Mobile Human Capital and Diffusion of Ideas Across Cities*

Shogo Sakabe†

November 11, 2022

Latest version available here

Abstract

I study how the internal migration of inventors affects local and aggregate growth through technological diffusion across cities. I propose a quantitative spatial theory of growth and knowledge diffusion through internal migration. My model highlights two mechanisms by which productivity growth can be higher in one city than in another: (1) agglomeration forces and (2) knowledge inflows through internal migration. Using data on U.S. cities, I find that the effect of knowledge diffusion explains approximately 40 percent of the spatial variation in productivity changes. I quantify the dynamic effects of place-based policies and find that reducing migration costs toward a small number of cities can improve aggregate efficiency while reducing spatial disparities in productivity across cities.

*I am grateful to David E. Weinstein, Donald R. Davis and Réka Juhász for their continued, patient guidance and generous support. I thank David Atkin, Vinayak Iyer, Sang Hoon Kong, Cameron LaPoint, Yuhei Miyauchi, Suresh Naidu, Diana Van Patten, Swapnika Rachapalli, Conor Walsh, Howard Zhang, and seminar participants at Columbia University for their helpful comments and suggestions. I thank the Center on Japanese Economy and Business (CJEB) at Columbia Business School for supporting this project by providing dissertation fellowships and the Nakajima Foundation for generous financial support.

†Department of Economics, Columbia University. Email: ss5122@columbia.edu; Website: https://shogosakabe.github.io/

1 Introduction

Innovation has been strikingly concentrated in a few cities, and the geographical concentration of highly skilled workers has played an important role in shaping spatial variation in innovation and economic growth (Carlino and Kerr, 2015). The unequal distribution of innovation and growth opportunities across regions has spurred policy discussions about how to attract high-skilled workers to areas that are not leading centers of technology. However, several studies indicate that place-based policies would have negative effects on aggregate efficiency (Moretti, 2021) and even spatial equality (Gaubert, 2018) in the presence of agglomeration externalities. In this paper, I explore the role of the internal migration of inventors and as an additional mechanism, knowledge spillovers across cities, both of which affect spatial variation in innovation and the effectiveness of place-based policies.

What is the effect of internal migration of high-skilled workers on spatial variation in growth? How do migration flows shape knowledge flows across cities? What is the effect of place-based policies that attract workers to particular locations? To answer these questions, I develop a quantitative spatial equilibrium theory of innovation, knowledge diffusion, and migration across cities.

I use a theoretical framework to structurally estimate the impact of the internal migration of inventors on knowledge diffusion and spatial growth. Based on the estimated model, I decompose the cross-sectional variation in productivity changes across cities into (1) the effect of density, or agglomeration externalities, on the generation of higher quality ideas and (2) gains from better access to technology through knowledge diffusion. My model implies that more than 40 percent of the spatial variation in productivity changes is explained by knowledge spillovers, and the remaining variation is attributable to agglomeration forces. This indicates that compared to other cities, denser cities have benefited more from agglomeration forces, as well as from knowledge spillovers from other cities. Lastly, I study the effectiveness of place-based policies that aim to attract workers to specific locations. I find that policies focused on reducing migration costs toward productive cities can simultaneously improve aggregate productivity and reduce spatial disparities by targeting cities that may not offer the best products but are experiencing dense migration inflows from other cities.

I model knowledge diffusion across locations as the outcome of interactions among producers in different industries or cities. In each city, a continuum of goods is produced. There are many potential producers of goods, and these producers have varying levels of productivity. The frontier of knowledge in a city is characterized by the productivity distribution of local producers who operate within the city. The average efficiency of producing goods in a city is summarized by its local stock of knowledge, and it evolves from a stochastic diffusion process. To make the diffusion process tractable, I extend the framework proposed by Buera and Oberfield (2020). which is compatible with the workhorse Ricardian trade model in Eaton and Kortum (2002). A producer draws her original idea from a city-specific distribution and combines it with knowledge acquired from other producers within a city and across cities. When learning from other producers, the quality of insights is subject to diminishing returns, implying that learning at more distinct locations leads to a higher rate of innovation, conditional on the stock of knowledge in those locations. I also allow for the efficiency of learning to depend on the inventor's current city as well as the city from which she is absorbing insights and knowledge. Modeling heterogeneity in the efficiency of learning is required to rationalize the observed proxy for the knowledge flows described below.

I estimate values for the local stock of knowledge and heterogeneity in the efficiency of learning that rationalizes the observed data on U.S. commuting zones (CZs) during 1976-2015. To do this, I incorporate the framework of knowledge diffusion into a quantitative spatial model that was recently developed in the literature (Redding, 2016; Monte et al., 2018). Workers live in a city for one period, and their children choose their own city in the next period. The children consider the idiosyncratic preference shocks, the spatial distribution of real income, and the costs of migrating from their parents' original location. I use data on wages, population, and land area at the CZ level to recover unobservable values for the stock of knowledge. I show that the parameters that govern heterogeneity in the efficiency of learning across cities can be inverted from the structure of the model using patent citations across locations.

This study is novel in several ways. First, accepting the model's implication of evolution in the stock of knowledge, I structurally estimate the impact of the following on local productivity changes: (1) population density in a city and (2) knowledge spillovers across U.S. cities. This enables me to investigate the separate contributions of agglomeration forces and knowledge diffusion to innovation. Importantly, I distinguish between the effect of intercity learning and agglomeration externalities in denser cities in relation to several mechanisms. These include local learning, matching, and sharing, as summarized in Duranton and Puga (2004), as well as innovation and patenting activity (Carlino et al., 2007; Moretti, 2021). In my empirical analysis, which will be described later in the paper, I show that the quality of original ideas drawn from a city-specific distribution is increasing in density, implying that agglomeration externality is generating new ideas, as documented in the literature. Furthermore, I show that inventors in a city tend to learn more from cities that send inventor migrants to the original inventors' city. Viewing this phenomenon through the lens of the model, it appears that migration-flow-related variation in producer composition affects the quality of insights. Since local producers combine their own fresh ideas with insights absorbed from other cities to develop new production technology, a complementarity exists between the quality of new ideas and the efficiency of intercity learning. The model quantifies that knowledge spillovers explain approximately 40 percent of the cross-sectional changes in the evolution of the stock of knowledge, and agglomeration forces explain the remaining 60 percent. In denser cities, both effects are stronger, and the effects complement each other to stimulate more rapid growth. This has an important implication for place-based policies. Relocating workers to cities where they gain more from intercity knowledge spillovers would improve the aggregate stock of knowledge. I examine this implication by simulating place-based policies that attract workers in specific locations.

Estimating the impact of inventor migration on knowledge flow is challenging. For example, concerned about endogeneity, firms may actively recruit inventors from cities whose knowledge they are seeking. To overcome endogeneity issues, I employ an instrumental-variable (IV) approach using violent crime rate changes and migration flows predicted by historical migration as instrumental variables for inventor migration flows across cities. The identification assumption is that these instruments affect labor flows across cities but are not related to current technology flows. The estimates indicate that doubling a share of migration flows from a given city leads to approximately a 54 percent increase in the patent citations attributed to that city. Using the crime rate changes as an instrument, I also estimate the elasticity of original ideas with respect to the density that relates to the generation of new ideas and agglomeration elasticity.

The second way in which the study is novel is that I explore the effect of spatial policies and consider the interaction between agglomeration externalities and knowledge spillovers that result from migration. In the model, the elasticity of productivity with respect to density would not be constant across cities due to the endogenous effect of migration on intercity knowledge spillovers. This is in contrast to studies on place-based policies with symmetric elasticity, such as (Glaeser and Gottlieb, 2009; Kline and Moretti, 2014), suggesting that implementing policies affecting workers' location choice will not result in any gains. It also differs from Fajgelbaum and Gaubert (2020), which investigated the static effect of place-based policies with spillovers across heterogeneous workers within a city. Instead, this paper estimates the dynamic effect of place-based policies with knowledge spillovers that result from migration across cities. The model implies that the aggregate effects of place-based labor subsidies are negligible, and local effects are also weak. I show that a reduction of migration costs toward a small set of cities can improve aggregate efficiency and equity.

Relation to the Literature

This paper is related to the literature on knowledge diffusion and technological adoption across countries and cities, including Kortum (1997), Eaton and Kortum (1999), Lind and Ramondo (2018), Buera and Oberfield (2020), Perla et al. (2021), and Liu and Ma (2021). My contribution to these studies is to allow for labor mobility across regions in the model and to quantify the effect of intercity spillovers through random encounters between workers migrating from different locations.

Kerr (2010) showed that the spatial reallocation of technology clusters following breakthrough inventions depends heavily on the mobility of the technology's labor force. Recent papers on the effects of immigration on local and aggregate innovation, including Arkolakis et al. (2020) and Prato (2021), have highlighted the importance of international migration as a mechanism for fostering knowledge spillovers and productivity growth. My work complements this line of inquiry by providing a theoretical framework featuring "diversification effects," whereby cities can grow by learning from more distinct locations and technologies on top of the scale effects and the productivity effects. Roca and Puga (2017) found that denser cities offer more valuable learning experiences for workers, and workers bring the dynamic gains from working in these cities with them when they relocate to other cities. In this study, I show that inventor relocation induces knowledge spillovers to workers in the destination city.

This study also contributes to literature on spatial growth models (Desmet and Rossi-Hansberg (2014); Desmet et al. (2018); Sollaci (2022); Nagy (2021); Berkes et al. (2022); Cai et al. (2022)). My study builds on this body of work by developing a model of endogenous knowledge diffusion through migration and studying the heterogeneous effects of this knowledge diffusion across cities.

This research also relates to voluminous literature on localized knowledge spillovers, including Jaffe et al. (1993), Glaeser (1999), Feldman and Audretsch (1999), Duranton and Puga (2001), Thompson and Fox-Kean (2005), Moser (2011), Kerr and Kominers (2015), Davis and Dingel (2019), and Moretti (2021). Notably, paper trails of patent citations demonstrate that inventors in large cities can absorb insights from very distant cities; this holds true for the geographical pattern of internal migration. This paper quantifies the role of knowledge spillovers that are shaped by both geography and labor movement across cities.

The rest of the paper is organized as follows. Section 2 develops the model. Section 3 describes the data. Section 4 desbribes estimation of the model. Section 5 explores the effect of place-based policies using the estimated model. Section 6 concludes the paper.

2 Model

This section develops a spatial equilibrium model for idea diffusion across cities. I build on Buera and Oberfield (2020)'s framework of technology and the Ricardian trade model with labor mobility provided by Redding (2016). In Section 4.3.1, I provide evidence for how inventor migration across cities affects intercity idea diffusion.

2.1 Idea Diffusion Across Cities

Consider an economy that consists of cities indexed by $i, j \in N$. Each city produces a continuum of goods $v \in [0, 1]$. For each good, there are many potential producers with different productivity levels q. The production function is given by

$$y(v) = q \cdot l(v) \tag{2.1}$$

where $l(v)$ is the labor input and $y(v)$ is the output of good v. In each city i at time t, productivity is drawn from a distribution $F_{i,t}(q)$. The parameter θ governs the productivity dispersion across goods.

Each period t, new ideas arrive at potential producers of each good in each city through random interactions between producers. A producer in city i adopts the most productive idea available to her. Each idea combines two random components. The first component is the producer's original idea drawn from a local idea distribution. I assume that the arrival rate of original ideas greater than ϵ is $G_{i,t}(\epsilon)$. The second component is an insight z_j drawn from the productivity distribution of each city $j = 1, 2, ...N$ through random encounters with a rate $\kappa_{ij,t}$. The productivity of the insight drawn from city j is given by $\epsilon \kappa_{ij} z_j^{\beta}$ for $\beta \in [0, 1)$. For the producer, the most productive new idea q' is

$$q' = \max_j \left\{ \epsilon \kappa_{ij,t} z_j^{\beta} \right\}.$$

The producer adopts the new idea if it is greater than her original idea, that is, $q' > q$.

2.1.1 Evolution of Productivity Distributions

Given the frontier of knowledge $F_{i,t}(q)$ and the local idea distribution $G_{i,t}(\epsilon)$, the frontier of knowledge at time $t+\Delta$ is expressed as follows:

$$1 - F_{i,t+1}(q) = [1 - F_{i,t}(q)] + F_{i,t}(q) \sum_{j \in N} \int_{t}^{t+\Delta} \int G_{i,\tau}\left(\frac{q}{\kappa_{ij,t} z^{\beta}}\right) dF_{j,\tau}(z)\, d\tau.$$

Rearranging and taking the limit as $\Delta \to 0$, the productivity distribution evolves according to the following:

$$\frac{d}{dt} \ln F_{i,t}(q) = -\sum_{j \in N} \int_{0}^{\infty} G_{i,t}\left(\frac{q}{\kappa_{ij,t} z_j^{\beta}}\right) dF_{j,t}(z_j).$$

I assume that the city-specific arrival rate of original ideas follows a power law. The CDF of the original ideas in city i is given by the following:

$$G_{i,t}(\epsilon) = 1 - \zeta_{i,t} \epsilon^{-\theta}, \tag{2.2}$$

and the initial productivity distribution for each city i follows a Fréchet distribution

$$F_{i,0}(q) = e^{-A_{i,0} q^{-\theta}} \quad A_{i,0} > 0,\ \theta > 1,$$

where $A_{i,t}$ is the scale parameter of city i's distribution at time t that evolves endogenously over time arising from intercity idea diffusion. Under these assumptions, I obtain

$$\frac{d}{dt} A_{i,t} = \zeta_{i,t} \Gamma(1-\beta) \sum_{j \in N} \left(K_{ij,t} A_{j,t}\right)^{\beta}, \tag{2.3}$$

where $\Gamma(\cdot)$ denotes the gamma function, and $K_{ij,t} \equiv \kappa_{ij,t}^{\theta/\beta}$.

2.1.2 Learning Across Cities

Conditional on adopting the new ideas, the probability that a producer in location i builds upon an insight from location j is as follows:

$$\phi_{ij,t} = \frac{K_{ij,t}A_{j,t}}{\sum_{k\in N} K_{ik,t}A_{k,t}} \tag{2.4}$$

See Section A.1 for the derivation. In the empirical analysis, I use observed patent citation shares as a proxy for $\{\phi_{ij,t}\}$.

2.2 Preferences and Income

In this subsection, I describe the model that extends the Ricardian trade model with labor mobility developed by Redding (2016). Each city $i \in N$ has its city-specific land supply, productivity, and amenities, and it differs from other cities in its geographical location. Each city i is endowed with a fixed supply of land H_i. Let $L_{i,t}$ denote the number of workers in city i at time t, and each worker is endowed with one unit of labor that is supplied inelastically. The total population of the economy as a whole $L_t = \sum_{i\in N} L_{i,t}$ evolves at an exogenous rate of $n_t = L_t/L_{t-1}$.

A worker who lives in city i has the following utility function:

$$U_{i,t} = b_{i,t}\left(\frac{C_{i,t}}{\alpha}\right)^{\alpha}\left(\frac{H_{i,t}}{1-\alpha}\right)^{1-\alpha}, \ \alpha \in (0,1), \tag{2.5}$$

where $C_{i,t}$ is the consumption of final goods, and $H_{i,t}$ is the residential land use. The idiosyncratic amenity shocks $b_{i,t}$ account for heterogeneous preferences for living in city i. I assume that the amenity shocks are drawn independently across cities and workers a Fréchet distribution:

$$J_{j,t}(b) = e^{-B_{j,t}b^{-\epsilon}} \ B_{j,t} > 0, \ \eta > 1. \tag{2.6}$$

The goods consumption index in city i takes the form of a constant elasticity of substitution (CES) over a continuum of goods $v \in [0,1]$:

$$C_{i,t} = \left[\int_0^1 c_{i,t}(v)^{\frac{\sigma-1}{\sigma}} dv \right]^{\frac{\sigma}{\sigma-1}}, \tag{2.7}$$

where σ is the elasticity of substitution between goods. The corresponding price index for goods consumption $P_{i,t}$ is given by the following:

$$P_{i,t} = \left[\int_0^1 p_{i,t}(v)^{1-\sigma} dv \right]^{\frac{1}{1-\sigma}}.$$

All workers residing in city i receive the same wage and make the same consumption and residential land choices. Utility maximization implies that workers spend a fraction $(1-\alpha)$ of their income on residential land. I assume that the expenditure on land in each city is redistributed in a lump sum to the workers residing there. The indirect utility is given by the following:

$$U_{i,t}(\omega) = \frac{v_{i,t}}{P_{i,t}^{\alpha} r_{i,t}^{1-\alpha}} b_{i,t}$$

where $v_{i,t}$ is income, which is the sum of the local wage and income from land rents in city i at time t. $r_{i,t}$ is the land rent for city i at time t.

Then, total income in each location equals the sum of the labor income and expenditure on residential land:

$$v_{i,t} L_{i,t} = w_{i,t} L_{i,t} + (1-\alpha) v_{i,t} L_{i,t} = \frac{w_{i,t} L_{i,t}}{\alpha} \tag{2.8}$$

Land market clearing requires that land income equals expenditure:

$$r_{i,t} = \frac{(1-\alpha) v_{i,t} L_{i,t}}{H_i} = \frac{1-\alpha}{\alpha} \frac{w_{i,t} L_{i,t}}{H_i} \tag{2.9}$$

2.3 Migration

At time $t-1$, $L_{i,t-1}$ adults reside in city i, and they have n_t child each. At the beginning of the period, t, and before the realization of idea diffusion, the children enter adulthood and choose where they want to reside to maximize their adult welfare. Migrating from city i to j costs $\mu_{ij,t} \geq 1$[1]. I also allow for idiosyncratic preference shocks related to how much each child values residing in each location in adulthood. A child who lives in city i chooses her location in the next period, according to the following:

$$U_{ij,t} = \max_j \frac{U_{j,t}}{\mu_{ij,t}} b_{j,t}$$

Each worker chooses the city that offers her the highest utility after observing her idiosyncratic amenity shock. The probability that a worker moves to city j conditional on living in location i in childhood is as follows:

$$\lambda_{ij,t} = \frac{\mathcal{B}_{ij,t} \left(v_{j,t} / P_{j,t}^{\alpha} r_{j,t}^{1-\alpha} \right)^{\epsilon}}{\sum_{k \in N} \mathcal{B}_{ik,t} \left(v_{k,t} / P_{k,t}^{\alpha} r_{k,t}^{1-\alpha} \right)^{\epsilon}} \tag{2.10}$$

where $\mathcal{B}_{ij,t} \equiv B_{j,t} \mu_{ij,t}^{-\epsilon}$ denotes the ease of migration from city i to j. For city pairs with zero migration flows, I assume that the migration cost is prohibitively high[2]. The measure of people moving from city i to j is given by the following:

$$L_{ij,t} = n_t \lambda_{ij,t} L_{i,t-1} \tag{2.11}$$

Summing (2.10) across origin cities, the measure of children who choose to live in a city i in their adulthood is given by $L_{j,t} = \sum_{k \in N} L_{kj,t}$.

In the presence of migration costs, welfare and real income are not equalized across cities. The local expected utility in a city can be written as a function of the weighted average of real income across cities with the weight being the ease of migration from that city to another city:

$$U_{i,t} = \delta \left[\sum_{j \in N} \mathcal{B}_{ij,t} \left(\frac{v_{j,t}}{P_{j,t}^{\alpha} r_{j,t}^{1-\alpha}} \right)^{\epsilon} \right]^{\frac{1}{\epsilon}} \tag{2.12}$$

where $\delta \equiv \Gamma\left(\left(\epsilon - 1 \right) / \epsilon \right)$ and $\Gamma\left(\cdot \right)$ denotes the Gamma function. Without migration costs, that is $\mu_{ij,t}^{-\epsilon} = 1, \forall i, j \in N$, the ease of migration becomes $\mathcal{B}_{ij,t} = B_{j,t} \cdot \mu_{ij,t}^{-\epsilon} = B_{j,t}$, which leads to the equalization of welfare across cities. With migration costs, workers in a city

[1]This nests the case of freely mobile workers, $\mu_{ij,t} = 1, \forall i, j \in N$,

[2]In calibration, I set $\mathcal{B}_{ij,t} = 0$ for city pairs (i, j) with $\lambda_{ij,t} = 0$.

with better migration options have higher expected utility.

2.4 Goods Trade

In each city, individual goods can be produced through labor by many producers according to a linear technology in (2.1). For a producer with productivity q in city j, the cost of providing one unit of the good to city i is as follows:

$$\frac{d_{ij}w_{j,t}}{q}$$

Producers engage in perfect competition. In equilibrium, the price index is given by the following:

$$P_{i,t} \propto \left(\frac{A_{i,t}}{\pi_{ii,t}}\right)^{-\frac{1}{\theta}} w_{i,t} \tag{2.13}$$

I assume that $\theta > \sigma - 1$, so that the price level is finite.

2.5 General Equilibrium

An equilibrium in this economy is characterized by a vector of prices and allocations, such that goods and factor markets clear in all periods. Formally, given the initial population $\{L_{i,0}\}$, the initial stock of knowledge $\{A_{i,0}\}$, geography characteristics $\{H_i, \mathcal{B}_{i,t}, K_{ij,t}, d_{ij}\}$, and a given path for the productivity vector $\{A_{i,t}\}$, an equilibrium is a vector of endogenous variables $\{L_{i,t}, \pi_{ij,t}, w_{i,t}\}$ that solve the following system of equations for all cities $i, j \in N$ and time periods t:

1. Each city's income must be equal to the expenditure for the goods produced in that city:

$$w_{i,t}L_{i,t} = \sum_{k\in N} \pi_{ki,t}w_{k,t}L_{k,t} \tag{2.14}$$

2. The share of expenditure of location i on goods produced by location j is expressed as follows:

$$\pi_{ij,t} = \frac{A_{j,t}\left(d_{ij}w_{j,t}\right)^{-\theta}}{\sum_{k\in N} A_{k,t}\left(d_{ik}w_{k,t}\right)^{-\theta}} \tag{2.15}$$

3. Each city's population is equal to the population arriving in that location. In other words, $L_{i,t} = \sum_{k\in N} L_{ki,t}$. From (2.10) and (2.11), and substituting (2.8) and (2.9), this implies that

$$L_{i,t} = n_t \sum_{k\in N} \left(\frac{\Phi_{ki,t}}{\Phi_{k,t}}\right) L_{k,t-1} \tag{2.16}$$

where

$$\Phi_{ki,t} \equiv \mathcal{B}_{ki,t}\left(\frac{A_{i,t}}{\pi_{ii,t}}\right)^{\frac{\alpha\epsilon}{\theta}}\left(\frac{L_{i,t}}{H_i}\right)^{-\epsilon(1-\alpha)}, \quad \Phi_{k,t} \equiv \sum_{j\in N} \Phi_{kj,t}.$$

The term $\Phi_{k,t}$ captures the attractiveness of migration options for children who are born in city k at time $t-1$.

Given parameters $\{\alpha, \sigma, \theta, \epsilon\}$ and local fundamentals $\{H_i, A_{i,t}, B_{i,,t}, d_{ij}, \mu_{ij,t}\}$, there exist a unique vector of wages and populations $\{w_{i,t}, L_{i,t}\}$ and a unique matrix of trade shares $\{\pi_{ij,t}\}$ that solve equations (2.14)-(2.16), as formally shown in Section A.3.

3 Data

I combine data from several U.S. sources. I use county-level employment $\{L_{i,t}\}$ and wages $\{w_{i,t}\}$ data from the Bureau of Economic Analysis (BEA) for the period 1976-2015. I aggregate county-level location data at the CZ level in 2000, based on Autor & Dorn (2013). I combine these datasets with Geographical Information Systems data to compute the bilateral distance between centroids of CZs. Using these data sources and data described below, I construct panel data for CZs at a 10-year frequency and estimate the model at the CZ level. I focus on the continental U.S. CZs. In other words, I do not include Hawaii, Alaska, Puerto Rico, or other U.S. territories. Hereafter, I will refer to CZs as cities.

3.1 Patent Citations and Internal Migration of Inventors

I use data of granted patents and patent application files from USPTO PatentsView, which covers the universe of U.S. patent filings and citations during 1976-2015. In the following, I describe how I identify the residential location of inventors and patent citations across CZs.

An inventor's location is based on the inventor's hometown information in the filing year of a patent. I exclude inventors filing a foreign address as their residential location. Let $L^R_{ij,t}$ denote the number of inventors migrating from city i to j in time t. Inventor migration is observed when the inventor files a patent and reports a residential location that differs from the residential location reported in the previously filed patent. The limitation of this approach is that I do not observe the exact timing of the inventors' migrations. Since I can only observe the residential location of inventors when they file a patent, I assume that the inventor did not move from the original location until she files another patent with different location information. To mitigate measurement errors due to a lack of precise information about the timing of inventor relocation, I aggregate the data according to a 10-year frequency. The number of inventors identified in this data is 701,110. Of these, 20 percent changed their residential location at least once in their patenting history[3][4]. Let $\psi^R_{ij,t}$ denote the share of inventor migrants from city i to city j out of employment in city j in period t, defined as[5]:

$$\psi^R_{ij,t} \equiv \frac{L^R_{ij,t}}{\sum_k L^R_{kj,t}}.$$

I construct CZ-to-CZ patent citation shares based on the residential location of each inventor in the following manner. To measure the effect of inventor migration on learning patterns across cities, I focus on the patent citations made by stayers, in other words inventors who never moved during their patenting career[6]. Suppose a patent p is filed by an inventor residing in city i. This patent cites another patent p' granted to an inventor in city i'. I count it as one citation to city i' made by city i. If a citing patent or a cited patent is filed by inventors residing in different locations, I assign the citation to each

[3]About 17% of movers returned to a CZ where they resided at least once during their patenting history, which resembles the return rate described by Prato (2021) for international migration between the U.S. and the European Union.

[4]In the migration data I used, 85% of inventor-year observations have information on the assignee. Among inventors with assignee information, such as their associated firms or institutions, approximately 75% of inventor-year migrations are associated with changing assignee.

[5]Inventors who never moved from city i in period t are counted as stayers, that is, $L^R_{ii,t}$.

[6]If an inventor only filed one patent, I assume she is a stayer.

citing-cited location pair. Then I count the number of citations made by city i to city j in period t, denoted by $c_{ij,t}$.

Patent citation shares across cities can inform us about knowledge diffusion conditional on innovation. I use patent citation shares as an empirical proxy for the bilateral learning probability $\{\phi_{ij,t}\}$ in (2.4). A patent citation share in city i to city j is defined as

$$\phi_{ij,t} = \frac{c_{ij,t}}{\sum_j c_{ij,t}}.$$

In section 4.3, I estimate the relationship between the patent citation shares and the inventor migration shares to inform the parameters $\{K_{ij,t}\}$ in (2.4).

3.2 Internal Migration of General People Across U.S. Cities

I construct annual migration flows data for general people between CZs from county-to-county migration flows provided by the Internal Revenue Service (IRS) 1990-2010, which covers more than 90% of the U.S. population (Schubert, 2021)[7]. I use migration flow data from the IRS as a proxy for migration shares of general people, $\{\lambda_{ij,t}\}$, to estimate the ease of migration $\{\mathbb{B}_{ij,t}\}$ to conduct counterfactual analysis in Section 5.

4 Estimation of the Model

This section describes an estimation procedure for model parameters, given the observed data. I fit the model to data from the period 1976-2015 using the average wage $\{w_{i,t}\}$ and employment $\{L_{i,t}\}$ data from BEA. For land supplies $\{H_i\}$, I use land areas at a CZ level. I estimate the model as follows. First, I choose central values for some parameters in the model based on the existing empirical literature. Second, I show I can invert the productivity $\{A_{i,t}\}$, the ease of migration $\{\mathbb{B}_{ij,t}\}$, and the bilateral learning parameters $\{K_{ij,t}\}$ given the assumed values of the parameters and observed data. Third, I structurally estimate the remaining parameters in the model that relate local productiv-

[7]Bilateral migration flows are recorded for county pairs with more than 10 tax returns and exemptions filed by movers.

ity changes to agglomeration externalities and the effect of knowledge diffusion through internal migration.

4.1 Assumptions About Parameter Values

The objective of this step is to choose the parameter values required to invert the model. For the share of final-goods consumption in residential consumption expenditure in (2.5), I use $\alpha = 0.75$, which is in line with Davis and Ortalo-Magné (2011). I assume the elasticity of substitution in (2.7) is $\sigma = 4$, which is consistent with literature such as Broda and Weinstein (2006). The amenity shape parameter ε in 2.6 is set to be $\varepsilon = 3$, which is in line with a study on the U.S. and Indonesia Bryan and Morten (2019).

I parameterize bilateral trade costs as a function of distance: $d_{ij}^{-\theta} = dist_{ij}^{-\theta\omega}$ where the composite parameter $(-\theta\omega)$ is the elasticity of trade flows with respect to distance. I use $(-\theta\omega) = -1.29$, which is the estimate of the gravity equations for the U.S. CZs provided by Monte et al. (2018). I assume $\theta = 4$, which is in line with the value estimated in Simonovska and Waugh (2014) and satisfies the assumption of $\theta > \sigma - 1$ for the price index in (2.13) to be well-defined. For the assumed value of θ, the elasticity of trade cost with respect to distance is $\omega = 0.32$.

4.2 Model Inversion

Given the assumptions about the parameter values $\{\alpha, \sigma, \theta, \omega, \varepsilon\}$ and observed data, I invert the productivity parameters $\{A_{i,t}\}$, the ease of migration $\{\mathcal{B}_{ij,t}\}$, and the bilateral learning rates $\{K_{ij,t}\}$ that rationalize the data. Proposition 1 below formalizes the inversion of the parameters.

Proposition 1. *There exist unique values (up to scale) of the stocks of knowledge $\{A_{i,t}\}$, the ease of migration $\{\mathcal{B}_{ij,t}\}$, and the bilateral learning parameters $\{K_{ij,t}\}$ that are consistent with data, given*

1. The model parameters $\{\alpha, \sigma, \theta, \omega, \varepsilon\}$ and parameterized trade costs $\{d_{ij}\}$.

2. Data on populations, wages, land supplies $\{L_{i,t}, w_{i,t}, H_i\}$, the probability of bilateral learning $\{\phi_{ij,t}\}$, and migration flow shares $\{\lambda_{ij,t}\}$.

Proof. See Section A.4. □

4.3 Estimation

The goal of this subsection is to estimate the value of β in (2.3). I consider the discrete analog of (2.3) when fitting it to the data[8]:

$$A_{i,t+1} - A_{i,t} = \zeta_{i,t}\Gamma\left(1-\beta\right)\sum_{j\in N}\left(K_{ij,t}A_{j,t}\right)^{\beta} \tag{4.1}$$

Given the assumed values of parameters and observed data, I invert local productivities $\{A_{i,t}, A_{i,t+1}\}$ based on 1[9]. In the following, I make structural assumptions that relate changes in the stock of knowledge to the agglomeration forces and internal migration of inventors. I then estimate β based on the structure of the model.

4.3.1 Migration and Knowledge Diffusion

I assume that the bilateral learning parameters $\{K_{ij,t}\}$ depend on migration inflows of inventors in the past period. The model proposes that children inherit knowledge and technology from their parents and bring their insights with them to whatever city they inhabit in adulthood. This assumption is given by

$$K_{ij,t} = \left(\psi^{R}_{ji,t-1}\right)^{\gamma}\varepsilon^{K}_{ij,t}, \tag{4.2}$$

where $\psi^{R}_{ij,t} = L^{R}_{ij,t}/L^{R}_{j,t}$ is the in-migration share of inventors moved from i to j out of j's population, and $\varepsilon^{K}_{ij,t}$ is a residual term. The parameter γ denotes the elasticity of knowledge flow with respect to inventor migration. Taking logs of both sides in (2.4) and

[8] A recent paper by Cai et al. (2022) provides a discrete-time generalization of the continuous-time representation of the evolution of the stock of knowledge in Buera and Oberfield (2020).

[9] I choose the normalization of $\{A_{i,t+1}\}$ such that $\min_i\{A_{i,t+1}/A_{i,t}\} = 1$. It ensures that the left hand side of (4.1) is non-negative.

using (4.2), I obtain the following:

$$\ln\left(\phi_{ij,t}\right) = \gamma \ln\left(\psi^{R}_{ji,t-1}\right) + \ln\left(\sum_{k\in N} K_{ik,t}A_{k,t}\right) + \ln A_{j,t} + \ln \varepsilon^{K}_{ij,t} \tag{4.3}$$

(4.3) relates patent citation shares to inventor in-migration shares.

I estimate the following regression

$$\ln\left(\phi_{ij,t}\right) = \gamma \ln\left(\psi^{R}_{ji,t-1}\right) + \delta_{i,t} + \delta_{j,t} + \varepsilon_{ij,t} \tag{4.4}$$

where the origin-time fixed effects include the local productivity $(A_{j,t})$ of origin city j. The destination-time fixed effects are $\delta_{i,t} \equiv \ln\left(\sum_{k\in N} K_{ik,t}A_{k,t}\right)$, and $\epsilon_{ij,t} \equiv \ln \varepsilon^{K}_{ij,t}$ is an error term. I use a 10-year panel of data on patent citations and inventor migration[10] in the following analysis.

While the fixed effects control for the city-level time-varying characteristics, estimating 4.4 by OLS may still be problematic because of potential endogeneity. For example, if firms in destination city i actively recruit inventors from the origin city j whose knowledge they are seeking, this could lead to reverse causality. For this reason, I instrument $\ln\left(\psi^{R}_{ij,t}\right)$ with a set of variables that would be independent of the current citation patterns between cities.

The first instrument is a change in the log of the violent crime rate $\Delta \ln\left(Crime_{j,t}\right) \equiv \ln Crime_{j,t} - \ln Crime_{j,t-1}$, where $Crime_{j,t}$ denotes the violent crime rate per 100,000 residents in the origin city j. I use violent crime rates during 1966-2015 from the Federal Bureau of Investigation Uniform Crime Reporting (UCR) Program Data, compiled by Kaplan (2021). The number of crimes is reported at an agency level, with some agencies covering multiple counties. In the main analysis, I use data on agencies whose primary location is in a single county and agency-years when the full 12 months of crime were reported to the UCR[11]. I then aggregate the number of violent crimes and population

[10] I choose a 10-year window for ”citation lags.” For example, for a patent filed in 2006, I only count citations made of other patents granted during 1996-2006.

[11] Among 24,578 agencies, 17,332 agencies satisfy these criteria. The results do not change when I use all agencies. For more details, see Appendix.

at a CZ level and compute the violent crime rates per 100,000 residents. Violent crimes include murder, assault, and robbery. Figure 1 shows the average log crime rates during 1966-1975 and the average log crime rate changes during 1976-2015.

The effect of a city's crime rate changes on its migration inflows can be positive or negative due to two offsetting effects. An increase in the crime rate in city i can be a push factor for emigrants if they are willing to move in response to declining local amenities. This effect would lead to an increase in the migration share of city i, or $\psi^R_{ij,t}$ in another city j. However, a decline in local amenities can also affect the initial location choice of workers. In the model, children choose their residence when they reach adulthood, and they prefer cities with better amenities, conditional on real income and migration costs. As a result, an increase in the crime rate exerts a downward pressure on the number of workers residing in the city, $L^R_{i,t}$. This would negatively affect the migration share of city i in other locations. These two effects determine the net effect of crime rate changes on migration flows.

The second instrument is the share of migrants predicted by the historical share of bilateral migration interacting with the number of inventors moving from origin city j in the current period. More formally, the instrument $\tilde{\psi}^R_{ij,t}$ for $\psi^R_{ij,t}$ is defined as follows:

$$\psi^R_{ij,t} = \frac{\tilde{L}^R_{ij,t}}{\sum_{k \in N} \tilde{L}^R_{kj,t}}$$

where

$$\tilde{L}^R_{ij,t} = \psi^R_{ij,t0} L^R_{i,t}.$$

For this specification, I use the period 1976-85 as t_0 and 1996-2015 to estimate (4.4) with origin fixed effects δ_j instead of origin-time fixed effects $\delta_{j,t}$. The identification assumption is that the historical migration shares contain information about a persistent component of migration costs, and this is not related to unobserved characteristics that affect the current technological relationship between cities.

4.3.2 Agglomeration Force

Next, I assume that the city-level shifters for the quality of own ideas, $\zeta_{i,t}$, is a function of the density in city i. This assumption is motivated by the empirical literature on the relationship between the size or density of a city (or a technology cluster) and its innovative activity[12]. For example, Carlino et al. (2007) documented that the number of patents per head increases with a city's population density. Moretti (2021) showed that an inventor's productivity increases when she moves to a denser technology cluster. Also, I assume that macroeconomic time trends in research productivity could also affect productivity changes, as shown in Bloom et al. (2020). More formally, I assume that

$$\zeta_{i,t} = \zeta_t \left(\frac{L_{i,t}}{H_i} \right)^{\rho} \varepsilon_{i,t}^{\zeta} \tag{4.5}$$

where ζ_t is a macroeconomic trend that affects the quality of ideas in all cities in time t, the parameter ρ captures the dynamic agglomeration externality in denser cities, and $\varepsilon_{i,t}^{\zeta}$ is an error term.

4.3.3 Estimation Procedure

I estimate β from (4.1) as follows. Given the estimates of $\{K_{ij,t}\}$ and γ in section 4.3.1, I obtain the values of $\{\zeta_{i,t}\}$ that satisfy (4.1) for $\beta \in [0,1]$:

$$\zeta_{i,t}\left(\beta\right) = \frac{\Gamma\left(1-\beta\right) \sum_{j \in N} \left(K_{ij,t} A_{j,t}\right)^{\beta}}{\left(A_{i,t+1} - A_{i,t}\right)} \tag{4.6}$$

Note that all values on the right-hand side of the equation are estimated in previous steps. Taking logs of both sides of (4.5), and rewriting ζ_t, ρ and $\varepsilon_{i,t}^{\zeta}$ as a function of β, I obtain

$$\ln \zeta_{i,t}\left(\beta\right) = \zeta_t\left(\beta\right) + \rho\left(\beta\right) \ln \left(\frac{L_{i,t}}{H_i} \right) + \varepsilon_{i,t}^{\zeta}\left(\beta\right) \tag{4.7}$$

Estimating (4.7) by OLS could be problematic because more inventive cities would

[12] Carlino and Kerr (2015) summarizes the recent literature on agglomeration and innovation.

attract more inventors, who would be anticipating agglomeration gains. Therefore, I again use city i's changes in the log of violent crime rates as an instrument for $\ln L_{i,t}$. I estimate (4.7) using data on CZs with at least one patent and citation.

I then estimate the parameters (β, ρ) jointly. For a given β, I estimate (4.7) and obtain the coefficient estimate $\rho(\beta)$ and residuals $\varepsilon^{\zeta}_{i,t}(\beta)$. I then choose the value of β that minimizes the squared sum of the residuals in (4.7) over cities and time periods, that is,

$$\beta^* = \arg\min_{\beta\in[0,1]} \sum_t \sum_{i\in N} \left[\varepsilon^{\zeta}_{i,t}(\beta)\right]^2 \tag{4.8}$$

After estimating β from (4.8), I obtain the corresponding value $\rho(\beta^*)$ as an estimate of ρ. Note that I include the case when $\beta = 0$, that is, knowledge diffusion has no effect on the evolution of the stock of knowledge and $\beta = 1$, which violates the assumption $\beta < 1$. This allows me to test whether changes in the stock of knowledge imply that the model validates the assumption of $\beta \in [0, 1)$.

4.4 Estimation Results

In this section, I first present results for the estimation of the parameters in the model. I begin with the estimation of the effect of inventor migration on knowledge flows, γ, and I describe the results for the structural estimation of parameters (β, ρ). The parameter choice and estimates are summarized in Table 3. Equipped with the parameter estimates, I decompose the cross-sectional variation in changes in the stock of knowledge into components explained by the following: (1) agglomeration forces and (2) knowledge diffusion.

4.4.1 Parameter Estimation

Crime rate changes at an origin have a negative effect on inventor emigrants from that location, and migration flows predicted by the historical migration shares have a positive effect, as shown in Table 1. The first column shows the results for the first-stage estimates of (4.4) using changes in log crime rates as an instrument, and the second column shows the results when I use both log crime rate changes and the log of predicted migration shares.

Patent citation shares follow inventor migration shares. Table 2 shows the estimation results using OLS, PPML, and IV. For the IV specifications, I report the first-stage F statistics as well as Hansen's J test of the overidentifying restrictions. These specifications pass the Hansen's J test; that is, I cannot reject the null hypothesis that the instruments are not correlated with the error terms. I choose $\gamma = 0.54$ in column (2) as my preferred estimate.

The structural estimation of β by (4.8) yields the estimate of $\beta = 0.39$. Given the estimate for β, I obtain the IV estimate of $\rho = 0.67$[13].

4.4.2 Decomposition of the Evolution of the Stock of Knowledge

To understand the quantitative importance of each component in (4.1), I decompose the cross-sectional variation in local productivity, following a commonly used procedure used in the international trade literature (e.g., Eaton et al. (2004); Hottman et al. (2016)). I use an operator Δ^c to denote the difference between a variable and its geometric mean over cities within a time period, such that $\Delta^c \ln x_{i,t} = \left[\ln x_{i,t} - \frac{1}{|N|}\sum_{k\in N} \ln x_{k,t}\right]$. Taking logs of (4.1) and expressing them relative to the geometric mean within time t, I obtain the following:

$$\Delta^c \ln (A_{i,t+1} - A_{i,t}) = \Delta^c \ln \zeta_{i,t} + \Delta^c \ln \left[\sum_{j\in N} (K_{ij,t}A_{j,t})^\beta\right] \tag{4.9}$$

Then, I regress each of the components of (4.9) on log productivity, using OLS as follows:

$$\Delta^c \ln \zeta_{i,t} = \delta^\zeta \Delta^c \ln (A_{i,t+1} - A_{i,t}) + \varepsilon^\zeta_{i,t} \tag{4.10}$$

$$\Delta^c \ln \left[\sum_{j\in N} (K_{ij,t}A_{j,t})^\beta\right] = \delta^K \Delta^c \ln (A_{i,t+1} - A_{i,t}) + \varepsilon^K_{i,t} \tag{4.11}$$

[13] In the first stage, regressing the log density of cities on changes in the log violent crime rates yields the coefficient (standard error clustered at CZ level) of -0.43 (0.07). The first-stage F statistic for the IV specification (4.7) is 54.

The OLS specification allocates the covariance terms among the components of (4.9) equally across those components, and it implies that $\delta^{\zeta}+\delta^{K}=1$. The coefficient estimates for the four equations above measure how much each component can explain variation in local productivity growth. I

Table 4 summarizes the results for the pooled regressions of equations (4.10)-(4.11) for all periods and each of the 10-year intervals separately. The variance decomposition indicates that 57% of the overall cross-sectional variation in the evolution of the stock of knowledge would be attributable to the city-specific shifter or agglomeration effect $\left(\delta^{\zeta}\right)$. The knowledge diffusion effect $\left(\delta^{K}\right)$ explains 43% of the variation.

Denser cities gained more from both agglomeration forces and knowledge diffusion. To visualize this, I plot the logarithm of each component on the right-hand side of (4.1) by city density in (4.9), given the estimated parameters. For each variable, I take the time-series average within each city. Blue (green) dots show a positive relationship between the average agglomeration forces (knowledge diffusion effect) and the average deviation of city density from the cross-sectional mean.

5 Place-Based Policies

In this section, I examine several counterfactual analyses, equipped with the estimates of parameters in the model. Let $x'_{i,t}$ and $\hat{x}_{i,t}=x'_{i,t}/x_{i,t}$ denote the value of variable $x_{i,t}$ in the counterfactual equilibrium and the counterfactual value relative to the value in the original equilibrium, respectively. In the counterfactual analyses, which require simulation of the counterfactual labor flows, I use data on migration flows $\{\lambda_{ij,t}\}$ and population from the IRS during 1990-2010, and I choose $t0: 1996-2005$ as an initial period and $t1: 2006-2015$ as the next period for the counterfactual analysis[14]. Due to a lack of data on inventor wages, I must assume a counterfactual change in inventor migration flows. I assume that a counterfactual change in inventor migration shares is equal to a counterfactual change in migration shares, $\hat{L}^{R}_{ij,t}=\hat{L}_{ij,t}$ when simulating

[14]Since the consistent measure of migration flows in the IRS data is only available for the period 1990-2010, I use the average IRS migration shares during 2006-2010 to approximate the model's migration shares for the period 2006-2015.

counterfactual changes in the bilateral learning parameters $\left\{\hat{K}_{ij,t}\right\}$[15].

I use employment shares in the initial period to compute aggregate productivity and welfare measures. The counterfactual change in aggregate productivity in time $t \in \{t0, t1\}$ is defined as follows: $\hat{A}_t = \sum_{i \in N} \left(L_{i,t0} / \sum L_{i,t0}\right) \hat{A}_{i,t}$. To understand the potential equity-efficiency tradeoffs, I measure spatial disparities by the coefficient of variation, defined as the cross-sectional variance divided mean of the stock of knowledge. The counterfactual change in aggregate welfare is defined as the share-weighted local welfare[16].

5.1 Regional Income Subsidies

I study the long-run effect of place-based policies that subsidize workers residing in a set of cities. A common objective for these kinds of policies is to reduce spatial inequality across cities while strengthening agglomeration externalities in target areas. A recent example of such a policy in the U.S. is the Regional Innovation program in the Creating Helpful Incentives to Produce Semiconductors and Science Act of 2022, which aims to create twenty regional technology hubs in areas that are not leading centers for innovation. I implement a one-period labor income subsidy that targets a set of twenty cities, which amounts to one percent of the GDP at the beginning of the initial period t_0 [17]. The effect of the policy is measured by the counterfactual changes in the economy in t_1 when the stock of knowledge is affected by the relocation of workers in t_0 through agglomeration forces and migration. I describe the details of the counterfactual policy implementation in Appendix C.1.

I consider three groups of cities as target locations for the place-based subsidy. The

[15] This assumption would not be innocuous. The underlying assumption is that changes in amenities and real income for inventors are proportional to changes in these values for general people. Extending this model by including two types of skills of workers is suggested for future work.

[16] This utilitarian approach to aggregate welfare was also used in studies such as Caliendo et al. (2019). Under free mobility, that is, $\mu_{ij,t1} = 1, \forall i, j \in N$, a change in the common welfare can be written as a function of the share-weighted average of real income changes across cities, where the weights depend on employment shares, as in Redding (2016).

[17] The proposed funding for the Regional Innovation program is ten billion dollars to create twenty regional technology hubs and one billion dollars for subsidizing ten economically distressed areas over four years during 2023-2027, with total costs of approximately 0.04% of GDP as of 2021. However, as shown below, the model implies small effects of a regional subsidy under a more intense policy program.

first policy targets the least productive cities with at least one inventor. This idea of targeting less productive cities resembles the idea behind the Empowerment Zone program in the U.S. and the ZFU program in France. Gaubert (2018) studied the general equilibrium effect of place-based tax incentives that subsidize firms in the smallest cities and showed the negative aggregate impacts on productivity and spatial equity due to a loss of agglomeration forces in midsize cities.

The second set of cities includes those that are the most productive. Targeting these cities would be based on recognition of agglomeration forces in innovative activities and because policymakers are primarily interested in aggregate efficiency. A recent empirical paper by Moretti (2021) showed that inventors become more productive when they relocate to a denser technological cluster, and a quantitative work by Sollaci (2022) argued that productive cities are heavily subsidized under the optimal R&D policy.

The third group of cities comes from a list of potential technology hubs proposed by Gruber and Johnson (2019) (hereafter, GJ) who proposed creating new technological hubs outside existing superstar cities in the U.S. They created rankings of cities based on education, working-age population, and housing prices[18]. During 2006-2015, this group's median employment and the stock of knowledge were at the 87th and 88th percentile of the distribution of U.S. cities, respectively. Thus, the cities may not be the most productive U.S. cities, but they are certainly already highly productive.

Results

The model predicts that place-based income subsidies will have small aggregate effects. One reason for this is that the observed migration flows imply prohibitive migration costs for most city pairs. When there is no observed migration flow from city i to j, (2.10) it suggests that the migration cost from i to j is infinitely high. Thus, in the counterfactual equilibrium, a worker in city i can move to city j only if the data revealed a positive

[18]They proposed a list of 102 potential cities to serve as technology hubs. Their candidate cities consist of single MSAs or multiple MSAs. I obtain their Technology Hub Index System (THIS) list from this website: `https://www.jump-startingamerica.com/102-places-for-jumpstarting-america`. From their list, I choose the top nineteen CZs that have at least one county in common with the counties in their top 11 individual and combined MSAs, and then I select the most populous CZ in their top 12th location, with 20 target CZs in total.

migration flow, that is, $\mathbb{B}_{ij,t} > 0$. Another reason is that for bigger cities, the implied subsidy rate is not high enough to attract many workers[19]. Table 5 shows the local and aggregate effects of the policy for each group of cities. The first three rows summarize local changes in employment, the stock of knowledge, and welfare in target cities in 2006-2015 relative to the original equilibrium. The last three rows show the aggregate effect on the stock of knowledge, real income, and the coefficient of variation of the stock of knowledge. Overall, a one-period regional subsidy does not have a long-run impact on the local or aggregate economy. The first column shows that local employment increases in the least productive cities by 3.4 percent, while an increase in the stock of knowledge is only 1.5 percent. The local welfare slightly declines because the local productivity gain does not offset a decline in the local real income. In the initial period, more workers are willing to relocate to target cities under labor subsidies. However, in the next period, the subsidy is no longer implemented, and the real income remains at a lower level relative to the original equilibrium, as it is costly to emigrate from the target cities. Aggregate effects are almost null, and this is true for other target areas as well.

Taken together, these results imply that a place-based labor subsidy may not be effective in improving aggregate equity or efficiency.

5.2 Reduction of Migration Costs

While a labor subsidy has a weak effect on the model, reducing migration costs could have local and aggregate impacts through changes in agglomeration forces and migration flows. To investigate this possibility, I examine a counterfactual policy that reduces migration costs toward target cities in the initial period t_0. In the next period t_1, migration costs have the same values as in the original equilibrium. I evaluate the local and aggregate impacts in the next period t_1 as the dynamic effects of the policy.

The first policy considers a finite reduction of migration costs toward target cities to zero from all other cities. For a target city j, this implies that $\mu'_{ij,t0} = 1, \forall i \in N, \mathbb{B}_{ij,t0} > 0$.

[19] The labor income subsidy rate that corresponds to the total cost of 1% of GDP in the model is 2 percent for the most productive cities and 11 percent for the GJ cities, whereas it is 1,230 percent for the least productive cities.

This policy does not allow workers to move from city k to j if there was no observed migration flow from k to j, or $\mathbb{B}_{ik} = 0$. In this counterfactual scenario, I assume it is extremely costly for highly skilled people to relocate in the absence of migration history from their location to the target cities. This could be because city j lacks industries or institutions that can utilize the kinds of knowledge available in city k. The second policy relaxes this assumption and considers free mobility toward target cities from all other cities. For a target city j, this implies that $\mu'_{ij,t0} = 1, \forall i \in N$.

Reducing the migration costs toward target cities has two separate effects on the evolution of the stock of knowledge. First, an increase in local employment improves agglomeration forces in target cities. Second, it changes the composition of migration inflows from other cities. Under free mobility toward target cities, the counterfactual value of the ease of migration becomes $\mathbb{B}'_{ij,t0} = B_{j,t}, \forall i \in N$ for target city j. Taking this into (2.10), the probability that a worker moves to city j becomes identical across all cities, or

$$\lambda'_{ij,t0} = \frac{\left(v'_{j,t0}/P'^{\alpha}_{j,t0} r'^{1-\alpha}_{j,t0}\right)^{\epsilon}}{\sum_{k \in N}\left(v'_{k,t0}/P'^{\alpha}_{k,t0} r'^{1-\alpha}_{k,t0}\right)^{\epsilon}} = \bar{\lambda}_{j,t0}, \forall i \in N.$$

As a result, the in-migration share from city i in the target city j becomes equal to the employment share of i. Thus, workers in target cities tend to interact more with migrants from larger cities. In the case of a finite cost reduction, in-migration share is equalized across origin cities with positive migration flows to target cities.

A crucial difference between the finite reduction of migration costs and the free mobility cases is that the latter affects the extensive margin of migration flows and hence, the diffusion of knowledge. Under free mobility to target cities, workers can access technology in other cities that was not available to them in the original equilibrium because of a lack of migration flows. On the other hand, a finite cost reduction policy only changes the intensive-margin composition of migration and knowledge flows within city pairs with positive migration flows in the original equilibrium.

Results

A finite reduction of migration costs toward less productive cities generates effects of similar magnitude to the work subsidy program considered in the preceding subsection. The first columns Table 6 show the effect of targeting the least productive cities. The effect on local employment is larger than that of the labor subsidy examined in the previous section. However, it does not greatly affect the spatial distribution of economic activity, and the aggregate effects are small. The parentheses under the changes in local employment and stock of knowledge show changes in the rank of the median value of target cities. The target cities' median employment (stock of knowledge) was at the 6th (4th) percentile in the distribution across cities in the original equilibrium, and it stays the same in the counterfactual equilibrium.

On the other hand, a reduction of migration costs toward productive cities has large aggregate impacts. Migration flows in the actual data imply that productive cities already generate strong migration forces, and these forces amplify positive shocks to the strength of agglomeration. The second column Table 6 shows that the reduction of migration costs toward the most productive cities leads to a 4.8 percent increase in aggregate welfare. Meanwhile, it leads to a 32.2 percent increase in the coefficient of variation of the stock of knowledge. The model predicts that improving mobility toward the most productive cities leads to higher growth and spatial disparities. This indicates an equity-efficiency tradeoff in innovation policy that favors the leading technological hubs. The third column shows that the policy can improve both aggregate equity and efficiency by targeting the productive but not the most productive cities. Targeting the selected cities from GJ's proposal improves aggregate welfare by 2.1 percent and reduces the spatial variation of productivity by 5.9 percent. The counterfactual policy implies that these cities can be new leading hubs: the median employment (the stock of knowledge) in these cities increases from the 87th (88th) to the 95th (95th) percentile of the spatial distribution.

My second counterfactual analysis for this policy regime indicates the importance of the extensive margin of migration flows in improving the productivity of *ex ante* less productive cities. Table 7 shows that employment levels and the stock of knowledge in the least productive cities can be dramatically improved when workers can freely move to these cities. This is because less productive cities tend to have sparse migration inflows from other cities, and opening the local labor market to the rest of the economy facilitates knowledge inflows. However, employment in these locations remains lower than in more productive cities, and the aggregate impacts are relatively small. The aggregate productivity and welfare effects of targeting bigger cities are smaller than in the case of the finite cost reduction policy. Under free mobility, workers also start learning from the least productive cities, and this creates a congestion effect that reduces productivity gains through knowledge diffusion.

To summarize, my model highlights that place-based reduction of migration costs can generate a Pareto improvement via targeting initially productive cities. The model points to the importance of reducing frictions in the extensive margin of migration inflows as a mechanism for stimulating new ideas and fostering growth in less productive cities.

6 Conclusion

This paper studies the effect of knowledge diffusion through the internal migration of inventors across U.S. cities on local and aggregate productivity growth and welfare. I provide evidence regarding the effect of inventor migration on knowledge flows to a destination, which informs my model of human capital accumulation through the interaction of workers across cities. I then develop a quantitative spatial equilibrium theory to explain the distributional consequences of intercity interactions, including trade flows, migration flows, and knowledge flows. Diminishing returns in interacting with others at different locations induce gains that arise from learning from more distinct locations. I estimate the model using data on U.S. cities and show that knowledge diffusion explains 40 percent of the cross-sectional variation in productivity changes; the remaining variation is attributable to agglomeration forces. In the counterfactual analysis, I show that a place-based reduction of migration costs targeting relatively *ex-ante* productive areas can induce a Pareto improvement and resolve the equity-efficiency tradeoff endemic to alternative place-based policy instruments. The model highlights the importance of improving the extensive margin of migration inflows for fostering productivity growth in less productive cities.

References

Treb Allen and Costas Arkolakis. Trade and the Topography of the Spatial Economy. *The Quarterly Journal of Economics*, 129(3), 2014.

Costas Arkolakis, Sun Kyoung Lee, and Michael Peters. European Immigrants and the United States' Rise to the Technological Frontier. *mimeo*, 2020.

Enrico Berkes, Ruben Gaetani, and Martı Mestieri. Cities and Technology Cycles. *mimeo*, 2022.

Nicholas Bloom, Charles I. Jones, John Van Reenen, and Michael Webb. Are Ideas Getting Harder to Find? *American Economic Review*, 110(4), 2020.

Christian Broda and David E. Weinstein. Globalization and the Gains From Varietystandard errors are in parentheses. *The Quarterly Journal of Economics*, 121(2), 2006.

Gharad Bryan and Melanie Morten. The Aggregate Productivity Effects of Internal Migration: Evidence from Indonesia. *Journal of Political Economy*, 127(5), 2019.

Francisco J. Buera and Ezra Oberfield. The Global Diffusion of Ideas. *Econometrica*, 88 (1), 2020.

Sheng Cai, Lorenzo Caliendo, Fernando Parro, and Wei Xiang. Mechanics of Spatial Growth. *mimeo*, 2022.

Lorenzo Caliendo, Maximiliano Dvorkin, and Fernando Parro. Trade and Labor Market Dynamics: General Equilibrium Analysis of the China Trade Shock. *Econometrica*, 87 (3), 2019.

Gerald Carlino and William R. Kerr. Chapter 6 - Agglomeration and Innovation. In Gilles Duranton, J. Vernon Henderson, and William C. Strange, editors, *Handbook of Regional and Urban Economics*, volume 5 of *Handbook of Regional and Urban Economics*. Elsevier, 2015.

Gerald A. Carlino, Satyajit Chatterjee, and Robert M. Hunt. Urban density and the rate of invention. *Journal of Urban Economics*, 61(3), 2007.

Donald R. Davis and Jonathan I. Dingel. A Spatial Knowledge Economy. *American Economic Review*, 109(1), 2019.

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

Klaus Desmet and Esteban Rossi-Hansberg. Spatial Development. *American Economic Review*, 104(4), 2014.

Klaus Desmet, Dávid Krisztián Nagy, and Esteban Rossi-Hansberg. The Geography of Development. *Journal of Political Economy*, 126(3), 2018.

Gilles Duranton and Diego Puga. Nursery Cities: Urban Diversity, Process Innovation, and the Life Cycle of Products. *American Economic Review*, 91(5), 2001.

Gilles Duranton and Diego Puga. Chapter 48 - Micro-Foundations of Urban Agglomeration Economies. In J. Vernon Henderson and Jacques-François Thisse, editors, *Handbook of Regional and Urban Economics*, volume 4 of *Cities and Geography*. Elsevier, 2004.

Jonathan Eaton and Samuel Kortum. International Technology Diffusion: Theory and Measurement. *International Economic Review*, 40(3), 1999.

Jonathan Eaton and Samuel Kortum. Technology, Geography, and Trade. *Econometrica*, 70(5), 2002.

Jonathan Eaton, Samuel Kortum, and Francis Kramarz. Dissecting Trade: Firms, Industries, and Export Destinations. *American Economic Review*, 94(2), 2004.

Pablo D. Fajgelbaum and Cecile Gaubert. Optimal Spatial Policies, Geography, and Sorting. *The Quarterly Journal of Economics*, 135(2), 2020.

Maryann P Feldman and David B Audretsch. Innovation in cities: Science-based diversity, specialization and localized competition. *European Economic Review*, 1999.

Cecile Gaubert. Firm Sorting and Agglomeration. *American Economic Review*, 108(11), 2018.

Edward L Glaeser. Learning in Cities. *Journal of Urban Economics*, 46(2), 1999.

Edward L. Glaeser and Joshua D. Gottlieb. The Wealth of Cities: Agglomeration Economies and Spatial Equilibrium in the United States. *Journal of Economic Literature*, 47(4), 2009.

Jonathan Gruber and Simon Johnson. *Jump-Starting America: How Breakthrough Science Can Revive Economic Growth and the American Dream*. PublicAffairs, 2019. ISBN 978-1-5417-6250-3.

Colin J. Hottman, Stephen J. Redding, and David E. Weinstein. Quantifying the Sources of Firm Heterogeneity. *The Quarterly Journal of Economics*, 131(3), 2016.

A. B. Jaffe, M. Trajtenberg, and R. Henderson. Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations. *The Quarterly Journal of Economics*, 108 (3), 1993.

Charles I. Jones. R & D-Based Models of Economic Growth. *Journal of Political Economy*, 103(4), 1995.

Charles I. Jones. Chapter 16 - Growth and Ideas. In Philippe Aghion and Steven N. Durlauf, editors, *Handbook of Economic Growth*, volume 1. Elsevier, 2005.

Jacob Kaplan. Jacob Kaplan's Concatenated Files: Uniform Crime Reporting Program Data: Offenses Known and Clearances by Arrest (Return A), 1960-2020, 2021.

William R. Kerr. Breakthrough inventions and migrating clusters of innovation. *Journal of Urban Economics*, 67(1), 2010.

William R. Kerr and Scott Duke Kominers. Agglomerative Forces and Cluster Shapes. *The Review of Economics and Statistics*, 97(4), 2015.

Patrick Kline and Enrico Moretti. Local Economic Development, Agglomeration Economies, and the Big Push: 100 Years of Evidence from the Tennessee Valley Authority standard errors are in parentheses. *The Quarterly Journal of Economics*, 129 (1), 2014.

Samuel S. Kortum. Research, Patenting, and Technological Change. *Econometrica*, 65 (6), 1997.

Nelson Lind and Natalia Ramondo. Trade with Correlation. Technical Report w24380, National Bureau of Economic Research, Cambridge, MA, 2018.

Ernest Liu and Song Ma. Innovation Networks and Innovation Policy. Technical Report 29607, National Bureau of Economic Research, Inc, 2021.

Ferdinando Monte, Stephen J. Redding, and Esteban Rossi-Hansberg. Commuting, Migration, and Local Employment Elasticities. *American Economic Review*, 108(12), 2018.

Enrico Moretti. The Effect of High-Tech Clusters on the Productivity of Top Inventors. *American Economic Review*, 111(10), 2021.

Petra Moser. Do Patents Weaken the Localization of Innovations? Evidence from World's Fairs. *The Journal of Economic History*, 71(2), 2011.

Dávid Krisztián Nagy. Quantitative economic geography meets history: Questions, answers and challenges. *Regional Science and Urban Economics*, 2021.

Jesse Perla, Christopher Tonetti, and Michael E. Waugh. Equilibrium Technology Diffusion, Trade, and Growth. *American Economic Review*, 111(1), 2021.

Marta Prato. The Global Race for Talent: Brain Drain, Knowledge Transfer and Growth. *mimeo*, 2021.

Stephen J. Redding. Goods trade, factor mobility and welfare. *Journal of International Economics*, 101, 2016.

Jorge De La Roca and Diego Puga. Learning by Working in Big Cities. *The Review of Economic Studies*, 84(1), 2017.

Gregor Schubert. House Price Contagion and U.S. City Migration Networks. *mimeo*, 2021.

Ina Simonovska and Michael E. Waugh. The elasticity of trade: Estimates and evidence. *Journal of International Economics*, 92(1), 2014.

Alexandre Sollaci. Agglomeration, Innovation, and Spatial Reallocation: The Aggregate Effects of R&D Tax Credits. *SSRN Electronic Journal*, 2022.

Peter Thompson and Melanie Fox-Kean. Patent Citations and the Geography of Knowledge Spillovers: A Reassessment. *The American Economic Review*, 95(1), 2005.

7 Tables

Table 1: First Stage Estimates

	$\ln(\psi^R_{ji,t-1})$	
	(1)	(2)
$\Delta \ln(Crime_{j,t-1})$	-0.159***	-0.081**
	(0.019)	(0.033)
$\ln(\tilde{\psi}^R_{ji,t-1})$		1.037***
		(0.012)
Observations	41,987	10,360
Adjusted R^2	0.61	0.87
Destination-Time FE	✓	✓
Origin FE	✓	✓

Notes: This table shows the first-stage regression results for the IV specification in (4.4). Standard errors in parentheses and clustered by destination CZ. ***p $<$ 0.01, **p $<$ 0.05, *p $<$ 0.1.

Table 2: Effects of Migration on Citation Shares

	$\ln(\phi_{ij,t})$					
	(1)	(2)	(3)	(4)	(5)	(6)
$\ln(\psi^R_{ji,t-1})$	0.542***	0.433***	0.540***	0.444***	0.427***	0.457***
	(0.010)	(0.009)	(0.160)	(0.009)	(0.008)	(0.009)
Observations	41,987	47,832	41,987	10,360	39,487	10,360
R^2	0.82	0.18	0.33	0.85	0.18	0.50
Destination-Time FE	✓	✓	✓	✓	✓	✓
Origin-Time FE	✓	✓		✓	✓	
Origin FE			✓			✓
Specification	OLS	PPML	IV	OLS	PPML	IV
First-Stage F			69			3854
Hansen's J Test (p-value)						0.31

Notes: This table shows the results for (4.4). First-Stage F shows the Kleibergen-Paap rk Wald F statistic for each of the IV specifications. Standard errors are in parentheses and clustered by destination CZ. ***p $<$ 0.01, **p $<$ 0.05, *p $<$ 0.1.

Table 3: Parameter Values

Parameter	Value	Description	Source
α	0.75	Consumption shares	Davis and Ortalo-Magné (2011)
β	0.39	Learning efficiency	(4.8)
γ	0.54	Knowledge flow elasticity wrt migration	(4.4)
ρ	0.66	Original idea elasticity wrt city size	(4.8)
σ	4	Elasticity of substitution	Broda and Weinstein (2006)
θ	4	Productivity shape parameter	Simonovska and Waugh (2014)
ω	0.32	Trade elasticity	Monte et al. (2018)
ϵ	3	Amenity shape parameter	Bryan and Morten (2019)

Notes: This table reports the estimated and calibrated parameters in the model.

Table 4: Variance Decomposition of the Evolution of the Stock of Knowledge

	Agglomeration (δ^{ζ})	Knowledge Diffusion (δ^{K})
All Periods	0.567	0.433
	(0.009)	(0.009)
$t1: 1986-1995$	0.496	0.504
	(0.017)	(0.017)
$t1: 1996-2005$	0.550	0.450
	(0.013)	(0.013)
$t1: 2006-2015$	0.644	0.356
	(0.015)	(0.015)

Notes: This table summarizes the estimation results for the variance decomposition in equations (4.10)-(4.11). The first row shows the results using the whole sample period from 1976 to 2015. The next three rows show the results for each of the separate time periods at a 10-year frequency. Standard errors are in parentheses.

Table 5: Dynamic Effects of a Regional Subsidy

	Bottom 20	Top 20	GJ 20
Local Employment	3.417%	0.016%	0.089%
Local Stock of Knowledge	1.470%	0.007%	0.037%
Local Welfare	−0.713%	−0.003%	−0.016%
Aggregate Stock of Knowledge	0.000%	0.000%	0.000%
Aggregate Welfare	0.000%	0.000%	0.000%
Coef. Var. Stock of Knowledge	−0.001%	0.005%	−0.002%

Notes: This table shows the results of the counterfactual analyses in Section 5.1. All values are expressed as changes from the original equilibrium. Each column shows the effect of a place-based labor income subsidy targeting a set of cities. The target cities in this analysis are as follows: (1) the least productive twenty CZs (Bottom 20), (2) the most productive twenty CZs (Top 20), (3) the top twenty CZs in the list of potential technology hubs proposed by Gruber and Johnson (2019) (GJ 20).

Table 6: Dynamic Effects of the Reduction of Migration Costs To Target Cities

	Bottom 20	Top 20	GJ 20
Local Employment	17.57%	4.19%	16.23%
	(6 → 6th pct.)	(99 → 99th pct.)	(87 → 95th pct.)
Local Stock of Knowledge	25.91%	101.96%	181.84%
	(4 → 4th pct.)	(99 → 99th pct.)	(88 → 95th pct.)
Local Welfare	−0.44%	12.44%	16.26%
Aggregate Stock of Knowledge	0.015%	36.10%	26.42%
Aggregate Welfare	0.002%	4.83%	2.11%
Coef. Var. Stock of Knowledge	−0.006%	32.20%	−5.94%

Notes: This table shows the results for counterfactual analyses in Section 5.1. All values are expressed as changes from the original equilibrium in $t1 : 2006 - 2015$. Each column shows the effect of a place-based labor income subsidy targeting a set of cities. The target cities in this analysis are as follows: (1) the least productive twenty CZs (Bottom 20), (2) the most productive twenty CZs (Top 20), (3) the top twenty CZs in the list of potential technology hubs proposed by Gruber and Johnson (2019) (GJ 20).

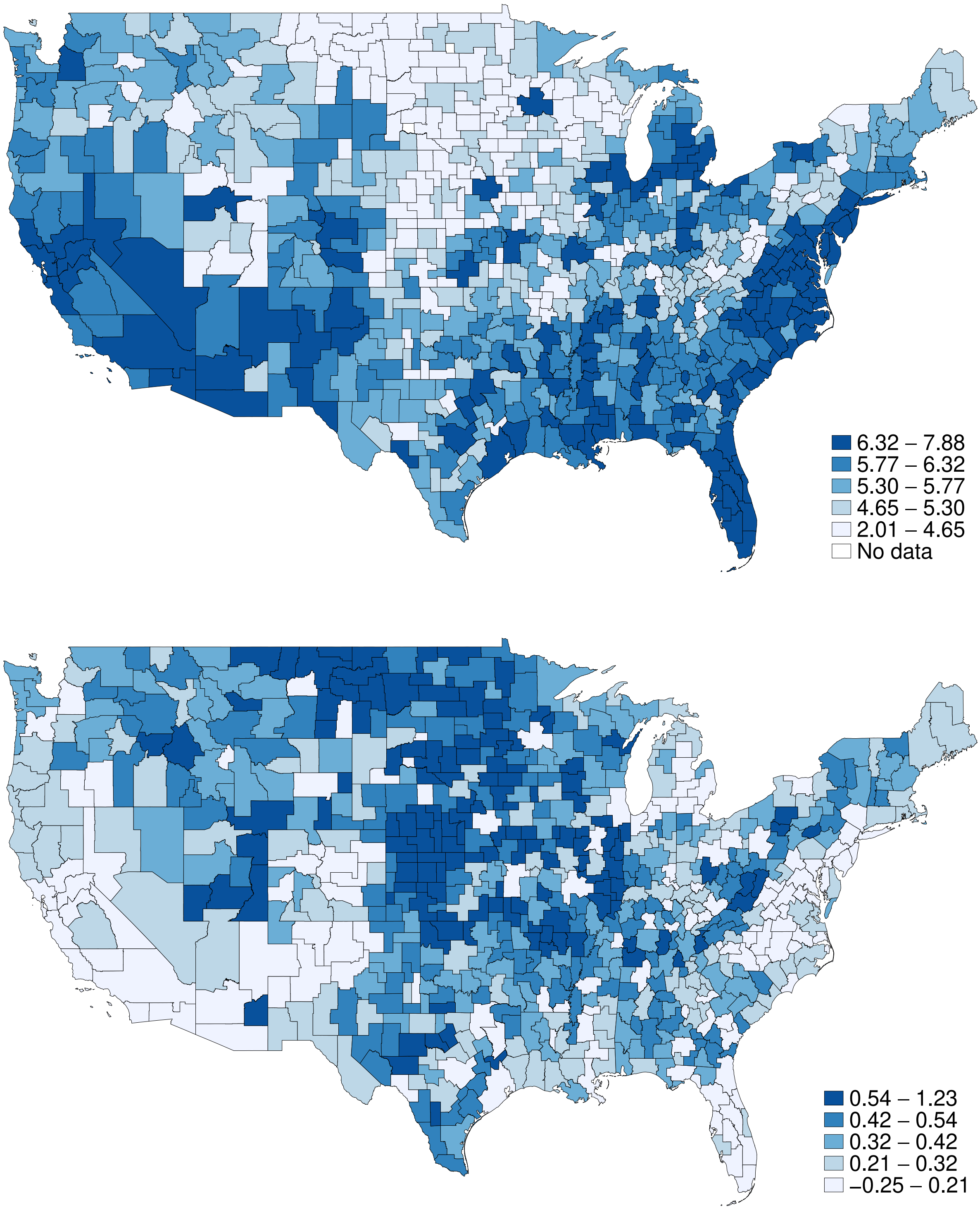
Table 7: Dynamic Effects of Free Mobility to Target Cities

	Bottom 20	Top 20	GJ 20
Local Employment	575.35%	5.40%	26.70%
	(6 $\rightarrow$ 38th pct.)	(99 $\rightarrow$ 99th pct.)	(87 $\rightarrow$ 96th pct.)
Local Stock of Knowledge	2330.85%	96.02%	252.31%
	(4 $\rightarrow$ 50th pct.)	(99 $\rightarrow$ 99th pct.)	(88 $\rightarrow$ 96th pct.)
Local Welfare	5.10%	11.40%	18.09%
Aggregate Stock of Knowledge	1.66%	33.77%	17.61%
Aggregate Welfare	0.10%	4.53%	1.71%
Coef. Var. Stock of Knowledge	−1.30%	32.74%	−8.27%

Notes: This table shows the results of counterfactual analyses in Section 5.2. All values are expressed as changes from the original equilibrium in $t1 : 2006 - 2015$. Each column shows the effect of a place-based labor income subsidy targeting a set of cities. The target cities in this analysis are as follows: (1) the least productive twenty CZs (Bottom 20), (2) the most productive twenty CZs (Top 20), (3) the top twenty CZs in the list of potential technology hubs proposed by Gruber and Johnson (2019) (GJ 20).

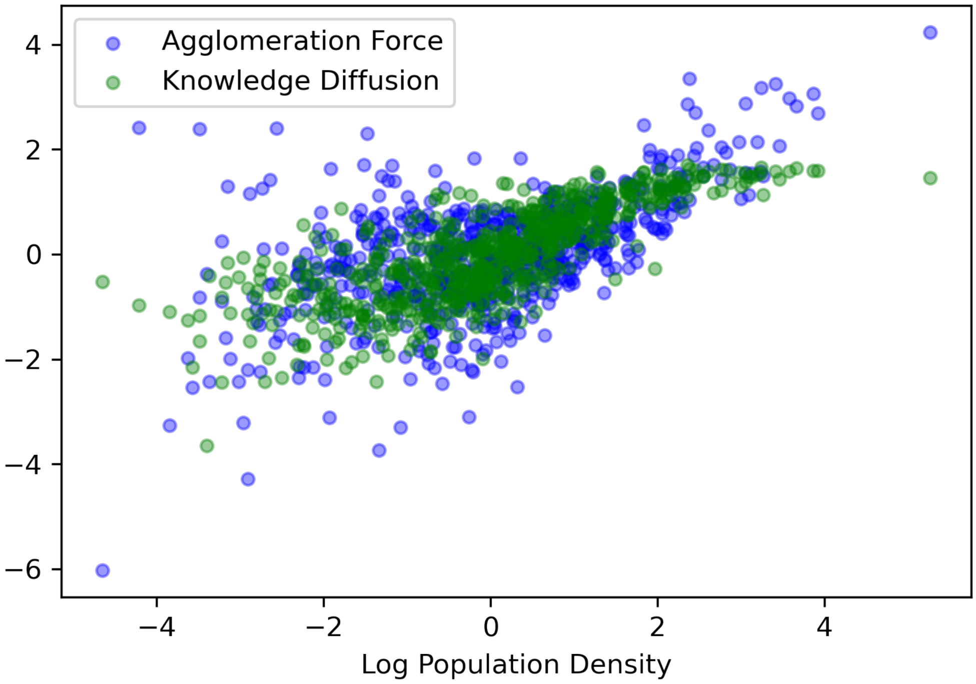
8 Figures

Figure 1: Average Log Crime Rates and Log Crime Rate Changes



Notes: The top panel shows the average log violent crime rates during 1966-1975. The bottom panel shows the average log violent crime rate changes from 1976 to 2015.

Figure 2: Average Changes in the Stock of Knowledge by Density



Notes: This figure shows the average value of each term over time in (4.9). Blue dots represent the agglomeration forces, and green dots show the migration force for each CZ.

Appendix

A Model

A.1 Proof of Equation (2.4)

Conditional on adopting, the probability that a producer in location i builds upon an idea from location j is as follows:

$$\phi_{ij,t} = \frac{\kappa_{ij,t}^{\theta/\beta} A_{j,t}}{\sum_{k \in N} \kappa_{ik,t}^{\theta/\beta} A_{k,t}}$$

Proof. I suppress the time subscript t for notational clarity. □

$$\begin{aligned}
P\left[\max_k \{q_{ik}(v); k \neq j\} \leq q_{ij}(v)\right] = P\left[\max_k \left\{\epsilon_i \kappa_{ik} z_k^{\beta}(v); k \neq j\right\} \leq \epsilon_i \kappa_{ij} z_j^{\beta}(v)\right] \\
=& P\left[\max_k \left\{\left(\frac{\kappa_{ik}}{\kappa_{ij}}\right)^{1/\beta} z_k(v); k \neq j\right\} \leq z_j(v)\right] \\
=& \int_0^{\infty} \Pi_{k \neq j} F_k\left(\left(\frac{\kappa_{ik}}{\kappa_{ij}}\right)^{1/\beta} z\right) dF_j(z) \\
=& \int_0^{\infty} \exp\left[-\kappa_{ij}^{-\theta/\beta} \sum_{k \neq j} A_k \kappa_{ik}^{\theta/\beta} z^{-\theta}\right] dF_j(z) \\
=& \int_0^{\infty} \exp\left[-\kappa_{ij}^{-\theta/\beta} \sum_{k \neq j} A_k \kappa_{ik}^{\theta/\beta} z^{-\theta}\right] \cdot \exp\left[-A_j z^{-\theta}\right] \cdot \left(\theta A_j z^{-\theta-1}\right) dz \\
=& A_j \int_0^{\infty} \exp\left[\left(-\kappa_{ij}^{-\theta/\beta} \sum_{k \in N} A_k \kappa_{ik}^{\theta/\beta}\right) z^{-\theta}\right] \cdot \left(\theta z^{-\theta-1}\right) dz \\
=& \frac{\kappa_{ij}^{\theta/\beta} A_j}{\sum_{k \in N} \kappa_{ik}^{\theta/\beta} A_k}
\end{aligned}$$

A.2 Proof of Equation (2.3)

The cumulative distribution evolves according to the following:

$$
\begin{aligned}
\frac{d}{dt}\ln F_{i,t}(Z) &= -\sum_{j\in N}\int_0^{\infty} G_t\left(\frac{Z}{\kappa_{ij,t}z^{\beta}}\right)dF_{j,t}(z)\\
&= -\zeta_{it}\sum_{j\in N}\int_0^{\infty}\left(\frac{Z}{\kappa_{ij,t}z^{\beta}}\right)^{-\theta}dF_{j,t}(z)\\
&= -\zeta_{it}Z^{-\theta}\sum_{j\in N}\kappa_{ij,t}^{\theta}\int_0^{\infty} z^{\beta\theta}dF_{j,t}(z)
\end{aligned}
$$

Under the distributional assumptions for $\zeta_{i,t}$ in (2.2), I get

$$
\frac{d}{dt}A_{i,t} = \zeta_{i,t}\Gamma(1-\beta)\sum_{j\in N}\kappa_{ij,t}^{\theta}A_{j,t}^{\beta}
$$

In discrete time, the scale parameter evolves according to the following:

$$
\begin{aligned}
A_{it+1} &= A_{i,t} + \zeta_{i,t}\sum_{j\in N}\kappa_{ij,t}^{\theta}\int_0^{\infty} z^{\beta\theta}dF_{j,t}(z)\\
&= A_{i,t} + \zeta_{i,t}\Gamma(1-\beta)\sum_{j\in N}\kappa_{ij,t}^{\theta}A_{j,t}^{\beta}
\end{aligned}
$$

A.3 Existence and Uniqueness

Proposition 2. *Given the land area, productivity and amenity parameters $\{H_i, A_{i,t}, B_{i,t}\}$, quasi-symmetric bilateral trade frictions $\{d_{ij}\}$ and migration costs $\{\mu_{ij,t}\}$ for all cities $i, j \in N$, there exist unique equilibrium populations $\{L_{i,t}\}$, wages $\{w_{i,t}\}$ and trade shares $\{\pi_{ij,t}\}$. Proof. The proof follows the same structure as in Allen and Arkolakis (2014), Redding (2016) and Monte et al. (2018).*

A.4 Model Inversion

Given the model parameters $\{\alpha, \sigma, \theta, \psi\}$, parameterized bilateral trade costs $\{d_{ij}\}$, and data on populations, wages, and land supplies $\{L_{i,t}, w_{i,t}, H_{i,t}\}$, there exist unique values of productivity parameters $\{A_{i,t}\}$ and amenity parameters $\{B_{i,t}\}$ that are consistent with the data up to scale for each time period t. It follows that there exist unique values of spatial meeting rates $\{\kappa_{ij,t}\}$ that are consistent with the data on $\{\phi_{ij,t}\}$ up to scale for each city i.

B Estimation

B.1 Estimation of Amenity Values

To estimate local amenity values $\{B_{i,t}\}$, I follow the steps described below.

1. Inverting the ease of migration $\mathbb{B}_{ij,t} = B_{j,t}\mu_{ij,t}^{-\epsilon}$ from the model.

2. Running the following fixed-effect regressions to recover the amenity value at destination j

$$\ln \mathbb{B}_{ij,t} = \delta_{j,t} + \varepsilon_{ij,t}$$

3. Estimating $\delta_{j,t}$ by PPML.

4. The free mobility counterfactual assumes $\mathbb{B}'_{ij,t} = \delta_{j,t}$.

C Counterfactual Analysis

C.1 Labor Income Subsidy

Consider local transfers that subsidize a particular set of cities, denoted by N_t^s, such that workers in city $i \in N_s$ can obtain a transfer of $s_i w_{i,t}$ per head in time t. The subsidy is paid for by a lump-sum tax levied on all workers in the economy. Let τ_t denote the amount of tax paid by workers per head, which is given by

$$\tau_t = \frac{\sum_{i \in N_t^s} s_{i,t} w_{i,t} L_{i,t}}{L_t}$$

Under the subsidy, the total income in city i in time t becomes

$$\begin{aligned} v_{i,t} L_{i,t} &= (1 + s_{i,t})\, w_{i,t} L_{i,t} + \frac{1-\alpha}{\alpha} \cdot \left\{ (1 + s_{i,t})\, w_{i,t} - \tau_t \right\} L_{i,t} - \tau_t L_{i,t} \\ &= \left\{ \frac{(1 + s_{i,t})\, w_{i,t} - \tau_t}{\alpha} \right\} L_{i,t} \end{aligned} \tag{C.1}$$

The rent is

$$r_{i,t} = \frac{1-\alpha}{\alpha} \cdot \frac{\left\{ (1 + s_{i,t})\, w_{i,t} - \tau_t \right\} L_{i,t}}{H_i} \tag{C.2}$$

Price index takes the same expression as (2.13):

$$P_i = \gamma \cdot w_i \left(\frac{A_i}{\pi_{ii}} \right)^{-\frac{1}{\theta}}, \quad \gamma \equiv \left[\Gamma \left(\frac{\theta - (\sigma - 1)}{\theta} \right) \right]^{\frac{1}{1-\sigma}}$$

From equations (C.1), (C.2), and (2.13), the real income can be written as follows:

$$\begin{aligned} v_{i,t} / P_{i,t}^{\alpha} r_{i,t}^{1-\alpha} &= \left\{ \frac{(1 + s_{i,t})\, w_{i,t} - \tau_t}{\alpha} \right\} \gamma^{-\alpha} \cdot w_{i,t}^{-\alpha} \left(\frac{A_{i,t}}{\pi_{ii,t}} \right)^{\frac{\alpha}{\theta}} \left(\frac{1-\alpha}{\alpha} \cdot \frac{\left\{ (1 + s_{i,t})\, w_{i,t} - \tau_t \right\} L_{i,t}}{H_i} \right)^{-(1-\alpha)} \\ &\propto \left\{ (1 + s_{i,t})\, w_{i,t} - \tau_t \right\}^{\alpha} \cdot w_{i,t}^{-\alpha} \left(\frac{A_{i,t}}{\pi_{ii,t}} \right)^{\frac{\alpha}{\theta}} \left(\frac{L_{i,t}}{H_i} \right)^{-(1-\alpha)} \end{aligned}$$

The migration shares under the income subsidies $\lambda^s_{ij,t}$ become

$$\lambda^s_{ij,t} = \frac{\mathcal{B}_{ij,t}\left(\left\{(1+s_{i,t}) - \frac{\tau_t}{w_{i,t}}\right\}^{\alpha}\left(\frac{A_{i,t}}{\pi_{ii,t}}\right)^{\frac{\alpha}{\theta}}\left(\frac{L_{i,t}}{H_i}\right)^{-(1-\alpha)}\right)^{\epsilon}}{\sum_{k\in N}\mathcal{B}_{ik,t}\left(\left\{(1+s_{k,t}) - \frac{\tau_t}{w_{k,t}}\right\}^{\alpha}\left(\frac{A_{k,t}}{\pi_{kk,t}}\right)^{\frac{\alpha}{\theta}}\left(\frac{L_{k,t}}{H_k}\right)^{-(1-\alpha)}\right)^{\epsilon}}$$

C.2 Counterfactual Aggregate Welfare Under Free Mobility

Under free mobility $\mu_{ij,t} = 1, \forall i,j \in N$ and given changes in productivity parameters $\left\{\hat{A}_{i,t}\right\}$, the common change in welfare between the two equilibria can be expressed as follows:

$$\hat{\bar{U}}_t = \left[\sum_{i \in N} \frac{L_{i,t}}{\bar{L}_t} \left\{ \left(\frac{\hat{A}_{i,t}}{\hat{\pi}_{ii,t}}\right)^{\alpha/\theta} \hat{L}_{i,t}^{-(1-\alpha)} \right\}^{\epsilon}\right]^{\frac{1}{\epsilon}}$$

C.3 Technological Autarky

In this counterfactual scenario, I assume that potential producers can only learn from local producers, that is, $K'_{ij,t} = 0, \forall i \neq j$, and $\bar{K}'_{i,t} = K'_{ii,t} = 1$. Given the stock of knowledge in the initial period, $A'_{i,t0} = A_{i,t0}$, the law of motion of $A'_{i,t}$ specified in (4.1) becomes:

$$A'_{i,t+1} - A'_{i,t} = \zeta'_{i,t} \cdot \Gamma\left(1 - \beta\right) \cdot \left(A'_{i,t}\right)^{\beta} \tag{C.3}$$

where $\zeta'_{i,t}$ is determined by the counterfactual city size $L'_{i,t}$ as in (4.7), in which I fix the value of the time fixed effect and the residuals as the original value: $\zeta'_{i,t} = \zeta_t \varepsilon_{i,t} \left(L'_{i,t}\right)^{\rho}$. A change in local productivity affects the allocation of labor across cities through changes in the relative income $\left\{v'_{i,t}\right\}$, while the values of the ease of migration $\{\mathcal{B}_{ij,t}\}$ are fixed at the original values.

The effect on the aggregate stock of knowledge is $\hat{A}_{t1} \equiv \sum_{i \in N} \left(L_{i,t0}/L_{t0}\right) \hat{A}_{i,t1} = 0.45$. Under the vector of counterfactual productivity, I get $\hat{U}_{t1} = 0.86$, indicating that the aggregate welfare loss without intercity idea flows is 14% relative to the original equilibrium during 2006-2015 under the counterfactual economy with technological autarky in 1996-2005[20].

To understand the magnitude of productivity and welfare gains from within-city and across-city knowledge flows, I decompose the log of productivity growth into two components:

$$\ln\left(\frac{x_{i,t1}}{x_{i,t0}}\right) = \ln\left(\frac{x'_{i,t1}}{x_{i,t0}}\right) + \ln\left(\frac{x_{i,t1}}{x'_{i,t1}}\right) \tag{C.4}$$

where the first term in the RHS of (C.4) accounts for the within-CZ effect, or the growth of variable x_i from the initial period t_0 to period t_1 under the counterfactual without intercity knowledge flows. The second term accounts for the across-CZ effect, considering the difference between the counterfactual productivity under technological autarky $\left(x'_{i,t1}\right)$ and the original equilibrium value $(x_{i,t1})$ in period t_1. (C.4) shows this decomposition by employment decile. For each employment bin, I take the weighted average of each term using employment shares within the bin. Black bars and gray bars show the first term

[20]This aggregate welfare effect is close to the welfare loss of 15% under free mobility.

and the second term in the RHS of (C.4), respectively, for productivity (panel (A)) and welfare (panel (B)). The within-CZ effects are non-monotonic in city size, reflecting the counteracting forces between the size effect and the diminishing returns in learning. To visualize this, rearranging (C.3), I obtain the following:

$$\frac{A'_{i,t1}}{A_{i,t0}} = 1 + \zeta_{i,t0} \cdot \Gamma\left(1-\beta\right) \cdot \left(A_{i,t0}\right)^{\beta-1}$$

which is increasing in $(\zeta_{i,t0})$ and decreasing in $(A_{i,t0})$ for the estimated parameter $\beta < 1$. The smallest cities have the lowest size effect, which offsets their higher growth effect due to the initially lower productivity. Conversely, the biggest cities have a sufficiently large population to offset diminishing growth[21]. Overall, the within-CZ effect explains 7-16% (19-30%) of the total productivity (welfare) growth, and the across-CZ effect explains more than 80% (60%) of the effects for all employment bins. For the aggregate productivity and welfare growth, the across-CZ effects explain 84% and 73%, respectively

[21]This diminishing effect of productivity on productivity growth is widely assumed in a class of macroeconomic models, as in Jones (1995), Kortum (1997), and Jones (2005). Empirically, Bloom et al. (2020) documents a sharp decline in research productivity at an aggregate level in recent years.

D Appendix Tables

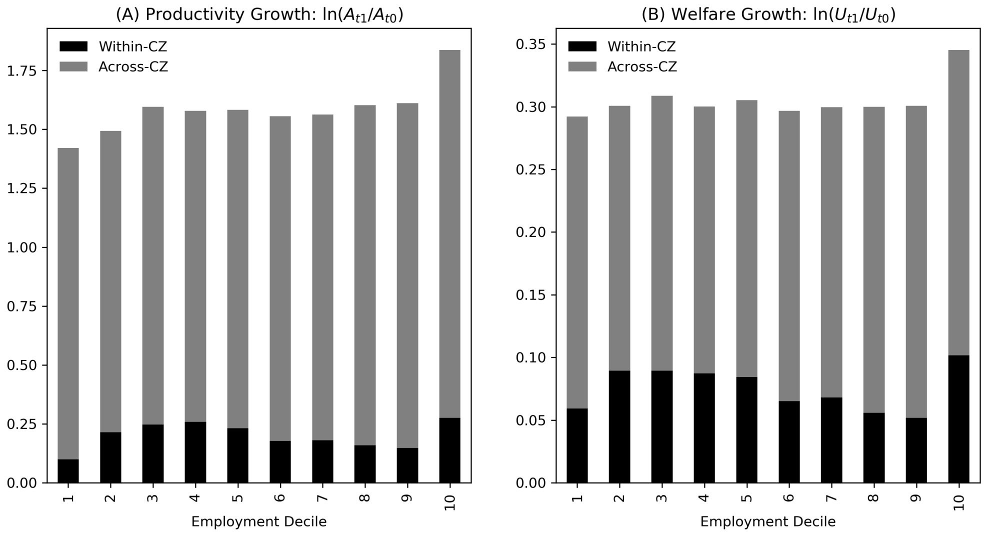
Table A.1: Dynamic Effects of Free Mobility Across All Cities

	Existing Flows	All City Pairs
Aggregate Stock of Knowledge	192.81%	281.05%
Aggregate Welfare	21.52%	24.76%
Coef. Var. Stock of Knowledge	−39.03%	−37.14%

Notes: This table shows the results of counterfactual analyses under free mobility across all cities. All values are expressed as changes from the original equilibrium in $t1 : 2006 - 2015$.

E Appendix Figures

Figure A.1: Decomposition of Productivity and Welfare Growth



Notes: This table shows the results of counterfactual analyses in (C.4).