function. The solver would just move on to the next guess of e' in such a case.

finite-horizon dynamic programming problem can be reformulated to an equivalent fixed-point problem:

$$V^0 = g\big(V^0\big)$$

and then to use the iteration: choose an initial guess $V^{0(0)}$, compute a sequence

$$V^{0(n+1)} = g\big(V^{0(n)}\big)$$

and obtain $V^{0(n)} \to V^0$. Here, $g(\cdot)$ is defined by a discounted sum of future values, thus increasing, concave, and differentiable. It is straightforward to show that there is a unique solution V^0 for any given e'. Moreover, this new step incurs little additional computational burden as it involves neither derivatives V_e^q nor numerical root-finding.

In sum, the FP iteration delivers the guess-consistent V. Armed with that, we can find V_e and evaluate the FOC. The following is the revised algorithm.

Algorithm 3

(i) Guess e'.

(ii) Find V by fixed-point iteration.

(iii) Find $V_e^0(e, \varepsilon)$ by backward induction, with V and e'.

(iv) Check if e' satisfies $u_c(I - e') = \alpha V_{e'}^0(e', \varepsilon')$.

Iterate on (i)-(iv) until e' solves the FOC. Both V and e' are obtained once the FOC is solved.

As in the simple case of no migration, this algorithm converges to the conventional VFI solution. The speed gain is again significant. I found Algorithm 3 is 255x faster than the VFI in an illustrative two-region case. Results are available upon request.

E. Additional Quantification Results

E.1 Tables

Table E.1: Calibration Fit: U.S. States

	Per Capita GDP (Model)	Data	Avg wage (Model)	(Data)	Pop Shares (Model)	(Data)
AL	1.413	1.413	0.714 1.081	0.714 1.081	0.017	0.017
AZ	1.727	1.727	0.639 1.231	0.639 1.231	0.017	0.017
AR	1.247	1.247	0.654 1.051	0.654 1.051	0.01	0.01
CA	1.961	1.961	0.73 1.349	0.73 1.349	0.088	0.088
CO	1.538	1.538	0.775 1.26	0.775 1.26	0.017	0.017
CT	2.286	2.286	0.818 1.394	0.818 1.394	0.013	0.013
FL	1.831	1.831	0.727 1.169	0.727 1.169	0.049	0.049
GA	1.679	1.679	0.747 1.171	0.747 1.171	0.032	0.032
HI	1.841	1.841	0.775 1.073	0.775 1.073	0.004	0.004
ID	1.258	1.258	0.729 1.047	0.729 1.047	0.005	0.005
IL	1.935	1.935	0.756 1.21	0.756 1.21	0.045	0.045
IN	1.592	1.592	0.76 1.144	0.76 1.144	0.026	0.026
IA	1.504	1.504	0.743 1.036	0.743 1.036	0.012	0.012
KS	1.318	1.318	0.718 1.098	0.718 1.098	0.011	0.011
KY	1.432	1.432	0.744 1.055	0.744 1.055	0.016	0.016
LA	1.651	1.651	0.775 1.139	0.775 1.139	0.016	0.016
ME	1.377	1.377	0.704 1.078	0.704 1.078	0.005	0.005
MD	1.976	1.976	0.837 1.388	0.837 1.388	0.019	0.019
MA	1.968	1.968	0.815 1.368	0.815 1.368	0.024	0.024
MI	1.631	1.631	0.721 1.174	0.721 1.174	0.04	0.04
MN	1.603	1.603	0.751 1.178	0.751 1.178	0.022	0.022
MS	1.225	1.225	0.674 0.968	0.674 0.968	0.01	0.01
MO	1.489	1.489	0.716 1.113	0.716 1.113	0.023	0.023
MT	1.03	1.03	0.721 0.848	0.721 0.848	0.003	0.003
NE	1.566	1.566	0.762 1.055	0.762 1.055	0.007	0.007
NV	1.977	1.977	0.815 1.19	0.815 1.19	0.007	0.007
NH	1.565	1.565	0.853 1.217	0.853 1.217	0.006	0.006
NJ	2.116	2.116	0.848 1.495	0.848 1.495	0.028	0.028
NM	1.199	1.199	0.675 1.075	0.675 1.075	0.006	0.006
NY	2.254	2.254	0.751 1.254	0.751 1.254	0.057	0.057
NC	1.713	1.713	0.668 1.09	0.668 1.09	0.033	0.033
ND	1.04	1.04	0.693 0.947	0.693 0.947	0.003	0.003
OH	1.623	1.623	0.74 1.103	0.74 1.103	0.048	0.048
OK	1.15	1.15	0.71 1.028	0.71 1.028	0.013	0.013
OR	1.514	1.514	0.728 1.138	0.728 1.138	0.013	0.013
PA	1.659	1.659	0.758 1.183	0.758 1.183	0.049	0.049
RI	1.778	1.778	0.763 1.193	0.763 1.193	0.004	0.004
SC	1.483	1.483	0.693 1.034	0.693 1.034	0.016	0.016
SD	1.435	1.435	0.705 0.972	0.705 0.972	0.003	0.003
TN	1.484	1.484	0.67 1.046	0.67 1.046	0.023	0.023
TX	1.699	1.699	0.704 1.169	0.704 1.169	0.069	0.069
UT	1.445	1.445	0.74 1.145	0.74 1.145	0.007	0.007
VT	1.204	1.204	0.782 1.036	0.782 1.036	0.003	0.003
VA	1.969	1.969	0.797 1.331	0.797 1.331	0.028	0.028
WA	1.998	1.998	0.84 1.213	0.84 1.213	0.022	0.022
WV	1.277	1.277	0.702 1.075	0.702 1.075	0.007	0.007
WI	1.456	1.456	0.798 1.112	0.798 1.112	0.024	0.024

All state-level moments are matched exactly.

Table E.2: Model Validation: TFP and Amenities

	TFP	Amenities	
	Herkenhoff et al. (2018)	Herkenhoff et al. (2018)	Amenities, Others
1980	0.280	-0.225	0.500
	(0.060)	(0.132)	(0.000)
1990	0.703	-0.076	0.415
	(0.000)	(0.615)	(0.004)
2000	0.743	-0.321	0.176
	(0.000)	(0.030)	(0.235)
2014	0.580	-0.550	
	(0.000)	(0.000)	

Spearman rank correlation coefficients. P-values in parentheses. For "Others", the 1980 and 1990 estimates are taken from Gabriel et al. (2003). The 2000 estimate is taken from Albouy (2008).

E.2 Figures

Figure E.1: ANI by Education: Levels

(a) High School or Less (b) Some College or More

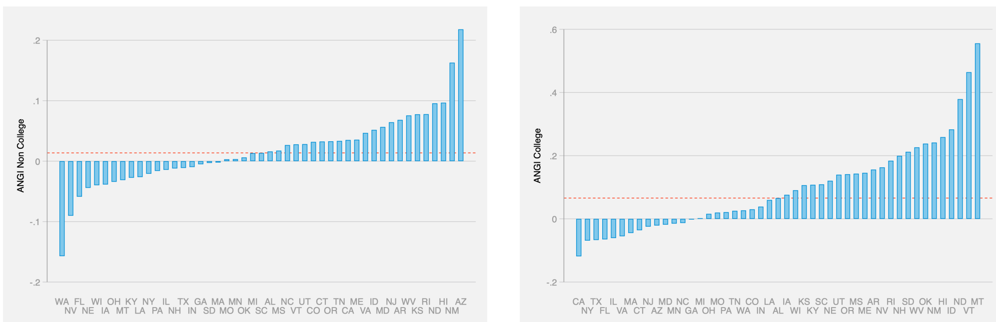


Figure E.2: RHCPC by Education: Levels

(a) High School or Less (b) Some College or More