Title
Essays in Urban Economics

Author
Truffa, Santiago

Publication Date
2017

Peer reviewed|Thesis/dissertation

Essays in Urban Economics

by

Santiago Truffa

A dissertation submitted in partial satisfaction of the
for the requirements of the degree of
Doctor of Philosophy
in
Business Administration
in
the Graduate Division
of the
University of California, Berkeley

Committee in Charge:
Prof. Ernesto Dal Bo (co-Chair)
Prof. John Morgan (co-Chair)
Prof. William Fuchs
Prof. Victor Couture
Prof.Cecile Gaubert

Summer 2017

Abstract

Essays in Urban Economics

by Santiago Truffa

Doctor of Philosophy in Business Administration

University of California, Berkeley

Prof. Ernesto Dal Bo (co-Chair) and Prof. John Morgan (co-Chair)

In Chapter I, titled "On the Geography of Inequality: Labor Sorting and Place-Based Policies in General Equilibrium", I study how city fundamentals, like amenities and housing restrictions, contribute to aggregate wage inequality through the sorting of heterogeneously skilled workers. I develop a "system of cities" model that features workers who differ along a continuum of skills and who compete for limited housing. This model is quantitatively tractable, and can replicate patterns in the dispersion of wages and housing prices both between and within cities. I calibrate this model to match different moments of the distributions of talent and wages for a cross-section of US cities, and I use it to understand the importance of sorting when accounting for patterns of regional inequality.

In Chapter II, titled "Urban Connectivity", I study how technological changes that affect the efficiency with which workers use their productive time in a city, can explain the increased spatial segregation in workers' skills and firms' productivity. I focus in an economy that produces knowledge and requires the matching of heterogeneous firms and workers. I provide a spatial equilibrium model that has the unique feature that allows for the sorting of a continuum of firms and workers where productive complementarities are city specific. I show that small changes in the connectivity of a city, can generate non-linear changes in city sizes and the level of skill segregation between cities. This suggest that small shocks to the productive environment of a city could account for the important changes we have observed in workers' skills and firms' productivity distributions.

Finally, in Chapter III, title Clustering to Coordinate: Who Benefits From Knowledge Spillovers?(joint work with William Grieser and Gonzalo Maturana), we study location and investment decisions by firms. We develop a model of knowledge sharing and derive the prediction that riskier and more complex industries experience the greatest gains from knowledge spillovers. Using tests that account for the non-randomness of location decisions, we find a strong positive relationship between industry risk or complexity and the clustering of: 1) firms' headquarters, 2) patent inventors, and 3) R&D expenses. Customer–supplier proximity is also significantly and positively related to industry risk and complexity.

Acknowledgements

I am indebted to my PhD advisors, Ernesto Dal Bo, John Morgan and William Fuchs, for their patience, guidance and support in achieving my doctorate milestone. Ernesto you were always there for me, from the beginning until the end, pushing me to give the best version of myself. John, I cannot tell you how much I enjoyed our conversations, where I always gain a new perspective on my projects and on research more generally. Willy, I have learn a lot from you. I have always counted on your deep insights and friendly support. Thanks for all the time we spent reading papers and discussing ideas.

Many other great scholars generously helped me along the way. I would like to especially thank Victor Couture, Cecile Gaubert, Rui de Figueiredo, and faculty members of the Business and Public Policy program whose suggestions, mentorship and support I benefited greatly from.

I am also thankful of the great co-authors I work with and hope to continue working, Gonzalo Maturana, William Grieser, Jose Antonio Espin-Sanchez, Yoonha Kim and Sheisha Kulkarni.

I am also deeply thankful to Maria Carkovic and the Sasakawa Foundation for generously funding my PhD years. I also thank my BPP peersboth current students and alums.

I am an exceptionally lucky person because of my family. I especially thank Dani my life companion who has not only given me the 2 most beautiful kids in the world, Luciano and Santino, but has always been there for me, in good times and also in bad times. Thanks for my parents in law, Enzo and Marcela, for all the loving support you have given us through all these years. Thanks to my parents, Santiago and Barbara, for giving me everything I needed to make my dreams come true.

Thank you all very much.

Contents

List of Figures

List of Tables

To Dani my love...

Chapter 1

On the Geography of Inequality: Labor Sorting and Place-Based Policies in General Equilibrium

1.1 Introduction

Rising wage inequality has been a defining feature of the US economy over the last few decades. This increase in inequality has been accompanied by changes in the organization of economic activity in space. This reorganization has generated a serious urban differentiation: cities that concentrate a higher fraction of high-skilled workers also feature higher wages and housing prices [1].

Prior research has shown that an important part of the spatial variation we observe in wages and housing prices is associated with the physical characteristics of cities. For instance, Gyourko et at [1] show that metropolitan statistical areas (MSAs) with low land availability tend to be more productive, and they attribute this effect to the sorting of high-skilled workers into these places. These same cities have been among the least likely to add new housing in the last couple of decades, and this could be exacerbating the regional divergence. This phenomenon has started a discussion (e.g., Hsieh and Moretti [2]) about the implications of local land use policies on overall growth and inequality.

In this paper, I study how city fundamentals contribute to wage inequality through the spatial distribution of the population. I develop a spatial equilibrium model that links city characteristics (e.g., amenities and housing supply) to the sorting of workers with a continuum of skills. This model is quantitatively tractable, and it is rich enough to replicate patterns in the dispersion of wages both between and within cities. The model features the existence of a unique equilibrium, which allows me to numerically quantify the general equilibrium effects of different types of place-based policies. I calibrate the model to match city-level moments for the distributions of talent and wages for a cross-section of US cities, and study how the spatial sorting of skills and the structure of wages

[1] Changes between cities have gone hand-in-hand with changes within cities, as more productive places have also become more unequal.

and housing prices are co-determined in equilibrium. In this context, I can turn off spatial sorting and evaluate what would happen to the economy in the absence of sorting effects. In the model, spatial sorting accounts for 7.5% of the aggregate wage dispersion, and makes the economy 1.9% more productive. Places that in equilibrium feature a tighter housing market (i.e., *superstar* cities) are on average 30% more productive and 40% more unequal due to sorting. Finally, I evaluate what would happen to the economy if houses were built in different cities, and I find that the economy would become 0.2–0.4% more productive from expanding the housing supply in *superstar* cities. However, I also find that relaxing housing constraints in superstar cities also tends to increase aggregate wage inequality by the same magnitude.

Regional inequality is a complex topic, and one that for several reasons requires general equilibrium reasoning. First, in a spatial equilibrium workers must not prefer other locations to their current one. This implies that population, wages, and housing prices in all cities should be considered simultaneously, making them interdependent. Second, wages and housing prices are inseparable from local agglomeration externalities and the distribution of skills available in the city (Combes et al.[3], Gennaioli et al. [4], Acemoglu and Dell [5], Van Nieuwerburgh and Weill [6]). Investigating this association requires a model in which labor productivity of each worker is endogenous to the location decisions of all workers. Third, to understand how city characteristics relate to regional inequality, we must disentangle how they affect agglomeration externalities and skill sorting separately. The reason is that a change in city characteristics (e.g., increasing the housing supply) could affect local labor productivity by increasing density and by changing the skill composition of the city.

I develop a general equilibrium model that allows for the sorting of heterogeneous workers with a continuum of skills in the presence of endogenous agglomeration externalities. In this model, workers care about their disposable income, which is their wages net of housing costs, and they care about the level of local amenities. Workers can freely choose where to live, but to access a city's amenities, they must consume one unit of housing. Since the number of houses per city is limited, workers compete for houses in bidding wars as discussed in Albrecht, Gautier, and Vroman [7]. By competing for limited housing, workers impose on each other a pecuniary congestion cost that depends on their skills through the bids they are willing to make. This differential externality can work as an endogenous gentrification force. The strength of this effect ultimately depends on the tightness of the housing market (i.e., how many buyers show up to compete for a house), and this effect directly links the features of a city's housing market to the characteristics of its talent pool. Since workers can freely move between cities, skill distributions are determined by a spatial equilibrium condition. Given that we have a continuum of skills, we can solve this functional problem using differential equations that make the model empirically tractable.

Workers in cities perform non-tradable tasks that combine in a single final good. Each task requires only labor and workers are differentiated in their labor productivity, which depends on workers' skills. Finally, workers can impose positive productive externalities on each another, and these externalities are ultimately a function of the endogenous number of workers producing in the city. The equilibrium of this economy yields, for each city, an endogenous wage distribution in which wages depend both on the local productivity and on the relative supply of skills.

In the calibration exercise, I use accepted parameter values from related literature, and the parameters that are unique to my specification, I recover from data. To do so, I constrain the model to generate city-level moments as close as possible to the equivalent empirical moments using the Current Population Survey (CPS) for 54 MSAs in the US in the year 2011. The calibrated model is consistent with several stylized facts. For instance, the model features a sorting of high-skilled workers into land-constrained cities. Given the higher supply of skilled workers, these cities specialize in high-productivity sectors, generating an endogenous correlation between low land availability and city productivity. Concurrently, land constrained cities have the property of having higher wage inequality driven by the interaction of endogenous relative prices and local agglomeration externalities. Furthermore, the calibrated model does a reasonable job of predicting moments that had not been targeted in the estimation, such as relative average house prices and city size. The model also predicts a positive correlation between average wages and wage dispersion that is consistent with the data.

Literature Review

This paper connects two important subjects in the urban economics literature. Extant research has shown that the physical geography of cities relates to their economic outcomes (Saiz [8], Ganong and Shoag [9], Hornbeck and Moretti [10]). A separate strand of the literature has focused on studying what determines the sorting of heterogeneous agent [2] [3]. Eeckhout et al. [20][21], Behrens and Robert-Nicoud [22] and Puga et al. [23] have developed discrete agent models that help theoretically explore how locations' fundamentals connect to productivity and inequality in cities [4]. My paper closely relates to Davis and Dingel [26] in using a continuum of skill types. I depart from their framework by providing a micro-founded housing market in which a restricted housing supply implies that, in equilibrium, a city may feature excess demand for housing (as in the superstar cities framework)[5]. As I am able to fully characterize and compute the unique equilibrium of the model using differential equations, I provide a tractable framework that is useful for quantitative policy evaluation.

[2] (Glasear [11], Shapiro [12], Couture [13], Albouy [14], Albouy and Seegert [15]) have empirically shown the importance of amenities in accounting for sorting patterns. I build on this literature, and I quantify the tradeoff between amenities versus restrictions on the housing supply.

[3] A related literature has explored the sorting of heterogeneous firms (Gaubert [16], Desmet and Rossi- Hansberg [17] and Behrens et al.[18], Serrato and Zidar [19]) to study the welfare implications of taxes and firm incentives. I complement this literature by focusing on the worker side. Further work is required to join these two threads in the literature.

[4] Frameworks that divide the workforce into discrete categories are empirically sensitive, since results depend on dichotomous definitions of what type of worker qualifies for each type of category. Indeed, Baum, Snow, Freedman, and Pavan [24] show that if we change the definition of high-skilled worker to a worker with some college education, some of the results shown by Diamond [25] no longer hold.

[5] To do so, I follow recent literature that models the housing market with bidding wars. For a review see Han and Strange [27].

This paper also relates to the literature that examines how city-level outcomes aggregate. Hsieh and Moretti [2] use a Rosen-Roback [28][29]framework to analyze the role of cities in aggregate growth. I contribute to this literature by providing a theory in which city productivity is endogenous to the interaction between sorting and local agglomeration externalities. A parallel literature has been studying wage inequality in cities (Baum-Snow and Pavan [30]). I provide a quantitative framework that can analyze the role of sorting when accounting for the dispersion of wages we observe within cities.

Finally, this paper also speaks to a growing literature that seeks to evaluate the aggregate consequences of place-based policies. Despite many examples of local program evaluations, it is hard to assess the general equilibrium effects of these types of policies [6]. I contribute to this literature by quantifying the aggregate potential consequences of local policy changes by means of the spatial sorting of heterogeneous workers.

This paper is organized as follows: Section 3.2 presents the model and the theoretical results. Section 1.3 discusses the data, describes the empirical estimation, and presents the main empirical results. Section 3.5 discusses policy implications and then concludes.

1.2. Model

I consider an economy that contains $N > 1$ cities. Cities have a heterogeneous endowment of housing supply S^i and amenities a^i. Non-tradable services are produced within each city.

1.2.1. Production within a City

The economy is populated by a continuum of workers with skill $s\epsilon[\underline{s}, \overline{s}]$ and these workers can move freely between cities. Denote $v^i(s) \geq 0$ as the endogenous supply of workers with skill s in city $i\epsilon\{1, ..N\}$.

Cities produce one final good, and producing that final good requires the aggregation of intermediate tasks which will be indexed by their skill intensity $\sigma\epsilon\Sigma = [\underline{\sigma}, \overline{\sigma}]$.

Production tasks are performed only through human capital, and workers vary in their productivity in these tasks. In particular, let $A(s, \sigma) > 0$ be the productivity of a worker of skill s in task σ [7]. $Y^i(\sigma) \geq 0$ is the endogenous output of task σ in city i and is given by,

$$Y^i(\sigma) = \int_{s\epsilon S} A(s, \sigma) L^i(s, \sigma) ds \tag{1.1}$$

[6] A notable exception is Kline and Moretti [31], who develop a methodology to estimate their aggregate effects.

[7] To capture the idea that high skill workers have a comparative advantage in more complex tasks, I assume as in Costinot and Vogel (2010), that productivity is log supermodular.

where $L^i(s,\sigma) \geq 0$ is the endogenous number of workers with skill s who work on task σ in city i.

The output of the final good is given by a Dixit-Stiglitz production function:

$$Y^i = \left\{ \int_{\sigma \epsilon \Sigma} E^i [Y^i(\sigma)]^{\frac{\varepsilon-1}{\varepsilon}} d\sigma \right\}^{\frac{\varepsilon}{\varepsilon-1}} \tag{1.2}$$

where, $\varepsilon > 1$ is the constant elasticity of substitution between tasks and E^i is a city-level productivity shifter that captures the effect of agglomeration externalities.

Total profit for the final good is given by,

$$\Pi^i = \left\{ \int_{\sigma \epsilon \Sigma} E^i [Y^i(\sigma)]^{\frac{\varepsilon-1}{\varepsilon}} d\sigma \right\} - \int_{\sigma \epsilon \Sigma} p^i(\sigma) Y^i(\sigma) d\sigma \tag{1.3}$$

where $p^i(\sigma)$ is the endogenous price of task σ in city i.

Finally, total profits for intermediate tasks are given by,

$$\Pi^i(\sigma) = \int_{s \epsilon S} [p^i(\sigma) A(s,\sigma) - w^i(s)] L^i(s,\sigma) ds \tag{1.4}$$

where $w^i(s)$ is the endogenous salary for a worker with skill s in city i.

1.2.2. Housing market

I follow a directed-search model as in Albrecht, Gautier and Vroman [7][AGV] to portray the housing market. Let $\theta^i = B^i/S^i$, be the tightness of the housing market in city i. Where B^i is the total number of workers that bid for houses in city i, and S^i is the total amount of houses for sale.

The game has several stages:

1. Buyers randomly arrive to compete for a house.
2. Each buyer has a private valuation x, which will be the buyer's wage in city i. Buyers do not observe the number of other visitors to the house.
3. As buyers arrive at a house they compete for it following a first-price auction (with an un-known number of competitors).

With a random number of buyers, an individual buyer's optimal bid is the weighted average of the buyer's optimal bids conditional on competing with $n = 0, 1, 2...$ as shown by AGV.

$$b(x) = \frac{\sum p_n F^n(x) b(x;n)}{\sum p_n F^n(x)}$$

where p_n is the probability of a buyer competing with n other buyers, and $b(x;n)$ is the optimal bid and $F()$ is the distribution of types in the market. I assume B^i and S^i are large enough so that, in the limit, the arrival rate of buyers visiting a particular seller follows a continuous Poisson process with parameter θ^i. This characterization yields the endogenous housing prices shown in Appendix 1.

The tightness of the housing market also determines the probability of obtaining a house in the city. Since buyers' valuations for a house are sampled from a common distribution, we have a "Poisson race" among different players who arrive following the same arrival rate θ^i. Thus, $e^{-\theta^i(1-F^i(w^i(s)))}$ is the probability that a buyer of skill s makes the highest bid, and thus wins the auction. Notice that θ^i depends on the total number of workers who arrive in equilibrium to produce in city i. I assume that workers have rational expectations, thus they correctly anticipate the value of θ^i in equilibrium before moving into a city.

1.2.3. Workers' Preferences

Workers' have homothetic preferences over their disposable income and the quality of local amenities. Homothetic preferences can be represented as $U = Ta^i log(x_i)$. For convenience, I will use the monotonic transformation $U = exp(U) = x_i e^{Ta^i}$, as all the properties of this utility representation hold for monotonic transformations. In this case, x_i is the disposable income and e^{Ta_i} is a utility shifter that depends on the amenities in that city a^i and on a deep preference parameter T, which captures the tradeoff between disposable income and the quality of local amenities. Workers enjoy amenities only in the case in which they can obtain a house in the city. When they cannot obtain a house, they will receive a reservation utility $\underline{u}(s)$ from living in the suburbs. Workers are risk-neutral in whether they get to live in the suburbs or not.

Disposable income, is given by wages net of housing costs. Let $b^i(w^i(s))$ be the optimal bid of a worker of skill s for a house in city i. Then, $x_i = w^i(s) - b^i(w^i(s))$.

Hence, the expected value that a buyer of skill s would receive from producing in city i is

$$U(w^i(s)) = (w^i(s) - b^i(w^i(s)))e^{Ta_i}e^{-\theta^i(1-F^i(w^i(s)))} + \underline{u}(s)(1 - e^{-\theta^i(1-F^i(w^i(s)))}) \qquad (1.5)$$

other workers in the city. In this context, using auctions to characterize the competition for limited housing can enormously simplify the problem. In particular as the optimal bid for a house follows from a first price auction, then

$$b^i(w^i(s)) = w^i(s) - \frac{F^i(w^i(s))}{f^i(w^i(s))}$$

This means that the disposable income of a worker is given by,

$$x_i = w^i(s) - b^i(w^i(s)) = \frac{F^i(w^i(s))}{f^i(w^i(s))}$$

Since an important fraction of labor rents accrue to land prices, the disposable income of a worker in city i is going to be equal to her virtual surplus (i.e., informational rents). This means that the level of wages disappears, and utility depends only on her relative wage in city i:

$$U^i(w^i(s)) = \frac{F^i(w^i(s))}{f^i(w^i(s))} e^{-\theta^i(1-Fi^i(w^i(s)))} e^{T*a_i} + \underline{u}(s)(1 - e^{-\theta^i(1-F^i(w^i(s)))}) \tag{1.6}$$

This characterization will allow us to further simplify the problem. Since wages are a monotone function of talent, we can perform a change of variables so we can work in the space of types. That is, there has exist a function V^i, such that $V^i(s) = F^i(w(s))$.

This means that

$$U^i(s) = \frac{V^i(s)}{v^i(s)} e^{-\theta^i(1-V^i(s))} e^{T*a_i} + \underline{u}(s)(1 - e^{-\theta^i(1-V^i(s))}), \tag{1.7}$$

where $V^i(s)$ is the endogenous cumulative distribution function (CDF) for talent in city i, and $v^i(s)$ the endogenous probability distribution function (PDF).

Notice the relevance of this change of variables and how it simplifies the overall problem. The utility of a worker of living in city i is only a function of fundamental city characteristics, and the endogenous distribution of skills in that city. This allows me to solve for the equilibrium in two separate parts: I solve for the spatial sorting first, then I solve the equilibrium for each city separately. This simplification allows the model to be computationally tractable.

1.2.4. Equilibrium

The equilibrium of this economy is driven by two requirements: a spatial equilibrium condition for each type of worker and a competitive equilibrium condition for each city. I first study how workers sort across space only as a function of city attributes. After solving for the spatial distribution of talent, I solve for wages and housing prices in each city.

Spatial Equilibrium: Sorting

Since there is free mobility, the utility of a worker of ability s must be equal across space.

I restrict attention to skills distributions on a close interval $[\underline{s}, \overline{s}]$ in which all cities share the same support of skills.

This is an important assumption that allows me to solve and characterize the spatial equilibrium of the economy. This condition tells us that, in equilibrium, all workers must be indifferent between all cities. This is consistent with empirical distributions of talent. Although we see differences between cities in the fraction of high-skilled to low-skilled workers, we still observe a positive mass of workers at every level of talent. Moreover, if we restrict attention to two cities, it is easy to show that, for any pair of non-overlapping skill distributions, this configuration would never be in equilibrium, since the lowest-skill worker in the high-skilled city will always have an incentive to move to the low-skilled city, where she would be the most skilled worker.

In order to use this model empirically, we must have a strategy to recover $\underline{u}(s)$. One possibility is that, given the costs of commuting to the center of the city, individuals capture only a fraction ϕ of living in city i. Thus, we could write utility in the following way:

$$U^i(s) = \frac{V^i(s)}{v^i(s)} e^{T * a_i} [e^{-\theta^i (1 - V^i(s))}(1 - \phi) + \phi] \tag{1.8}$$

Notice that ϕ captures workers' sensitivity to not finding space in the city. If workers could commute at no cost, then ϕ would equal 1 and therefore housing restrictions would not matter. Given this preference representation, changes in T and ϕ are observationally equivalent for empirical purposes, since both speak to the tradeoff between disposable income and amenities. I make the identifying assumption $\phi = 0$. Theoretical results would not change for any $0 < \phi < 1$.

For each worker of skill s, it must be the case that

$$EU^1(s) = EU^2(s) = .. = EU^N(s) \quad \forall s \epsilon [\underline{s}, \overline{s}] \tag{1.9}$$

Notice that, if this condition holds, the condition must hold for any monotonic transformation of the expected utility, and it must hold for any linear combination of such transformations. In particular, the above condition could be translated into a system of equivalent conditions as follows:

$$ln[EU^i(s)] = \frac{1}{N-1} \sum_{j \neq i} ln[EU^j(s)] \ \forall s \epsilon [\underline{s}, \overline{s}] \quad \forall i \epsilon [1, .., N] \tag{1.10}$$

$$ln[\frac{V^i(s)}{v^i(s)} e^{-\theta^i (1 - V^i(s))} e^{T a_i}] = \frac{1}{N-1} \sum_{j \neq i} ln[\frac{V^j(s)}{v^j(s)} e^{-\theta^j (1 - V^j(s))} e^{T a_j}]; \tag{1.11}$$

Using the properties of the log function, we can rearrange the above system into

$$ln[v_i] = ln[\prod_{j \neq i} v_j^{1/N-1}] \frac{V_i}{\prod_{j \neq i} V_j^{1/N-1}} e^{-\frac{1}{N-1}\sum_{j \neq i} \theta^j (1-V^j(s)) + \theta^i (1-V^i(s))} e^{T[a_i - \frac{1}{N-1}\sum_{j \neq i} a_j]}]; \tag{1.12}$$

The system above cannot be easily solved numerically or characterized by a closed-form solution. Nevertheless, if we make a linear approximation to one term of the above expression, the system is easy to characterize and solve. We must rearrange this system of differential equations in such a way that we can write all derivatives on the left side of the equation and write all primitives on the right side, as follows:

$$v_i = F(\vec{V}, \gamma) \tag{1.13}$$

In order to do this, we will approximate the geometrical average of probability density functions (PDFs), where $0 < vj << 1$, with its arithmetic average.

That is,

$$\prod_{j \neq i} v_j^{1/N-1} \equiv G_i$$

represents the geometric average of the probability density functions that are different from i. By taking a second-order Taylor expansion of G_i, I approximate the geometric average by an arithmetic average[8]. In particular, it is the case that

$$G_i \approx A_i - \sigma^2/2$$

where $A_i = \frac{1}{N-1} \sum_{j \neq i} v_j$ and $\sigma^2/2$ is the sample variance.

I can now take advantage of the fact that this is a closed system, hence the sum of densities must be equal to the overall number of workers in this economy:

$$\sum_{j \neq i} v_j(s) + v_i(s) = v(s); \forall s \tag{1.14}$$

where $v(s)$ is the exogenous aggregate PDF.

Thus I can rewrite

$$G_i = \frac{1}{N-1}(v - v_i) - \sigma^2/2$$

[8]Note that we are dismissing the third-order terms, which are very close to zero in the case of probability densities.

This means that I can rewrite the above expression in such a way that the density function is a function of all the probability distributions. By exponentiating both sides of the equation, we arrive at

$$v_i = [\frac{1}{N-1}(v - v_i) - \frac{\sigma^2}{2}][\frac{V_i}{\prod_{j\neq i} V_j^{1/N-1}} e^{-\frac{1}{N-1}\sum_{j\neq i}\theta^j(1-V^j(s))+\theta^i(1-V^i(s))} e^{T[a_i - \frac{1}{N-1}\sum_{j\neq i} a_j]}]; \tag{1.15}$$

Finally, we can separate this expression, such that probability density functions are on the LHS,

$$v_i = \frac{1}{(1 + \frac{1}{N-1}\frac{V_i}{\prod_{j\neq i} V_j^{1/N-1}} e^{-\frac{1}{N-1}\sum_{j\neq i}\theta^j(1-V^j(s))+\theta^i(1-V^i(s)} e^{T[a_i - \frac{1}{N-1}\sum_{j\neq i} a_j]})}$$

$$[\frac{1}{N-1}v - \frac{\sigma^2}{2}][\frac{V_i}{\prod_{j\neq i} V_j^{1/N-1}} e^{-\frac{1}{N-1}\sum_{j\neq i}\theta^j(1-V^j(s))+\theta^i(1-V^i(s)} e^{T[a_i - \frac{1}{N-1}\sum_{j\neq i} a_j]}].$$

Two things must be noted here. First, from the above expression, I consider the problem of solving for the endogenous PDF as equivalent to the problem of solving for a mixing probability, since $\beta_i(s) \equiv v_i(s)/v(s)$ [9]. Thus, if I am to determine the endogenous PDF in equilibrium, I could determine the mixing strategy that players should follow in equilibrium. Second, as stated above, the above expression can be represented as

v$_i = F(\vec{V}, \gamma)$ (1.16)

which means that each probability density is a function of the vector of all probability distributions and a set of parameters γ.

Now, it is crucial to note that since this must hold for every point, and since there is a continuum of skills and types, I can consider the properties of a continuous function. On the other hand, given that, by definition, $v^i = \frac{\partial V}{\partial s}^i$, there is a system of ordinary differential equations. This simplifies everything, since by establishing certain properties about the function $F(\vec{V}, \gamma)$, I can actually say something about the existence and uniqueness of an equilibrium. This characterization will allow me to numerically compute the equilibrium for different sets of parameters.

[9] Workers must be indifferent between all cities, thus, in equilibrium, each worker must choose a mixing probability $b(s)$ by which each worker of skill s randomizes between cities. If all workers follow the same mixing strategies, then, in equilibrium, we should expect to see the distributions predicted by such mixing probabilities.

The above characterization will allow me to numerically compute the equilibrium for different sets of parameters.

Proposition 1.1. *There exists a unique spatial equilibrium to the system of cities model*

Proof. see the Appendix. □ □

Given that this is an ODE, and the right side of the equation is continuously differentiable, there exists a function that can solve for the spatial equilibrium. Let us refer to this unique function as $\vec{v}$.

Competitive Equilibrium

Once I have solved the spatial equilibrium above, I can compute the competitive equilibrium for each independent city.

Lemma 1.2. (Competitive Equilibrium)

In a competitive equilibrium, all firms maximize their profits and markets clear. This equilibrium can be characterized by a continuous and strictly increasing matching function $M^i : S \to \Sigma$ *such that (i)* $L^i(s, \sigma) > 0$ *if and only if* $M^i(s) = \sigma$ *and (ii)* $M^i(\overline{s}) = \overline{\sigma}$ *and* $M^i(\underline{s}) = \underline{\sigma}$.

We can characterize the matching function and wage schedule by a pair of differential equations for each city:

$$\frac{dM^i}{ds} = \frac{A[s, M^i(s)]V^i(s)}{\left[\int_{s \epsilon S} w^i(s)v^i(s)ds\right]\left[p^i(M^i(s))/E^i(M^i(s))\right]^{-\varepsilon}} \tag{1.17}$$

$$\frac{dLnw^i(s)}{ds} = \frac{\partial LnA[s, M^i(s)]}{\partial s} \tag{1.18}$$

where $v^i(s)$ *is the endogenous probability density function and* $V^i(s)$ *the cumulative density function for city* i *that was solved from the spatial equilibrium.*

Proof. See Costinot and Vogel [32] lemma 3. □ □

Two-Cities model

Given that for the general system of cities I can only characterize an approximation, I will study the case with two cities, in which the approximation coincides with the exact solution of the model. As I am able to provide analytic results for this specific case, these will be informative of the general properties of the model. In particular, we want to understand the role of housing restrictions in the model, and how these translate into different allocations of talent and consequent wage differentials between cities.

Proposition 1.3. (Monotone likelihood ratio property) *For two cities that have the same level of amenities, whenever $\theta^1 > \theta^2$, the monotone likelihood property for the distribution of talents in City 1 versus City 2 holds:*

$$v^1(s')v^2(s) \geq v^2(s')v^1(s) \qquad \forall s' > s$$

Proof. See Appendix 2. □ □

This property captures the idea that City 1 (i.e., the city with a more inelastic housing supply) will have relatively more high-skill workers than City 2.

Proposition 1.4. (Skill upgrading)

For two cities that have the same level of amenities, whenever $\theta^1 > \theta^2$,

then $M^1(s) \leq M^2(s) \quad \forall s$

Proof. See Appendix 2. □ □

From a worker standpoint, moving from City 2 to City 1 implies a task downgrading. From a task standpoint, each task will be performed by workers of higher ability in City 1 versus City 2. The intuition for this result is very simple, as the relative supply of high-talent workers is higher in City 1; and, since markets clear, more tasks must be performed by high-skilled workers. This shifts the M schedule downward.

Proposition 1.5. (Existence and uniqueness of competitive equilibrium)

There exist a unique equilibrium for each competitive equilibrium in a city

Proof. Given existence and uniqueness of the sorting equilibrium, we must prove the existence and uniqueness of the competitive equilibrium. Again, given that we have a system of ODE, it is clear that for any pair of functions A() and B() that are continuous and of bounded variation, then we can write $dM = F^M(M, w, \gamma)$ and $dw = F^w(M, w, \gamma)$. Given that F^M and F^w are continuous and of bounded variation, there exists a unique equilibrium for the system of ordinary differential equations. □ □

1.3. Calibration

To calibrate the model I will follow a simple strategy. I will have two types of parameters. General parameters that have already been estimated in related literature's, and new parameters that are specific to my theoretical specification. For the general parameters, I use accepted parameter values from related literature, and the parameters that are unique to my model I recover from data.

There are two parameters specific to my model. The first parameter is one that quantifies the relative taste for amenities T. The second is one that captures the complementarity between workers' skills and job complexity A. Then there are two general parameters we can retrieve from the literature, an elasticity of substitution between services or skills ε, and a parameter that captures local agglomeration externalities η.

First I will describe each parameter and the functional assumptions used in the exercise. Next, I will describe the data used in the calibration. I will then discuss the estimation techniques employed and finally, I show my results.

Parameters and Functional Assumptions

We first need an aggregate distribution of skills for the economy, which we will then "sort" between cities. We will set the domain of this aggregate skill distribution to be $[0, 1]$ and assume that the distribution of skills in the economy $v(s)$, will follow a truncated normal, centered in 0.5 with dispersion 1.

Taste for amenities: T

The first parameter we must recover is the taste for amenities parameter,T , which shifts the utility of workers by e^{Ta^i} , determining worker's preferences for places. This parameter quantifies the tradeoff between housing tightness and the quality of services provided by a city (as well as other intangibles such as *weather*or *natural beauty*). A higher taste for amenities means that workers tend to prefer living in that region, thus they are willing to risk going into a city, even when facing a tough housing market that could leave them living in the periphery. Although we assume that the taste for amenities is the same for all workers, the trade off this entails is distinct for different type of workers. In particular, high-skilled workers who know they can get a house in the city center and still have higher disposable income will tend to value the amenities of a city more.

Given that this parameter fundamentally speaks to how workers of different skills sort between cities, we will recover it from empirical distributions of skills.

Technology shifter: A

One important theoretical assumption is that high-skilled workers have a comparative advantage in performing more complex tasks, and that complementarity must be log-supermodular.

I will use a simple parametric characterization for empirical purposes: $A(s, \sigma) = e^{As\sigma}$, where A is a technology shifter which is recovered from the data.

The parameter A captures the advantage that higher-skilled workers might have over lower-skilled workers. This parameter does not affect how workers sort between cities, but it directly affects the level and dispersion of wages. We will recover this parameter from the observed distributions of wages by city.

Agglomeration externalities: η

The model assumes that density can make workers more productive. For each city, we impose an endogenous productivity shifter, which is a function of city size. I denote E^i to be a city-level productivity shifter, where $E^i = g(B^i)$, and $g()$ is a function of the endogenous number of workers that come to produce in city i, B^i.

Moretti and Klein [31] show that the elasticity of agglomeration externalities with respect to density is constant. So we will model agglomeration externalities following a power function such that $g_i(B^i) = (B^i)^\eta$, where η is the agglomeration parameter, which is set to 0.08, as estimated by Moretti and Klein (2014).

Notice that agglomeration externalities affect workers differently. High-skilled workers see higher productivity gains from these local shifters.

Elasticity of substitution between services: ε

In the equilibrium of this economy, the final parameter that we must numerically compute is an elasticity of substitution between services or tasks for the Dixit–Stiglitz function. In the literature, several authors have estimated an elasticity of substitution between skilled and unskilled workers in the range of [1,2] (see Katz and Murphy [33] and Ciccone and Peri [34]) . On the other hand, Hsieh and Klenow [35] use an elasticity of substitution between manufacturing goods equal to 3. For the main results, I use an elasticity of substitution equal to 2. Nevertheless, results are consistent for this elasticity varying between 1 and 3.

Data

I use the Current Population Survey (CPS) for March 2011. The CPS provides the wages for each MSA as well as the number of years of completed education, which I use as a proxy for talent. Due to the many limitations of using education as a proxy for talent (see Bacolod et al [36]), I also use alternative talent data as a robustness check (see Appendix 4). Results are very robust to the use of this alternative talent data.

The model also requires other sources of external information. For instance, the model requires a proxy for the relative value of amenities in each MSA a^i. I use the hedonic parameters computed by Albouy [14] [10] to proxy for the amenity value in each city. The advantage of using this measurement is that it is divided into two parts: and *endogenous*

[10] He developed a methodology that can derive hedonic measures of local productivity and local amenities from data such as local wages, housing prices, and taxes. The quality of life measure positively correlates to measures of natural amenities relating to climate and geography.

productivity element (dependent on the skill composition of the workforce) and a *quality of life* element. I use the second part of the amenity index, since it is exogenous to the sorting of talent.

The model also requires an exogenous measure of the housing stock in each MSA S^i. I use the housing supply elasticities estimated by Saiz [8][11], and assume that the supply of new houses in each MSA is proportional to their respective elasticity multiplied by the housing stock. We use the part of the elasticity that is determined by geographical restrictions.

Finally, to check the quality of the fit of this model, I evaluate the performance of this calibrated economy when predicting non-targeted moments. To do this, I specify two variables that are endogenous to my framework: the relative housing prices and the relative population size for each MSA. To proxy for the relative value of housing in each MSA, I use the Zillow price index for the year 2011. The population for each MSA is taken from the 2010 Census.

Indirect Inference Estimation

Since this theory does not feature a closed-form solution, I must resort to alternative procedures to recover the parameters for workers' taste for amenities and the technology shifter. I intend for the model to generate moments of the skill distribution that are as close as possible to the observed mean and variance of skills by city. Analogously, I also intend for the mean and variance of the wage schedule to be as close as possible to the one observed for each city.

Given that I can solve the theoretical model in two separate parts (i.e., the spatial sorting of workers and the competitive equilibrium in each city), I can also divide my estimation procedure into two parts. First, I solve the spatial sorting problem, in which I must recover only one parameter: taste for amenities. Once I achieve an endogenous distribution of skills for each MSA, I can solve for the distribution of wages in each city afterward, because the cities produce non-tradable services. From these wage distributions, I can recover the second parameter: the skill–technology complementarity.

Recover T from talent data

I simulate the model for different parameter values of T such that the model's predictions for the cross-section are as close as possible to the data. I define a Wald-type loss function that weights the distance between the predicted and the observed mean and variance of the talent distribution for each MSA.

The indirect inference estimator is thus given by,

[11] This measure stems from satellite-generated data on terrain elevation and the presence of water bodies to estimate the amount of developable land in each MSA.

$$\hat{T} = argmin_T L(T) = (\rho - \hat{\rho}(T))' W (\rho - \hat{\rho}(T))$$

where ρ are the data moments, $\hat{\rho}(T)$ are the simulated moments, and W is a positive definite weighting matrix, that captures relative city size.

I use a grid search procedure and divide the estimation process into four consecutive steps:

Step 1: Partition the parameter space of T[12]

Step 2: Run the model and compute auxiliary vector $\hat{\rho}(T)$.

Step 3: Compute the criterion function $L(T)$.

Step 4: Repeat steps 1-3 to minimize $L(T)$.

Using this process, I find that the parameter that minimizes this distance is equal to 7[13].

Notice that the taste parameter requires two different kinds of exogenous data, amenities a^i and housing S^i, to uniquely predict two features of the skill distribution: shape and size.

Technology parameter (A)

I follow the same strategy as in the first part of the calibration.Now, however, I can solve the model for wages for each city independently. I start by choosing a parameter value for both the agglomeration parameter and the elasticity of substitution (as already discussed). I also take the distributions of skills by city as given from the above problem. I simulate the model for different parameter values of A such that the model's predictions for the distribution of wages in each city are as close as possible to the data. As before, I define a Wald-type of loss function, which is a weighted distance between predicted and observed moments for the distribution of wages in each MSA. As before, I focus attention of the first and second moments of each distribution. The indirect inference estimator is thus given by

$$\hat{A} = argmin_A L(A) = (\rho - \hat{\rho}(A))' W (\rho - \hat{\rho}(A)),$$

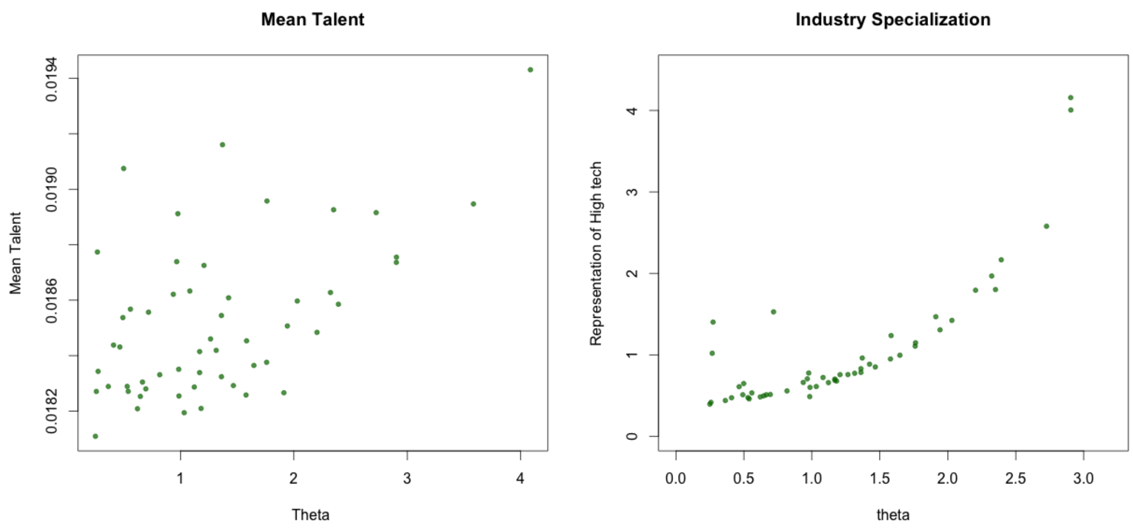
where ρ are the data moments, $\hat{\rho}(T)$ are the simulated moments, and W is a positive definite weighting matrix that captures relative city size.

Following the same procedure as in the first part, I find that the parameter that minimizes this distance is equal to 1.36.

[12]I let T vary from -100 to 100. I took steps of size 0.1, to ran the program for 2000 possible values of T.

[13]Which is very consistent with the 7.4 that we find using the Lumosity sample.

FIGURE 1.1: Sorting and Specialization in Superstar Cities



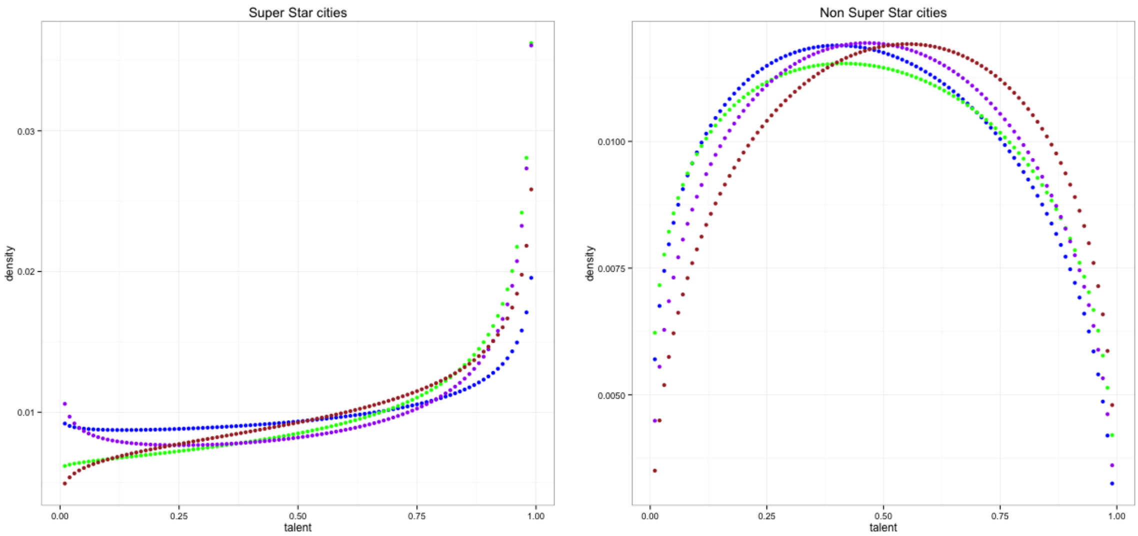
Calibration Results

Internal structure

With the calibrated model, I explore the numerical predictions in terms of the relations between the main endogenous outcomes; specifically, the distribution of wages, housing prices, and talent by MSA versus fundamental city characteristics. In particular, I want to explore how these variables change as a function of housing market tightness. To do so, I measure housing market tightness with *theta*, which is the equilibrium ratio between buyers and sellers in a given city. Note that theta speaks to both of the dimensions considered as city characteristics, since high-theta cities are the result of a combination of high amenities and a restrictive housing supply.

Figure (1) Panel (a) shows that, in equilibrium, more talented workers tend to concentrate in high-theta cities, which I will refer to as *superstar* cities. The average talent of workers increases as we move to cities that feature a tighter housing markets. Although all cities produce all goods in equilibrium, the fraction of these goods will depend on the endogenous supply of talent available in each local economy. Cities will be able to specialize. While all cities must produce all tasks, the relative output between intermediate tasks may vary between cities. Panel (b) plots the degree of specialization for each city. The model predicts that superstar cities tend to specialize in tasks that require more talent and tasks for which technology is used more intensively.

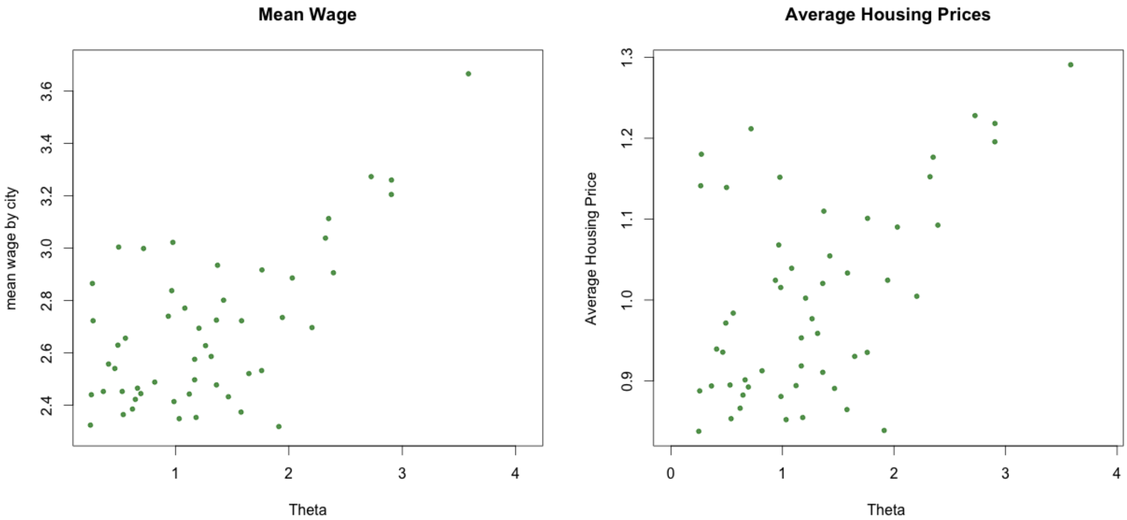
Figure 1.2: Talent Distributions



Panel (a) plots the distribution of skills for superstar cities. Panel (b) plots the distribution of skills for non-superstar cities

Figure (2) shows the skill distributions predicted by the model for a subset of high-theta cities and the skill distributions predicted by the model for low-theta cities. *Superstar* cities feature talent polarization. The skill distributions have "fat tails" both for high- and low-skilled workers.On the contrary, low-theta cities have the property of having bell-shaped talent distributions. This translates into important economic differences between these places.

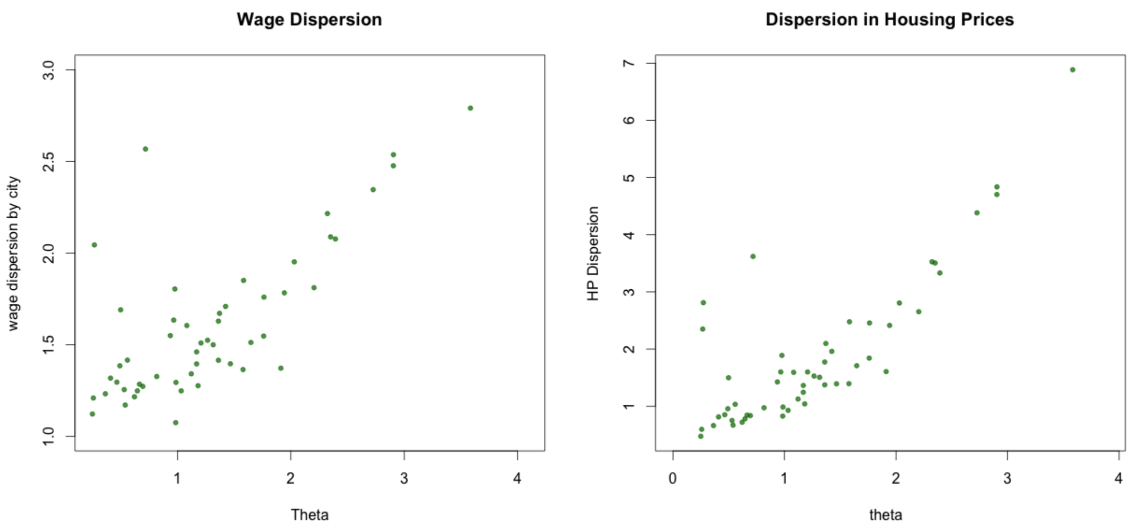
Figure 1.3: Wage and Housing Prices in Superstar Cities



Panel (a)average wage in each MSA . Panel (b) housing price in each MSA .

Figure (3) shows how these differences in talent distribution render significant differences in local productivity. Superstar cities have higher average wages, and housing prices also increase monotonically as a function of housing tightness. Sorting not only generates between-city differences; they also cause important variation within cities.

Figure 1.4: Wage and Housing Price Dispersion in Superstar Cities



Panel (a) plots the wage dispersion in each MSA. Panel (b) plots the housing price dispersion in each MSA.

Figure (4) shows that superstar cities are also more unequal places. Although workers must be indifferent between cities, their disposable income varies dramatically. It is important to note that this is fundamentally a consequence of the distribution of workers that live in these places.

Figure (5) Panel (a) plots the disposable income of workers in the 20th percentile across cities. Low-skilled workers must be compensated for living in superstar cities, since they have very low chances of finding a house within the MSA, thus they do not get to enjoy their amenities.The low supply of workers makes their relative talent very valuable, therefore this is reflected in relative prices and wages. Figure (5) Panel (b) shows how disposable income varies for workers in the 80th percentile of the talent distribution. Although these workers have higher wages, they also tend to pay a higher fraction of their wages for living in the city center. As they are very sensitive to amenities, they are willing to sacrifice some income to enjoy urban living. Figure (5) Panel (c) plots the ratio of disposable income between 80th-percentile workers and 20th-percentile workers. The difference in disposable income decreases dramatically as we move into cities that feature a tighter housing market. Although superstar cities are more unequal places because of their bimodal talent distributions, they tend to diminish differences in disposable income between high-and low-skilled workers.

External Validity

We must investigate capacity of the calibrated model for predicting patterns that were not targeted in the original exercise. To evaluate this, I focus on two important predictions: the relative size of cities and their relative housing prices.

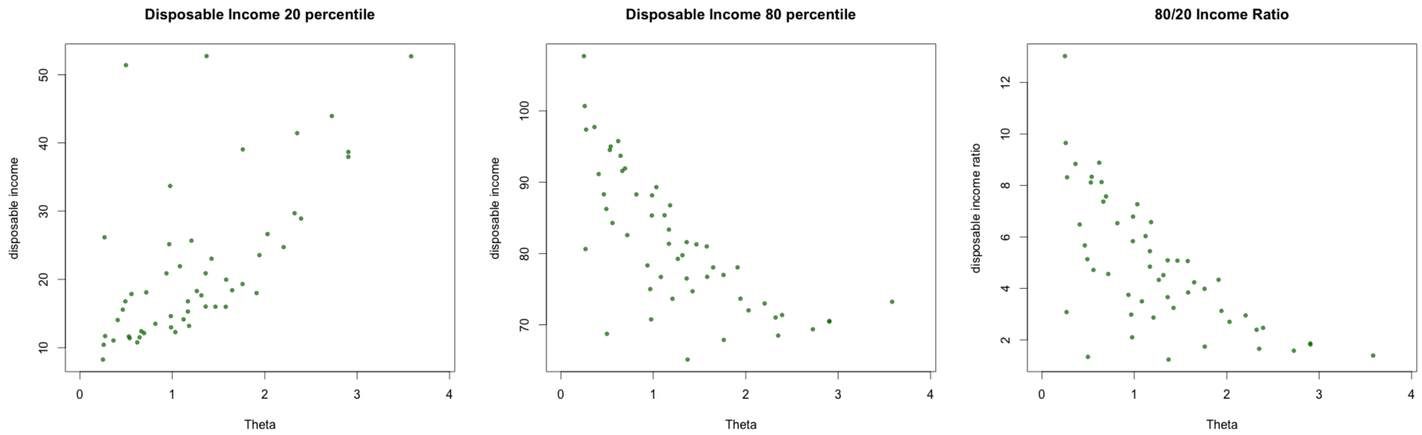
Figure (6) plots the predicted-versus-real city size and average housing prices. The model performs well in terms of predicting both relative city sizes and relative housing prices.

The calibrated model can also predict certain patterns in the correlation structure of the data, which was not originally targeted. Notice that the calibration exercise involved targeting the level of wages and wage dispersion independently (not their correlation). Figure (7) shows the correlation that the calibrated model predicts versus the correlation observed on the data using the CPS.

Unpacking Sorting Effects

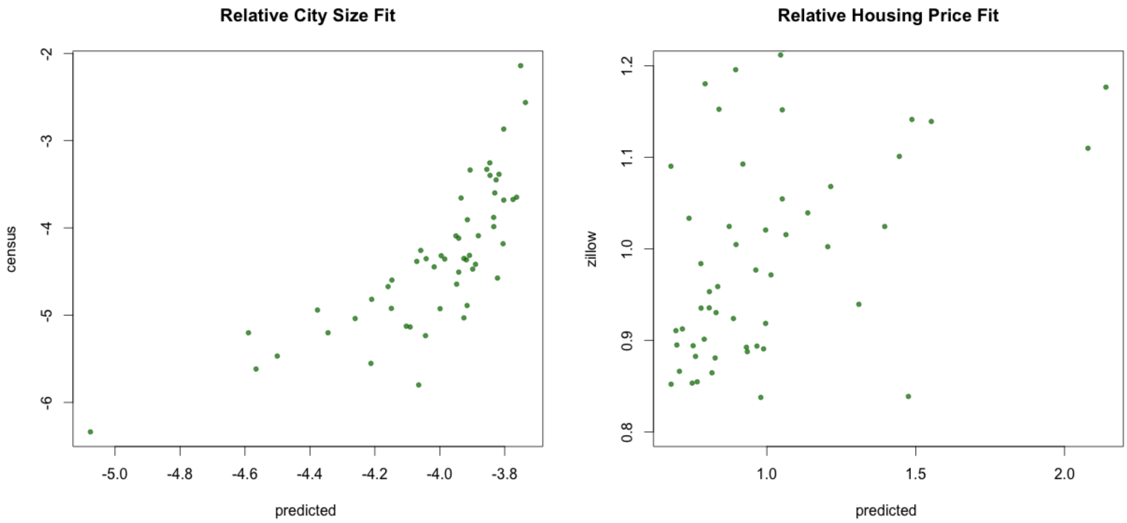
Since the location decisions of workers are not driven by local productivity differences, but rather by city characteristics, I can separate what fraction of the wage dispersion is due to the sorting of heterogeneous agents between cities and what fraction of the wage dispersion is due to local agglomeration externalities. For the first decomposition, I fix the size of each city and impose the same skill distribution everywhere. For the model (without sorting) I recalculate the dispersion in wages and housing prices both between and within cities.

Figure 1.5: Disposable income by city type



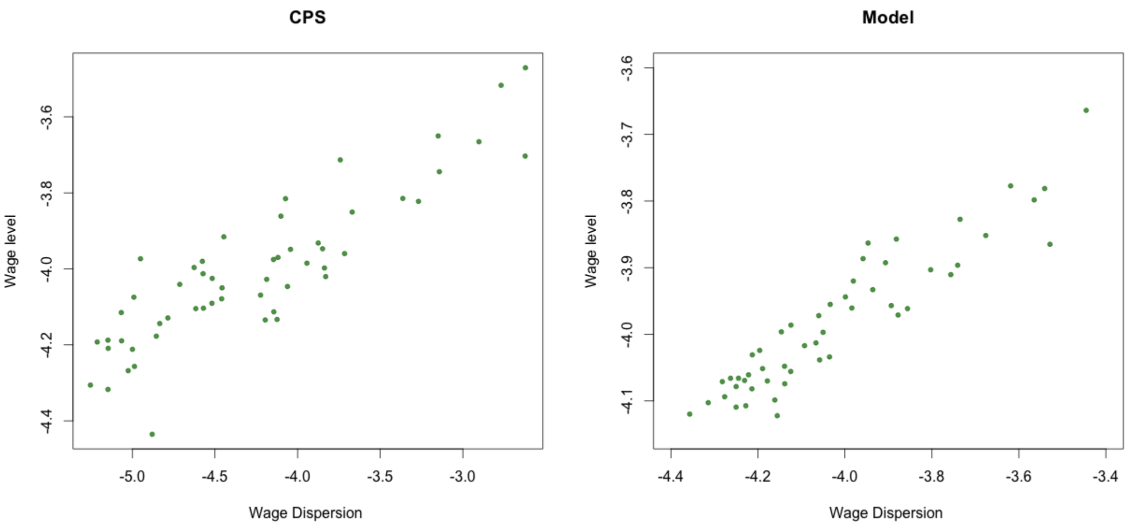
Panel (a) disposable income 20th percentile worker by city type.
Panel (b) disposable income 80th percentile worker by city type.
Panel (c) disposable income ratio between 80th percentile and 20th percentile worker by city type.

Figure 1.6: Non-targeted Variables



Panel (a) observed relative city sizes (from Census 2010) versus predicted relative city sizes. Panel (b) observed relative average housing price from the Zillow housing price index versus the relative housing prices predicted by the model.

Figure 1.7: Internal Correlation Structure: Correlation Between Level of Wages and Wage Dispersion



This figure shows the correlation structure between wage level and wage dispersion predicted by the model versus the correlation observed in the data. Graphs are in logarithmic scale. Panel (a) plots the relation between average wages and wage dispersion for the March 2011 CPS. Panel (b) plots the relation between average wages and wage dispersion as predicted by the model.

When I undo the sorting, the total wage dispersion drops by 7.5% and housing price dispersion drops by 5.7%. Despite the decline in the dispersion of housing prices and wages, the overall economy experiences a fall in aggregate productivity. Total GDP falls 1.9%. This is because when we allow for sorting, most productive workers tend to cluster in larger cities. In larger cities, workers' marginal productivity is higher because of agglomeration externalities. By undoing the sorting, we are reallocating very productive workers to smaller cities. Overall, this translates into lower aggregate productivity.

I can decompose the aggregate effect further to examine events at the city level. Figure (8) shows a decomposition of how sorting changes the average wage, wage dispersion, and average housing price as a function of city characteristics (represented by theta).

Panel (a) shows the change in the average wage of a city, both allowing for and not allowing for sorting. *Superstar* cities experience an average wage increase of around 20–40% with sorting, whereas non-superstar cities oscillate between 10% and 10%. This result indicates how important sorting is to generate between-city variation. Panel (b) plots the changes in wage dispersion within cities as a function of city characteristics. Wage inequality is between 20% and 40% higher in cities with tighter housing markets due to talent sorting. Finally, Panel (c) shows a similar pattern for average housing prices in these metro areas. Talent sorting increases average housing prices by 20–40% in cities that feature a more restricted housing supply.

Another important decomposition I analyze is the result of the absence of agglomeration externalities. By setting the agglomeration externality parameter to zero, I can evaluate what would happen to an economy that features sorting in the absence of agglomeration externalities. I then examine the way in which sorting and agglomeration externalities interact to generate city-level changes. To do so, I compute the change in average wages and wage dispersion when we remove sorting in a world that features agglomeration externalities versus a world without agglomeration externalities. I find that changes in average wages between cities is, on average, 17% higher in the presence of agglomeration externalities, and I find that the changes in within-city wage dispersion is, on average, 7% higher in the presence of agglomeration externalities. Overall, these two forces seem to complement each other.

Place-Based Policies: National Housing

In this section, I explore the general equilibrium consequences of policies that take place at the city level. In this context, the model allows me to capture population responses to changes in local environments. Through the sorting of the population, these interventions can have important aggregate effects for both aggregate productivity and inequality.

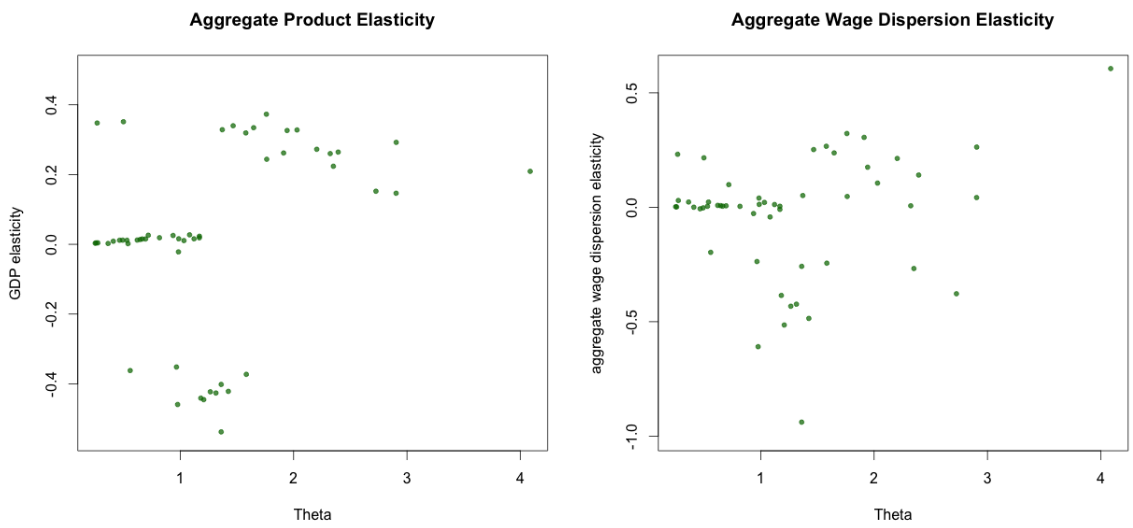
I want to evaluate the aggregate consequences of changes in the local housing supply. To do this, I independently "shock" each city with an increase in housing supply equivalent to 1% of the national housing stock. I then compute the new aggregate effects on the economy.

Figure (9) Panel (a) shows the percentage changes in aggregate product following an intervention in different types of cities. We want to evaluate what would happen if we relax the supply of houses in urban environments that feature a tighter housing market. First, I find that, when building in cities where housing constraints are not binding (e.g., cities where theta is below 1), the housing shock does not generate considerable effects. In these cities, aggregate effects are very close to zero.

FIGURE 1.8: Sorting Effect by City Type

Panel (a) predicted change in the average wage. Panel (b) predicted change in wage dispersion. Panel (c)predicted change in average housing price.

Figure 1.9: Aggregate Effects of Shocking Housing Supply by City Type



Panel (a) percentage change in average wages. Panel (b) percentage change in wage dispersion.

Second, by expanding cities where housing constraints are binding but not that important (i.e., theta is above 1 but below 1.5), aggregate productivity decreases by approximately 0.4%. Note that expanding housing does not change the number of workers in the economy, but rather the characteristics of the local housing markets. By changing these characteristics, the same workers are reallocated between cities. These re-allocations could imply higher or lower productivity as the result of the interaction between skill sorting and endogenous local agglomeration externalities. By relaxing housing constraints in these types of cities, we make the cities more attractive, thus motivating workers from larger urban centers to move to these places.

Overall, this makes the economy less productive, since these cities do not feature important agglomeration externalities. Finally, by expanding the supply of houses in *superstar* cities (where theta is above 1.5) the economy grows 0.2–0.4%. By allowing constrained cities to grow, we give more workers access to these productive environments. As these cities grow, they also become more productive. Yet, the sorting of skilled workers moderates these effects. The expansion of the housing supply mostly changes the incentives of low-skilled workers to move into these urban environments. As low-skilled workers decide to move in, they generate a negative congestion in the city they move to; meanwhile, the cities they are departing from become more appealing to all other workers. In equilibrium, this implies a transfer of high-skilled workers from big cities to smaller ones.By doing so, their aggregate productivity decreases.

Although it is the case that expanding superstar cities generates gains in aggregate productivity, this expansion also gives rise to higher wage inequality. Figure (9) Panel (b) plots the change in ex post aggregate wage dispersion after a change in housing supply in cities with different levels of housing tightness. As before, a shock to the supply of houses in places where housing constraints are not binding has no effect on aggregate inequality.

Yet, as we move from cities with thetas above one, we find that inequality is higher for cities that feature tighter housing constraints. In particular, these cities will see an influx of low-skilled workers. As the relative supply of high-skilled workers decreases in large cities, their salaries increase. At an equal rate, low-skilled workers will see their salaries depressed by the expansion of their relative talent supply. Overall, relaxing housing constraints in superstar cities generates important gains in aggregate production. But, these policy changes also entail adjustments in the composition of the local labor force. Sorting can attenuate the potential gains from expanding superstar cities and while also incurring a cost, since these talent reallocations translate into higher levels of wage inequality.

1.4. Conclusion

In this paper, I examine how city characteristics such as housing supply and local amenities affect the sorting of heterogeneous agents between cities. I developed an urban macro model in which cities have a restricted supply of houses, and workers with a continuum of skills compete for limited space through bidding wars. In this context, the pecuniary congestion costs that heterogeneous workers impose on each other operate as an endogenous driver for gentrification. The model has a unique equilibrium that can be calibrated to match different moments of the talent and wage distribution for a cross-section of US cities.

Overall, my numerical simulations stress that the sorting of heterogeneously skilled workers can generate sizable aggregate effects both on productivity and on inequality, and these results are mostly driven by the interaction between sorting and local agglomeration externalities. I examine what fraction of the wage variation in the US can be explained by talent sorting, and I find that the sorting of heterogeneous workers accounts for 7.5% of the total variation and generates considerable differences between cities. Sorting mostly affects cities that feature tighter housing markets, making them between 20% and 40% more productive as they would otherwise be in the absence of this mechanism. Although sorting is an important contributor to overall wage dispersion, it also positively contributes to aggregate productivity. If sorting is removed, aggregate production falls 1.9% as a consequence of highly skilled workers moving away from large cities (subject to high agglomeration externalities) to smaller cities.

Finally, I use the calibrated model to evaluate different types of place-based policies. I find that policies designed to improve labor mobility (e.g., the expansion of housing in constrained urban centers) can have unintended consequences, given the sorting of heterogeneously skilled workers in the presence of local agglomeration externalities. For instance, I estimate that the expanding housing supply in cities with tighter housing markets increases productivity between 0.2% and 0.4%, but this increase is mitigated by sorting, as high-skilled workers tend to relocate to cities where they are less productive. Although these policies aim to reduce inequality by facilitating the spatial mobility of workers, I show that expanding larger cities can produce an unexpected increase in aggregate wage inequality.

Chapter 2

Urban Connectivity

2.1 Introduction

Many authors have suggested that globalization and technology would mean the death of proximity, as people would no longer need to live close to each other. Yet, in an economy where the skills and knowledge of the workforce have become fundamental drivers for growth, cities and the connectivity they offer seem to be more important than ever. Acting as brokers of tacit knowledge, cities have served the role of platforms facilitating connections and allowing for the creation of new ideas. In this context, technology instead of dispersing firms and workers in space, could be working as a pull force by changing the efficiency with which firms and workers use their time and increasing the productive value of proximity.

We have witnessed an important rise in spatial disparities (what Moretti [37][38][39] called the Big Divergence). Certain cities have become very productive featuring high wages and housing prices. These same cities have also seen a higher concentration of high skilled workers and productive firms. This suggests that the value of proximity has not only increased for everyone, but it is also the case that it has change differently for different types of firms and workers.

The modern city is a very different place from what it used to be. Current mobile technologies allow people to easily coordinate, perform multiple tasks and constantly communicate and connect via a plethora of different digital platforms. Is it that we want to try a new restaurant or find a new job, everyday there seems to be a new technological solution for things to happen faster. All of these changes imply that workers can use their time more efficiently. The value of relaxing the time constraint of a worker depends on how productive that worker can be with his time. If so, the gains from proximity in a more productive environment are higher for a worker with higher skills. This poses a natural question. To what extent connectivity enhancing technologies can explain the higher sorting and selection we observe in cities?

For several reasons this is a difficult question. First of all, as firms and workers can freely move between places, we need to take into consideration the interplay between different regions using a spatial equilibrium perspective. Secondly, the urban literature has analyzed this problem either from the perspective of firms (Gaubert [16]) or from the perspective of workers (Diamond [25]). Incorporating the strategic behavior of firms and workers in a competitive environment has so far proven elusive.[1].

What drives the patterns of co-agglomeration? One possibility is that there are common features in the city (such as the supply of housing or local amenities) that both firms and workers value. This can make firms and workers coordinate in a city, although they might be making their decision independently of each other. A second possibility is that the decision of firms influences the decisions of workers and vice-versa. If there are productive complementarities between firm and workers, then their decisions might influence each other. Given the circular nature of this process, it could perfectly be the case that a small change in the relative value of proximity can trigger important changes in equilibrium. Of course both explanations might be feasible, and thus are hard to disentangle. To explore the relative importance of both ideas we need a spatial equilibrium model that can capture the strategic interaction between firms and workers, and separately account for how their sorting decisions are driven by physical features of the city and by complementarities in production.

In this paper I provide such a model. Think of an economy that produces knowledge. Firms have problems that need to be solved, and for doing so they require to be match with a workers with different levels of skills. The production function for knowledge is super modular, as high skilled workers have a comparative advantage in solving more difficult problems, which are available in more productive firms. This competitive advantage will be partly driven by the efficiency with which workers can use their time within a firm. This complementarity will be captured by a technological parameter we will refer as urban connectivity. What is crucial here is that urban connectivity is a feature of space. Different cities can provide different levels of urban connectivity.

There are going to be 2 cities that will originally differ in their supply of housing and in the quality of local amenities. There is also going to be a continuum of firms and workers that need to decide where to locate. Since knowledge is tacit, workers cannot observe a firm type nor firms can observe a worker type unless they meet and engage in face-to- face interactions. So before deciding where they want to live they do not know with what firm or with what worker they will be matched. Since firms and workers have rational expectations they can infer ex ante how the distributions of firms and workers will look in equilibrium. Once in the city they will be matched randomly, so their expected productivity in a city will depend on the equilibrium distributions of workers and firm types.

In this model, competition will play a crucial role. Since there is a limited amount of houses, workers will have to compete through bidding auctions, to be able to consume a house and enjoy the productive and consumption amenities that a city provides. Similarly, since there is a limited amount of workers (specially high skilled workers), firms will have to compete for them, again through competitive auctions. What is crucial in this setting is that by competing through auctions firms and workers generate pecuniary externalities that are dependent on firms and workers types. Secondly, given that we have a continuum

[1] An important exception is Beherens et al [18]

of worker skills and firms types, once we impose a spatial equilibrium condition, meaning that all firms and all workers need to be indifferent in expected value between the two cities, we will be able to characterize the equilibrium of this economy with a system of differential equations.

This is important for two reasons. First, under certain regularity conditions I can prove that there exists a unique equilibrium to the system of differential equations. Secondly, I will be able to compute this system numerically, which makes this type of model useful for empirical analysis. The main result of the model is that small changes in connectivity can generate important changes in the equilibrium distributions of firm and worker types between cities. This is due to the circular nature of how firms and workers decide to locate. Since workers are sensitive to firms types and firms are sensitive to workers types, small changes in time use efficiency can be dramatically amplified in equilibrium. The main objective for this model is in the near future is to quantify to what extent changes in co-agglomeration patterns can be attributed to changes in urban connectivity.

This paper relates to the literature studying what determines the sorting of heterogeneous agents (Glasear et al [40], Shapiro [12]], Bacolod [36], Eeckhout et al [21][20]), and the sorting of heterogeneous firms (Desmet and Rossi- Hansberg [17]). This paper contributes by exploring interaction between these forces, and how these relate to technological changes that have taken place in the working environments of cities.

This paper is organized as follows. Section II describes the model and the main theoretical results. Section III describes future empirical applications and identifications strategies. Finally, section IV concludes.

2.2. The Model

The model has several parts. First we will describe the production function in this economy, which requires firms and workers to meet so they can solve problems. We will then present the problem firms face when deciding where to locate. Next, we will present the housing market and workers' preferences for places to shed some light into how they make their location decisions. Finally, we describe the spatial equilibrium for this economy and present the main results.

2.2.1. Production function

This is a market for skills (as in Garicano, Fuchs and Rayo [41]). Think of a firm of type/skill z which has a problem that needs to be solved. Problems arise with a random difficulty $x \epsilon U[0, 1]$. The firm will be able to solve the problem if $x \leq z$. Otherwise, it will have to look in the labor market for a worker with the right set of skills o help them out. Workers have skill m. In case a worker has a skill level above the problem at hand, then he will be able to solve the problem.

Communicating ideas within an organization is costly. The cost of communicating ideas can be though as the efficiency with which workers use their time. Workers are constraint in the number of hours available during a working day to carry on productive activities. In this regard, the cost of communicating ideas directly relates with the efficiency with which workers can connect and coordinate with firms. Lets denote h how many units of time it takes a worker to solve a problem. We will refer to this parameter as the urban connectivity parameter. Moreover, assume that this connectivity parameter is a characteristic of space. That is, different cities might provide firms and workers different capabilities so that they can have an affluent use of their time.

Whenever a problem is solved, the firm gets $1 and 0$ when the problems is not solved.

The probability that a worker can solve a problem, conditional on the fact that the firm could not solve it is given by:

$$Pr(x < m|x > z) = \frac{m - z}{1 - z} \tag{2.1}$$

Which follows directly from the properties of the uniform distribution. Denote $n(z)$ as the number of firms that each worker can interact with in 1 unit of time.

Where,

$$n(z) = \frac{1}{h(1 - z)} \tag{2.2}$$

Once a firm is matched to a worker, the team will produce:

$$V(m, z) = \frac{m}{h(1 - z)} \tag{2.3}$$

What is crucial of this representation is that payoffs are super-modular. Which means that a higher skilled workers are more productive when matched to a more productive firms.

2.2.2. Matching in the city

Firms and workers ex-ante do not know which workers or firms they will encounter in the city. This is so, because ex-ante knowledge is tacit. Every workers looks the same from the perspective of a firm and every firm looks the same from the perspective of a worker. Firms can only observe a worker type, once they meet. Thus we will assume that before going to a city any firm can be randomly matched to interact with any worker. Therefore, firms deeply care about the statistical properties of the distribution of workers in that city. Although they cannot tell for sure who they will encounter in a city, they know that given the pool of workers they are to face, he probability of finding the talent they need will depend on the statistical properties of the talent distribution.

The same is going to be true for workers. They do not know which firms they will work with, yet ex-ante they know that the chances of finding a productive firm to work in will depend on the pool of firms in any given city. This assumption is capturing two fundamental features in the market for ideas. One is that knowledge is tacit, thus firms and workers require face-to-face interactions for truly learning each other's types. Secondly, although there is randomness in labor outcomes, firms and workers will have an accurate prior of the general characteristics of the labor market they want to be part of. When a firm chooses to locate in Silicon Valley or New York, they do not for sure whom they will end up working with, yet they have a very clear sense of the general properties of the types of workers they might encounter in different markets.

Housing prices might play an important role in screening talent for firms. Low skilled workers will not be willing to pay the high rental prices they would face in a city like San Francisco or New York, unless they know they have the required talent and productivity to make a decent living in such places. This means that conditional on observing who leaves where, firms will be able to "statistically discriminate" the quality of the workers they are to find. The same logic would applied for workers. Firms that are not very productive, will be priced out from very productive cities. Thus a worker, conditional on observing a firm's location, will be able to "statistically discriminate" or form an expectation of the type of firm they might encounter in different places..

2.2.3. Firms

In the market for ideas, firms will compete for the limited talent they will find in each city. Given that firms cannot observe the talent of a worker before meeting them, we will assume the following process:

1. Firms pick a (random) worker from the pool of workers in the city. Firms ex-ante cannot observe type, they can only correctly predict the properties of the talent distribution of workers in equilibrium. When the firm meets with a worker, they can observe their talent m.

2. If $m > z$ then the firm makes a bid for a worker.

3. Since many firms arrive to compete for a worker, they all make a bid and the highest bidder gets the worker.

3. If the firm cannot acquire talent, then gets an outside option of 0. We assume that each firm can get to bid for a worker once.

In a large labor market, the arrival rate with which firms approach a worker follows a Poisson process given by the tightness of the labor market:

$$\theta^i_{LM} = \frac{F^i}{W^i} \tag{2.4}$$

Where F^i is the (endogenous) total number of firms competing for workers in city i and W^i is the total number of workers producing in city i

When firms compete for workers, they do so following a first price auction, where firms offer workers salaries and the firm which offers the highest salary gets the worker.

The willingness to pay or bid for a worker is going to be a fraction of all the rents the firm can produce with that workers in their team. Let ϕ^L represent the fraction of total output $V(m, z)$ a firm is willing to pay for a worker. Then their optimal bid would be given by:

$$b(\phi^L V) = \phi^L V - \frac{\tilde{F}(\phi^L V)}{\tilde{f}(\phi^L V)} \tag{2.5}$$

That is, the firm offers its willingness to pay, minus an informational rent (or virtual surplus), where $\tilde{F}$ is the cumulative distribution and $\tilde{f}$ the probability distribution of the relative willingness to pay of firms.

Notice that the willingness to pay of a firm of type z is a monotonic function of its type. Thus, we can make a change of variables and re-write the above expression as:

$$b(\phi^L V) = \phi^L V - \frac{F(z)}{f(z)} \tag{2.6}$$

Where F is the cumulative distribution for firm types and f is the probability distribution for firm types.

Firms are risk neutral, and their utility is their expected income. The utility a firm would get from being matched with a worker of type m is given by:

$$U(I) = I = V - b(\phi^L V) = (1 - \phi^L)V + \frac{F(z)}{f(z)} = (1 - \phi^L)\frac{m}{h(1-z)} + \frac{F(z)}{f(z)} \tag{2.7}$$

Now the problem is that ex-ante a firm does not know with what worker is going to be matched to, nor if once is matched with a given worker, is going to be able to win the bid. This means that there are two forces that go on opposite directions. On the one hand, better firms attract better workers, which makes the given city more attractive. Yet better firms also imply tougher competition for a limited talent. This entails an interesting trade-off that will depend on each firm's type. In equilibrium, this counterbalancing forces imply that for all firm types, we should find a positive mass them everywhere.

Notice that as firms offer a fraction of the revenue they can make, then the probability of a firm winning a bid, is equal to the probability of that firm having the higher type when arriving to compete for a given worker. The probability of that event is therefore given by:

$$Pr(winning) = e^{-\theta_{LM}(1-F(z))} \tag{2.8}$$

That represents the probability of an event where all competing firms that randomly arrive to compete for a worker, all have a willingness to pay lower than mine.

The value that a firm would expect to get from moving to a city i would then be equal to:

$$E[U^i(z)] = \{(1-\phi^L)\frac{u_i^m}{h_i(1-z)} + \frac{F^i(z)}{f^i(z)}\}e^{-\theta^i_{LM}(1-F^i(z))} \tag{2.9}$$

Where $E^i(m) = u_i^m$ is the first moment of the talent distribution of workers in city i.

2.2.4. Housing Market

Firms face a limited supply of workers in a city, and so do workers face a limited supply of space (or housing in a city). Thus for workers to be able to live and produce in a city they will need to compete for limited housing. In order to provide a micro-foundation for the housing market, we follow a directed-search model as in [Albrecht, Gautier and Vroman [7]] (AGV) to portray the housing market.

Let $\theta_H^i = \frac{W^i}{S^i}$, be the tightness of the housing market in city i. Where W^i is the total number of workers that bid for houses in city i, and S^i is the total amount of houses for sale.

Houses will be homogeneous. The utility a workers gets from consuming housing comes only from the fact that it allows them to access the productive and consumption amenities a cities provides.

The housing game will have several several stages:

1. First workers (acting as interested buyers) randomly arrive to compete for a house.

2. Each worker has a private valuation x, which will be a fraction of the workers wage in city i.C 3. As workers arrive at a house they compete for it following a first-price auction (with an un-known number of competitors).

The tightness of the housing market also determines the probability of obtaining a house in the city. Since workers' valuations for a house are sampled from a common distribution, we have a Poisson race among different players who arrive following the same arrival rate θ_H^i. Thus,

$$e^{-\theta^i_H(1-\tilde{G}^i(w^i(m)))} \tag{2.10}$$

is the probability that a worker of skill m offers the highest bid, and thus wins the auction.

Where $\tilde{G}^i(w^i(m))$ is the endogenous distribution of wages in city i. Notice that θ_H^i depends on the total number of workers who arrive in equilibrium to produce in city i.

We assume that workers have rational expectations, thus they correctly anticipate the value of θ_H^i in equilibrium before moving into a city.

The expected wage of a worker of talent m in city i will be given by:

$$w^i(m) = b(\phi^L V) = \phi^L \frac{m}{h_i(1-z)} - \frac{F(z)}{f(z)} \tag{2.11}$$

We will assume that a worker's willingness to pay for a house, will be a constant fraction ϕ^H of his wage, thus his optimal bidding function will be given by:

$$b(\phi^H w^i(m)) = \phi^H w^i(m) - \frac{\tilde{G}(\phi^H w^i(m))}{\tilde{g}(\phi^H w^i(m))} = \phi^H [\phi^L \frac{m}{h_i(1-z)} - \frac{F(z)}{f(z)}] - \frac{\tilde{G}(\phi^H w^i(m))}{\tilde{g}(\phi^H w^i(m))} \tag{2.12}$$

That is, the worker offers its willingness to pay, minus an informational rent (virtual surplus), where $\tilde{G}$ is the cumulative distribution and $\tilde{g}$ the probability distribution of the relative willingness to pay of workers for houses.

Notice that the willingness to pay of a worker of type m for a house is a monotonic function of its skill level. Thus, we can make a change of variables and re-write the above expression as:

$$b(\phi^H w^i(m)) = \phi^H [\phi^L \frac{m}{h_i(1-z)} - \frac{F(z)}{f(z)}] - \frac{G(m)}{g(m)} \tag{2.13}$$

Where G is the cumulative distribution for workers' skills and g s the probability distribution for workers' skills.

2.2.5. Workers' preferences

Workers' have homothetic preferences over their disposable income and the quality of local amenities. The idea is that cities do not only offer the capacity to generate income and thus consumption, but also feature other intangible assets such as parks, good weather or a vibrant cultural environment which can also influence workers' location decisions. Homothetic preferences can be represented as $U = Ta^i log(x_i)$. For convenience, I will use the monotonic transformation $U = exp(U) = x_i e^{Ta^i}$, as all the properties of this utility representation hold for monotonic transformations. In this case, x_i is the disposable income and e^{Ta_i} is a utility shifter that depends on the amenities in that city, a^i, and on a deep preference parameter T, which captures the tradeoff between disposable income and the quality of local amenities. We assume workers consume a composite good, which is tradable and which price we normalize to 1.

Workers enjoy amenities only in the case in which they can obtain a house in the city. When they cannot obtain a house, they will receive a reservation utility $\underline{u}(m) = 0$ from living in the suburbs. In case a worker does not get a house in the city, they will still be part of the local labor force, yet they face an important dis-utility from the fact that they have to commute to consume and produce. The model does not require $\underline{u}(m) = 0$, yet this is a practical assumption for future empirical applications. Workers are risk-neutral with respect to whether they get to live in the suburbs or not.

Disposable income, is given by wages net of housing costs, $x_i = w^i(m) - b^i(w^i(m))$. Hence, the expected value that a worker of skill m would get from locating in city i is given by :

$$U(w^i(m)) = (w^i(m) - b^i(\phi^H w^i(m)))e^{Ta_i}e^{-\theta^i_H(1-G^i(m))} \tag{2.14}$$

$$U(w^i(m)) = [(1-\phi^H)(\phi^L \frac{m}{h_i(1-z)} - \frac{F(z)}{f(z)}) + \frac{G(m)}{g(m)}]e^{Ta_i}e^{-\theta^i_H(1-G^i(m))} \tag{2.15}$$

Before moving into a city, a worker does not know with which firm it will be matched. Since workers and firms are matched randomly in the city, the expected value of producing there is going to be given by:

$$E[U^i(m)] = [(1-\phi^H)(\phi^L \frac{m}{h_i} E_z(\frac{1}{1-z}) - E_z(\frac{F(z)}{f(z)})) + \frac{G(m)}{g(m)}]e^{Ta_i}e^{-\theta^i_H(1-G^i(m))} \tag{2.16}$$

Where $E^i(m) = u^m_i$ is the first moment of the talent distribution of workers in city i.

2.2.6. Spatial equilibrium for 2 cities

Firs of all, for ease of exposition, we will assume that there are only 2 cities in this economy. The model can easily be extended to an arbitrarily large number of cities. Given that we already found the expected value for firms and workers of locating in different cities, we will look for an spatial equilibrium for this economy. Given the fact that firms and workers can freely move between cities, it has to be the case that in equilibrium all workers and all firms need to be indifferent between locating in city 1 or locating in city 2. This implies that the utility of firms and workers must be equal across cities.

We focus on a closed economy with a fix supply of firms and workers. Let $F(z)$ be the Cumulative Distribution Function (CDF) for all the firms in the economy, and let $f(z)$ be its Probability Distribution Function (PDF). Similarly, let $G(m)$ be the Cumulative Distribution Function (CDF) for all workers in the economy, and let $g(m)$ be its Probability Distribution Function (PDF).

We need to solve a "functional" problem, that is we need to find a partition of $F(z)$, $F^1(z)$ and $F^2(z)$ such that $F^1(z) + F^2(z) = F(z)$. Similarly, we need to find a partition of $G(m)$, $G^1(m)$ and $G^2(m)$ such that $G^1(m) + G^2(m) = G(m)$

The primitives of the problem will be the original skill distributions and the characteristics of each city. Cities will feature a fix supply of housing and an amenity level. Cities will also differ in their productive capabilities which we call urban connectivity. The efficiency with which workers can make use of their working time, will be a feature of space and will be captured by the technological parameter h_i which we refer as urban connectivity in city i.

The spatial equilibrium condition for firms will be given by:

$$[(1-\phi^L)\frac{u_1^m}{h_1(1-z)}+\frac{F^1(z)}{f^1(z)}]e^{-\theta^1_{LM}(1-F^1(z))}=[(1-\phi^L)\frac{u_2^m}{h_2(1-z)}+\frac{F^2(z)}{f^2(z)}]e^{-\theta^2_{LM}(1-F^2(z))} \tag{2.17}$$

$$[(1-\phi^L)\frac{u_1^m}{h_1(1-z)}+\frac{F^1(z)}{f^1(z)}]e^{\theta^2_{LM}(1-F^2(z))-\theta^1_{LM}(1-F^1(z))}=[(1-\phi^L)\frac{u_2^m}{h_2(1-z)}+\frac{F^2(z)}{f^2(z)}] \tag{2.18}$$

and since $F^1(z)+F^2(z)=F(z)$ and $f^1(z)+f^2(z)=f(z)$,

$$[(1-\phi^L)\frac{u_1^m}{h_1(1-z)}+\frac{F^1(z)}{f^1(z)}]e^{\theta^2_{LM}(1-F(z)+F^1(z))-\theta^1_{LM}(1-F^1(z))}=[(1-\phi^L)\frac{u_2^m}{h_2(1-z)}+\frac{F(z)-F^1(z)}{f(z)-f^1(z)}] \tag{2.19}$$

Note two things. First of all, the above expression is a function of two un-knowns, a function F^1 and its derivative f^1 . If we re-arrange all the terms of the above expression and write:

$$f^1=H(F^1,\gamma) \tag{2.20}$$

we can characterize the functional problem as a differential equation. This differential equation will depend on the moments of the distributions of workers and a set of parameters γ.

$$\frac{(1-\phi^L)}{1-z}[\frac{u_1^m}{h_1}e^{\theta^2_{LM}-\theta^1_{LM}+F^1(z)(\theta^2_{LM}+\theta^1_{LM})-\theta^2_{LM}F(z)}-\frac{u_2^m}{h_2}]= \tag{2.21}$$

$$\frac{F(z)-F^1(z)}{f(z)-f^1(z)}-\frac{F^1(z)}{f^1(z)}e^{\theta^2_{LM}-\theta^1_{LM}+F^1(z)(\theta^2_{LM}+\theta^1_{LM})-\theta^2_{LM}F(z)} \tag{2.22}$$

From the expression above, we can get a quadratic expression for f^1.

In order to solve the above expressions lets call

$A=\frac{(1-\phi^L)}{1-z}[\frac{u_1^m}{h_1}e^{\theta^2_{LM}-\theta^1_{LM}+F^1(z)(\theta^2_{LM}+\theta^1_{LM})-\theta^2_{LM}F(z)}-\frac{u_2^m}{h_2}]$

$a=F(z)-F^1(z)+F^1(z)e^{\theta^2_{LM}-\theta^1_{LM}+F^1(z)(\theta^2_{LM}+\theta^1_{LM})-\theta^2_{LM}F(z)}$

and

$b = f(z)F^1(z)e^{\theta^2_{LM} - \theta^1_{LM} + F^1(z)(\theta^2_{LM} + \theta^1_{LM}) - \theta^2_{LM}F(z)}$

Then

$$f^1(z) = \frac{-Af(z) + a + \sqrt{(Af(z) - a)^2 + 4Ab}}{2A} \tag{2.23}$$

Similarly, as all workers need to be indifferent between locating in city 1 or city 2, the spatial equilibrium condition for workers will be given by:

$$[(1-\phi^H)(\phi^L \frac{m}{h_1} E_{z1}(\frac{1}{1-z}) - E_{z1}(\frac{F(z)}{f(z)})) + \frac{G^1(m)}{g^1(m)}]e^{Ta_1}e^{-\theta^1_H(1-G^1(m))} = \tag{2.24}$$

$$[(1-\phi^H)(\phi^L \frac{m}{h_2} E_{z2}(\frac{1}{1-z}) - E_{z2}(\frac{F(z)}{f(z)})) + \frac{G^2(m)}{g^2(m)}]e^{Ta_2}e^{-\theta^2_H(1-G^2(m))} \tag{2.25}$$

$$[(1-\phi^H)(\phi^L \frac{m}{h_1} E_{z1}(\frac{1}{1-z}) - E_{z1}(\frac{F(z)}{f(z)})) + \frac{G^1(m)}{g^1(m)}]e^{(a_1-a_2)T + \theta^2_H - \theta^1_H + G^1(m)(\theta^2_H + \theta^1_H) - \theta^2_H G(m)} = \tag{2.26}$$

$$[(1-\phi^H)(\phi^L \frac{m}{h_2} E_{z2}(\frac{1}{1-z}) - E_{z2}(\frac{F(z)}{f(z)})) + \frac{G^2(m)}{g^2(m)}] \tag{2.27}$$

Since we have a closed economy it also has to be the case that $G^1(m) + G^2(m) = G(m)$ and $g^1(m) + g^2(m) = g(m)$, thus

$$(1-\phi^H)[[(\phi^L \frac{m}{h_1} E_{z1}(\frac{1}{1-z}) - E_{z1}(\frac{F(z)}{f(z)})) + \frac{G^1(m)}{g^1(m)}]e^{(a_1-a_2)T + \theta^2_H - \theta^1_H + G^1(m)(\theta^2_H + \theta^1_H) - \theta^2_H G(m)} - (\phi^L \frac{m}{h_2} E_{z2}(\frac{1}{1-z}) - E_{z2}(\frac{F(z)}{f(z)}))] \tag{2.28}$$

$$= \frac{G(z) - G^1(z)}{g(z) - g^1(z)} - \frac{G^1(m)}{g^1(m)}e^{(a_1-a_2)T + \theta^2_H - \theta^1_H + G^1(m)(\theta^2_H + \theta^1_H) - \theta^2_H G(m)} \tag{2.29}$$

Once again, we need to re-arrange all terms so that we can write $g^1 = \tilde{H}(G^1, \tilde{\gamma})$, so that we can characterize the functional problem as a differential equation.

Note that again this differential equation will depend on the moments of the distributions of firms and a set of parameters $\tilde{\gamma}$.

From the expression above, we can get a quadratic expression for g^1.

Lets denote

$\tilde{A} = (1-\phi^H)[[(\phi^L \frac{m}{h_1} E_{z1}(\frac{1}{1-z}) - E_{z1}(\frac{F(z)}{f(z)})) + \frac{G^1(m)}{g^1(m)}]e^{(a_1-a_2)T + \theta^2_H - \theta^1_H + G^1(m)(\theta^2_H + \theta^1_H) - \theta^2_H G(m)} - (\phi^L \frac{m}{h_2} E_{z2}(\frac{1}{1-z}) - E_{z2}(\frac{F(z)}{f(z)}))]$

$\tilde{a} = G(m) - G^1(m) + G^1(m)e^{(a_1-a_2)T + \theta^2_H - \theta^1_H + G^1(m)(\theta^2_H + \theta^1_H) - \theta^2_H G(m)}$

and

$\tilde{b} = g(m)G^1(m)e^{(a_1-a_2)T+\theta_H^2-\theta_H^1+G^1(m)(\theta_H^2+\theta_H^1)-\theta_H^2 G(m)}$

Then

$$g^1(m) = \frac{-\tilde{A}g(m) + \tilde{a} + \sqrt{(\tilde{A}g(m) - \tilde{a})^2 + 4\tilde{A}\tilde{b}}}{2\tilde{A}} \tag{2.30}$$

Finally, note that the differential equations are interdependent, as the mean of the skill distribution of workers is an argument for the equilibrium condition of firms, and the mean of the skill distributions of firms is an argument for the equilibrium condition of workers. This inter-dependency can generate non-linear response to changes in parameter values as we will show.

2.2.7. Results

We will present mainly two theoretical results. First we will show that this economy features the existence of an equilibrium which can be unique. This implies that quantitative simulations of the model can be useful for empirical applications. The second result is to show that small changes in urban connectivity, that is the relative efficiency with which workers can use their working time, can generate important changes in the concentration of high skilled workers and productive firms in those cities.

Proposition 2.1. *There exits and equilibrium to this economy, which under certain conditions can be unique*

Proof: Note that the equilibrium of this economy is completely characterized by a system of differential equations for $< f^1(z), g^1(m) >$. As long as $f^1 = H(F^1, \gamma)$ and $g^1 = \tilde{H}(G^1, \tilde{\gamma}$ are continuous and have bounded derivatives, then there exits a unique equilibrium the system. This will be the case, whenever $F(z)$ and $G(m)$ are continuous functions, with positive density in their entire support.QED

Proposition 2.2. *Small changes in urban connectivity in one city, can generate important effects on the system overall.*

Proof: Lets first write the system of differential equations.

$$\begin{aligned} g^1(m) &= \tilde{H}(G^1, f^1, \tilde{\gamma}) \\ f^1(z) &= H(F^1, g^1, \gamma) \end{aligned} \tag{2.31}$$

Now evaluate how a change the the urban connectivity of city 1 would affect the system of equations

$$\begin{aligned} \frac{\partial g^1}{\partial h_1} &= \frac{\partial \tilde{H}(G^1,f^1,\tilde{\gamma})}{\partial G^1}\frac{\partial G^1}{\partial h_1} + \frac{\partial \tilde{H}(G^1,f^1,\tilde{\gamma})}{\partial f^1}\frac{\partial f^1}{\partial h_1} \\ \frac{\partial f^1}{\partial h_1} &= \frac{\partial H(F^1,g^1,\gamma)}{\partial F^1}\frac{\partial F^1}{\partial h_1} + \frac{\partial H(F^1,g^1,\gamma)}{\partial g^1}\frac{\partial g^1}{\partial h_1} \end{aligned} \tag{2.32}$$

re-arranging terms we get,

$\frac{\partial g^1}{\partial h_1} = \frac{1}{(1-\frac{\partial H}{\partial g^1}\frac{\partial \tilde{H}}{\partial f^1})}[\frac{\partial \tilde{H}}{\partial G^1}\frac{\partial G^1}{\partial h_1} + \frac{\partial \tilde{H}}{\partial f^1}\frac{\partial H}{\partial F^1}\frac{\partial F^1}{\partial h_1}]$QED

What is critical here to note, is that small changes in the urban connectivity of a city, can generate dramatical shifts in the system. This effect is twofold. First it implies that the affected city will grow. What is more important though, is that that the growth will come prominently from high skilled workers and more productive firms, generating a high level of segregation and spatial inequality.

2.3. Conclusion

We study how urban connectivity can affect the spatial segregation we observe in workers' skills and firm productivities. We focus in an economy that produces knowledge, and to do so needs to match heterogenous firms and workers. We provide a spatial equilibrium model that has the unique feature that allows for the sorting of a continuum of firms and workers where productive complementarities are city specific characteristics. The model, under certain conditions, has a unique equilibrium that can be used for empirical applications. The main theoretical result, is that small changes in the connectivity of a city, can generate non-linear changes in city sizes and the level of skill segregation between cities.

This model formalizes a under-explore channel, that can explain why cities are moving further apart in the skill composition of their workers and in the productive differences between their firms. This suggest that small changes in the productive environment of a city, which could be driven by technology, could account for big changes in equilibrium distributions.

In future versions of this paper, we hope to empirically explore these issues, and quantify the role that relative changes in urban connectivity between cities can account for the divergent outcomes we have have recently observed.

Chapter 3

Clustering to Coordinate: Who Benefits from Knowledge Spillovers? (with William Grieser and Gonzalo Maturana)

3.1. Introduction

How do firms benefit from agglomeration, and why do some industries cluster more than others? Extant research provides compelling evidence that industrial activity is spatially concentrated and that such *agglomeration* generates gains in firm and worker productivity. Since the work of [42], part of this effect is often attributed to the notion that geographic concentration facilitates the spread of tacit knowledge. However, few theories provide guidance to empirically test when this mechanism is more prevalent. Furthermore, while productivity gains are generated through the actions of firms, the precise relationship between the drivers of agglomeration and corporate decisions has not been widely explored. In this paper, we develop a specific mechanism through which knowledge spillovers improve corporate decisions, and we provide empirical evidence consistent with the proposition that this mechanism is an important driver of agglomeration.

We start with a framework in which investment decisions can be conceptualized as games of incomplete information in which payouts depend on the actions of related firms. When firms hold heterogeneous and incomplete information regarding these investment payouts, the sharing of private information can reduce project uncertainty. This feedback generates the incentive for related firms to coordinate investment decisions. However, there are costs to coordinating, since knowledge is scattered and difficult to convey. In this context, dense urban environments can be considered as a technology that reduces coordination costs, thereby facilitating the transfer of knowledge and improving project selection.

In equilibrium, our model predicts that the investment benefits that arise from the spatial concentration of firms are larger in relatively more uncertain industries that invest in relatively more complex projects. In particular, coordination costs arise endogenously as a byproduct of informational frictions. Firms that invest in relatively more complex projects (i.e., projects that require the coordination of many complementary assets) and operate in relatively more uncertain environments face the highest coordination costs. While feedback between firms' actions makes it difficult to solve for equilibrium outcomes, the global games framework developed by [43] and [44] offers a useful approach for circumventing these difficulties. Furthermore, the global games framework allows us to capture the miscoordination that arises from the formation of higher-order beliefs between firms, which may induce inefficient investment.[1] In the context of our model, the ability to share private knowledge translates into firms gaining precision, which mitigates the miscoordination problem and, in turn, improves investment decisions. Thus, using the global games framework, we provide a theoretical explanation motivated through firm behavior to explain why certain industries exhibit stronger degrees of clustering than others.[2]

Our model predicts that we should observe a greater degree of clustering for relatively uncertain (i.e., riskier) and relatively complex industries. It is challenging to test this prediction empirically, since the benefits of coordination are difficult to measure. However, a firm's location decision is informative of these benefits through revealed preferences. An additional challenge is that, even in the absence of agglomeration externalities, we should not expect firms to locate randomly. Specifically, spatial concentration depends on the size of a given industry and the general concentration of the population at large, which may be determined by a variety of factors, such as local amenities. Thus, a well-designed test of clustering (i.e., localization) should control for these confounding factors. We conduct tests of localization based on continuous density estimates, as introduced by [46], which account for these issues.[3]

We develop proxies for industry risk and complexity by combining two industry-level metrics: (1) stock return volatility and (2) a measure of worker skill.[4] We collect geographic coordinates for the ZIP codes of the corporate headquarters for more than 9,000 firms in 24 industries, and we find a strong positive relationship between industry clustering and industry risk/complexity at close distances (between 0 and 20 miles). Moreover, we do not find a relationship between headquarters location and industry risk/complexity at longer distances, consistent with the predictions of our model.

Another possible concern with our empirical test is that we rely on corporate activity being concentrated in headquarters locations. While headquarters location provides a first approximation for the locality of a firm's activity, a firm may also conduct some

[1] [44] explain higher-order beliefs as "... players' beliefs about other players' beliefs, players' beliefs about other players' beliefs about other players' beliefs, and so on."

[2] [45] provide a recent example of the global games framework applied to financial economics.

[3] Duranton and Overman develop a test of localization based on kernel density estimations of bilateral distances between establishments in an industry. To control for industry size and population concentration, they construct counterfactuals by generating "pseudo-industries" through randomly sampling firms from the full set of possible locations according to industrial organization and general population conditions.

[4] The intuition behind this choice is that industries that invest in relatively more complex projects require workers that are better trained and better educated than firms that serve less complex markets. We use a ranking of occupations from the U.S. Department of Labor.

of its operations elsewhere. To mitigate this concern, we conduct localization tests as described above using inventor patent locations, and we find similar results.

Our empirical findings are consistent with the underlying mechanism of our model (i.e., knowledge spillovers) affecting industry clustering through firm investment decisions. Nevertheless, the results are also consistent with other possible mechanisms. For example, thicker labor markets can facilitate the reemployment of workers in riskier industries in which the probability of being laid off is higher. To provide further reassurance that knowledge spillovers are an important driver of heterogeneity in industry clustering, and to further investigate its relationship with corporate decisions, we examine a different dimension of corporate decision making, which should be especially prevalent to our knowledge-sharing story. Specifically, we focus on the abnormal clustering of R&D expenses across industries. We find that R&D in riskier/complex industries exhibits a greater degree of clustering than headquarters locations, consistent with a knowledge channel driving industry clustering through facilitating project implementation. While our evidence increases our confidence that knowledge spillovers are an important component of firm decision making, we acknowledge that we cannot address all labor market stories, and we acknowledge that multiple mechanisms could simultaneously be at play.[5]

Finally, we study a setting in which the predictions of our model should be particularly prevalent: Customer–supplier proximity. We analyze a sample of more than 2,300 customer–supplier pairs in which the customer accounts for at least 10% of the supplier's total sales. Thus, in this sample, supplier investments depend heavily on the investments of customers. We find that suppliers are 7.9 percentage points more likely to locate within 20 miles of customers in industries with the highest risk/complexity relative to customers in industries with the lowest risk/complexity. This 7.9 percentage point increase is economically large, representing 89.3% of the sample average propensity for suppliers to locate near customers. Moreover, when suppliers sell to customers in industries with the highest risk/complexity, their headquarters tend to be 150 miles closer, on average, to the customer when compared to suppliers that sell to customers in industries with the lowest risk/complexity.

Our paper contributes to the literature on the benefits of knowledge spillovers.[6] Specifically, we add to research that provides the microfoundations of knowledge diffusion. [48], [49], and [13] show that localization alleviates the costs of exchanging ideas.[7] [51] and [52] focus on human capital externalities through learning. We use the insights from the global games literature to show that the sharing of knowledge may facilitate coordination among firms and workers. The literature also provides suggestive empirical evidence that knowledge spillovers play an important role in the clustering of certain industries.[8] We contribute to the empirical literature by providing evidence consistent with knowledge spillovers leading to different degrees of clustering according to theoretically motivated

[5]For example, higher-ability workers may benefit more from larger cities.

[6]See [47] for a survey of the theory.

[7]Similarly, [50] focus on the consequences of costly idea exchange in a system of cities.

[8]For example, [53] and Buzard et al. [54] show that innovation is more spatially concentrated than manufacturing. [55] show a rapid spatial decay in the benefits of knowledge spillovers in the advertising industry.

industry characteristics (i.e., risk and complexity). We find corroborating evidence in three distinct samples.

Additionally, we contribute to the literature on the determinants of agglomeration externalities. [42] classifies three main forces that could drive agglomeration: labor market pooling, input sharing, and knowledge spillovers. A number of studies have established correlations between agglomeration measures and industry characteristics in an attempt to uncover the underlying economic mechanisms that drive agglomeration.[9] However, the literature is just beginning to highlight the relative influence of the three agglomeration forces (see Ellison, Glaeser, and Kerr 57, Rosenthal and Strange 59). While it is not our goal to disentangle the relative importance of these agglomeration mechanisms, and while we recognize that multiple forces are probably at play, we provide insights regarding the specific mechanisms through which knowledge is transferred between firms.

Our paper also relates to the literature on corporate investment. By selecting projects with the highest net present value, corporate investment policy facilitates one of the most important functions of financial markets: the allocation of capital to its highest-valued use. However, valuing projects with uncertain cash flows can be difficult when information is noisy or signals are imprecise (e.g., [60]). Our model suggests that one plausible channel for firms to reduce uncertainty, and thereby improve project selection, is to cluster with related firms in order to facilitate communication and the sharing of private information.[10]

Finally, as we provide a rationale for the value of knowledge spillovers, this paper relates to the literature in urban economics that studies why workers are more productive in cities (e.g., Moretti 39, Glaeser and Maré 40). More broadly, our work also relates to the classical literature that links knowledge externalities with economic development (e.g., Lucas 62, Romer 63) and, more specifically, to more recent work on the relation of agglomeration and growth (e.g., Rossi-Hansberg and Wright 64, Davis, Fisher, and Whited 65).

The rest of the paper is organized as follows: Section 3.2 describes the model, along with its main result. Section 3.3 describes the framework, data, and measures used in the empirical tests. Section 3.4 presents the empirical results. Section 3.5 concludes.

3.2. The Model

The transmission of knowledge plays a growing role in an economy in which services and technology are becoming increasingly important. Face-to-face interaction greatly facilitates the transfer of knowledge, which is embedded in workers and is difficult to transfer ([52]). Thus, although information can flow more freely than ever, distance continues to

[9] For example, [56] show that innovation activity is more concentrated than industrial activity. Likewise, [57] and [58] find substantial heterogeneity in patterns of industry agglomeration using establishment data from the U.S. and the U.K., respectively.

[10] This is consistent with the recent empirical findings in [61], who suggest that corporate investment is highly sensitive to the investment of local firms, after controlling for industry conditions.

play a pivotal role in the transfer of knowledge: Firms and cities seek to minimize the transaction costs of acquiring knowledge by coordinating specialized workers, by allowing them to communicate, and by creating an environment for them to learn from each other. Next, we introduce a model that captures these features.

3.2.1. Basic Framework

In this section, we introduce our model, which builds on the global games framework of [44].[11] Consider a continuum of firms that face a binary investment decision: To invest (I) or not to invest (NI). We assume that actions are complementary, consistent with the idea that complex goods and services require the combining of dispersed knowledge, and the idea that knowledge is non-excludable by nature. That is, the profitability of an investment increases with the investment of other firms. More specifically, the payoff of the investment opportunity is given by

$$U = \begin{cases} \theta + l - 1 & if \quad I \\ 0 & if \quad NI, \end{cases} \tag{3.1}$$

where θ is a random variable that represents the realization of uncertain investment and l is the number of other firms that also decide to invest. Without loss of generality, we normalize non-investment payoffs to zero.

The informational environment is characterized by a publicly observed signal (e.g., information) and a privately observed signal (e.g., tacit knowledge). All firms observe an unbiased, but noisy, signal: $y_i = \theta + \varepsilon_i$, where $\varepsilon_i \backsim N(0, \tau^2)$ and τ relate the fundamental uncertainty of the industry. Additionally, each firm independently observes a private signal $x_i = \theta + \nu_i$, where $\nu_i \backsim N(0, \sigma^2)$ and σ relate to the precision of private assessments. By construction, all noise terms are independently distributed, and therefore they are uncorrelated.[12]

As demonstrated by Morris and Shin, the symmetric equilibrium of these games are fully characterized by a switching strategy in which firms *invest* whenever the expected value of the realization of the outcome θ is higher than some threshold κ, and choose to *not invest* otherwise. In this setting, the utility representation has the important feature that, since payoffs are linear and all signals are independent from one other, studying a two-player game is equivalent to studying the continuous and more general case of a continuum of firms.[13] Thus, for ease of exposition, and without loss of generality, we focus

[11] This framework has found useful applications in different areas of economics and finance where there are informational frictions in coordination. This includes applications in the context of Currency Crises, Bank Runs, and Debt Runs (see [66]).

[12] Thus, any correlation that might arise in equilibrium will not stem from the informational structure but from endogenous strategic considerations.

[13] Consider a firm that observes a signal x and thinks that all other firms will follow the same switching strategy with threshold κ. Because the realization of all signals are independent conditional on θ, the firm's expectation of the proportion of other firms that will infer an expected value of θ lower than κ will be equal to the probability that the firm assigns to any other firm that observes a signal with a value lower than κ.

on characterizing the equilibrium of the two-player game. The normal form representation of the game is given by the following payoff matrix:

	I	NI
I	θ,θ	$\theta-1$,0
NI	0,$\theta-1$	0,0

(3.2)

Given the observed signals y and x and the properties of the normal distribution, each firm's expectation of θ is

$$\bar{\theta} = \frac{\sigma^2 y + \tau^2 x}{\sigma^2 + \tau^2}, \tag{3.3}$$

which is the average of both signals, weighted by their noise-to-signal ratio. Similarly, the standard deviation of θ is

$$\hat{\sigma} = \sqrt{\frac{\sigma^2\tau^2}{\sigma^2 + \tau^2}}. \tag{3.4}$$

As mentioned above, each firm will follow a switching strategy $s()$, which is a function of its posterior:

$$s(\bar{\theta}) = \begin{cases} I & if \quad \bar{\theta} > \kappa \\ NI & if \quad \bar{\theta} \leq \kappa. \end{cases} \tag{3.5}$$

In the symmetric equilibrium of this game, firms' strategies depend on their beliefs about other firms' strategies. Therefore, each firm must anticipate other firms' private signals.

Lemma 3.1. *The equilibrium of the game is given by the implicit equation*

$$\kappa + -\Phi\left\{\gamma(\kappa - y)\right\} = 0, \tag{3.6}$$

where $\gamma \equiv \frac{(\sigma^2/\tau^2)}{\sqrt{\frac{2\sigma^2\tau^2+\sigma^4}{\sigma^2+\tau^2}}}$.

Proof. See Appendix A. □

The value of γ is such that $\gamma \leq 2\pi$.[14]

[14] This assumption guarantees that–following an Iterated Elimination of Strictly Dominated Strategies–a unique equilibrium exists (see Morris and Shin 44 for the proof). We revisit this assumption below.

3.2.2. Information Inefficiency

If firms are using information efficiently, then actions and beliefs should adjust to changes in public signals on a one-to-one basis. Any difference between actions and beliefs is due to the strategic effects induced by high-order beliefs about the public signal. To characterize this strategic miscoordination, we define the variable *Informational Inefficiency* as the ratio between the equilibrium action and private beliefs. We are interested in how this ratio adjusts to changes in public information. Let $I(y)$ be the *Informational Inefficiency*:[15]

$$I(y) = \kappa/\bar{x} - 1. \tag{3.7}$$

Note that *Informational Inefficiency* captures the cost of coordinating actions when knowledge is dispersed, which arises from the incompleteness of the informational environment. Therefore, the degree to which the inefficiency adjusts to changes in the public signal is given by

$$dI(y) = \frac{\partial \kappa}{\partial y} / \frac{\partial \bar{x}}{\partial y}, \tag{3.8}$$

where

$$\frac{\partial \kappa}{\partial y} = -\frac{\frac{\sigma^2}{\tau^2} + \sqrt{\gamma}\phi()}{1 - \sqrt{\gamma}\phi()}$$

and

$$\frac{\partial \bar{x}}{\partial y} = -\frac{\tau^2}{\sigma^2}.$$

Thus,

$$dI(y) = \frac{1 + \frac{\tau^2}{\sigma^2}\sqrt{\gamma}\phi()}{1 - \sqrt{\gamma}\phi()}. \tag{3.9}$$

Equation 3.9 yields two important results. First, the coordination cost (the function $dI(y)$) reaches its maximum when the observed signal is 1/2. It is when the informational content of signals is low that the coordination cost is higher. Second, the coordination cost is increasing in τ. That is, the higher the level of prior uncertainty, the higher the coordination failure.

3.2.3. The Value of Proximity

One commonly proposed mechanism behind the benefits of agglomeration is based on the notion that ideas "flow in the air" (e.g., Lucas 62). However, formalizing this notion

[15] Note that when actions and beliefs coincide, the inefficiency is zero by construction.

in a theoretical and empirical framework has proved challenging (Duranton and Puga 47). In our theoretical framework, *knowledge spillovers* can be thought as firms sharing their private signals.[16] This sharing provides firms with a sufficient statistic to guide their decision-making process. As this unbiased communication takes place, the sufficient statistic will follow a sampling distribution of private signals. As firms share their private assessments with n other discrete unbiased firms, the sample of private signals $\bar{x}$ will be given by[17]

$$\bar{x} \sim N(\theta, \sigma/\sqrt{n}). \tag{3.10}$$

High-density regions facilitate the sharing of private signals. In particular, the signal becomes more precise as the number of clustering firms increases. In the limit,[18]

$$\hat{\sigma} = \sigma/\sqrt{n} \rightarrow_{n \rightarrow \infty} 0. \tag{3.11}$$

Proposition 3.2. *In sufficiently dense environments, knowledge spillovers generate gains from coordination that are higher for more uncertain industries.*

Proof. See Appendix A. □

Everything else constant, as the precision of the private signal increases, each firm relies less on the public signal. The strategic miscoordination induced by the public signal gradually disappears as it becomes irrelevant.

Proposition 3.2 has implications for empirical analysis. As firms in uncertain or riskier industries benefit to a greater extent from knowledge spillovers, a higher degree of spatial concentration should be observed in these industries. Moreover, given the initial assumption of complementary investments, the model implicitly relates to the production of complex goods or services that require the coordination of a high number of specialized firms. Importantly, these industry features are fundamental characteristics of industries that arise endogenously from the model.

Next, we provide empirical evidence consistent with the implications above. First, we start by describing the empirical framework, our data, and our constructed measures. Then, we present the empirical results.

[16] This, in turn, can be thought as a transfer of knowledge. Proximity only affects the assessment of private signals; it does not affect the assessment of public signals (which consist of more general information).

[17] Note that since we have a continuum of players, each discrete and independent signal that gets added to the sample does not change the overall beliefs about aggregate distributions.

[18] Note that as $\sigma \rightarrow 0$, $\gamma \rightarrow 0$. This limit is consistent with the region defined by Assumption 3.2.1.

3.3. Empirical Framework and Data

3.3.1. Kernel Density Estimations and Industry Localization

As a first step, we construct a measure of agglomeration at the industry level. We use the methodology in [46, hereafter DO], who develop a test of localization based on kernel density estimations of bilateral distances between establishments in an industry. More specifically, they estimate the following function for each industry A:

$$\hat{K}_A(d) = \frac{1}{n_A(n_A-1)h} \sum_{i=1}^{n_A-1} \sum_{j=1}^{n_A} f\left(\frac{d-d_{i,j}}{h}\right), \tag{3.12}$$

where $d_{i,j}$ is the Euclidean distance between the locations of establishments i and j in industry A. The number of establishments in an industry is denoted by n_A. The function f is a Gaussian kernel density with bandwidth h. Note that equation 3.12 considers $\frac{n_A(n_A-1)}{2}$ unique bilateral distances and generates a density distribution for all potential distances. In particular, industries with a high degree of agglomeration will tend to have high values of $\hat{K}_A(d)$ at lower distances.

Although the kernel density provides useful information about the distribution of the different localities in an industry, it does not provide the full picture. Even if a value of $\hat{K}_A(d)$ at a given distance in a given industry appears to be high, it cannot be concluded that the value is *abnormally* high without comparing it to the appropriate counterfactual. In particular, the comparison with other industries may not be informative, as spatial concentration depends on both the size and the concentration of the industries as well as the general population density. To address this issue, DO construct counterfactuals by generating 1,000 pseudo-industries of equivalent size as the industry of interest by random sampling from the full set of possible locations. From these simulations, DO construct confidence intervals for each industry and distance. In particular, let $\bar{K}_A(d)$ be the upper limit for the 95% confidence interval. DO define the following index of localization:

$$\gamma_A(d) \equiv max\left(\hat{K}_A(d) - \bar{K}_A(d), 0\right). \tag{3.13}$$

A positive value of $\gamma_A(d)$ (i.e., when the kernel density exceeds the upper bound of the 95% confidence interval) indicates a departure from randomness, subject to stylized industry concentration and overall population characteristics. Therefore, industry A exhibits localization at distance d.[19]

[19] DO also define an index of dispersion using the 95% confidence interval lower limit. However, since we focus on localization and not on dispersion, we refrain from this calculation.

3.3.2. Data and Measures

To test the implications of our model, we draw on a number of data sources. In this section, we discuss each data source, along with the construction of our empirical measures.

3.3.2.1. Firms and Locations

Information on firm headquarters locations, industry characteristics, and financials come from Compustat. To calculate our kernel density estimates, we collect geographic coordinates from the ZIP code for each firm's headquarters from 2000 to 2012.[20] We restrict the sample of firms to those firms headquartered in the contiguous United States. The sample includes a total of 9,167 firms.[21] Extant research has indicated the significance of headquarters location, which provides a useful first approximation for the location of a firm's activities (e.g., Dougal, Parsons, and Titman 61, Pirinsky and Wang 68).[22]

3.3.2.2. Industry Risk

We conduct our analysis at the Fama and French 48 industry classification level. We use price stock volatility as a proxy for industry-level risk. To construct this measure, we use data from the Center for Research in Security Prices' (CRSP) Monthly Stock File. This file contains monthly stock-price information for more than 30,000 U.S. publicly traded firms. For each industry classification, we construct a series of value-weighted monthly returns from 2000 to 2012. Then, we compute industry volatility as the standard deviation of each series of returns.

3.3.2.3. Industry Complexity

We use a measure of worker skill as a proxy for industry-level complexity. The intuition behind this choice is that industries that focus on more complex products and markets require workers that are better trained and educated than firms that serve less complex markets. We use data from the Occupational Information Network (O*NET), a website that contains detailed information provided by the U.S. Department of Labor in a survey of randomly sampled U.S. workers for each occupation. O*NET classifies each occupation into one of five skill categories according to the degree of preparation needed, stating that "[e]very occupation requires a different mix of knowledge, skills, and abilities, and is performed using a variety of activities and tasks."

[20] We start in 2000 and not earlier because the productive structure of the U.S. economy has recently undergone important shifts (Herrendorf et al. 67). This "Structural Transformation" can be understood as the reallocation of economic activity from agriculture to manufacturing and, recently, to knowledge services. Thus, in recent years, the forces in our model should be more salient.

[21] Descriptive statistics for these firms are presented in Appendix B.

[22] However, a portion of a firm's operations may take place outside of its headquarters location. To mitigate this concern, in Section 3.4.3.4.3, we conduct a robustness analysis using corporate inventor locations, which are not confined to a firm's headquarters.

The skill level of occupations range from“little or no preparation needed” (Job Zone 1) to “extensive preparation” (Job Zone 5). Job Zone 1 includes occupations that may require a high school diploma or GED, little or no previous work-related skill required, and a few days to a few months of on-the-job training.[23] Job zone 5 includes occupations that typically require a master’s degree, Ph.D., M.D., or J.D.; extensive skill, knowledge, and experience; but typically little on-the-job training, because most occupations in this category assume that the worker already possesses the necessary skills and knowledge.[24]

To aggregate the O*NET skill measures to the industry level, we use National Industry-Specific Employment and Wage Estimates from the Bureau of Labor and Statistics (BLS) Occupational Employment Statistics (OES) database. The OES survey data contain the number of people employed at each occupation for each 4-digit NAICS industry. We calculate the total industry cost of input (wage) for each occupation by multiplying the annual mean wage of the occupation by the number of people employed in an industry at that occupation. We then create a wage-weighted average skill for each 4-digit NAICS code, using the job zone assigned to each occupation according to the O*NET database. To aggregate our wage-weighted skill measure to broader industry classifications (e.g., Fama–French 48 industry portfolios), we compute the average skill level across the 4-digit NAICS contained in the broader industries.

3.3.2.4. Industry Risk/Complexity Index

Recall that uncertainty or riskiness is a necessary but non-sufficient condition for the implications of the model to hold. A fundamental assumption of the model is the need for coordinating complementary investments. Therefore, the production of *complex* goods or services in *uncertain* environments benefits to a larger extent from agglomeration externalities. For this reason, we construct an index to capture both risk and complexity based on the two metrics above. More specifically, the volatility and skill metrics are standardized and averaged.[25] Then, the resulting values are normalized so that the index ranges from 0 to 1.

3.4. Kernel Density Plots

3.4.1. Initial Inspection

Figure 3.1 plots the kernel density estimations (equation 3.12) for the highest and lowest risk/complexity industries (i.e., “Electronic equipment” and “Meals, restaurants, and hotels,” respectively). Consistent with the predictions of our model, the probability of two firms being located within 20 miles of each other is about four times larger in the

[23]Some examples of occupations in this category include taxi drivers, amusement and recreation attendants, and nonfarm animal caretakers.

[24]Examples of occupations in Job Zone 5 include lawyers, sports medicine physicians, surgeons, treasurers, and controllers.

[25]In this way, both measures, which initially differ in levels, become comparable.

highest risk/complexity industry than in the lowest risk/complexity industry. In fact, most of the differences between the two densities are driven by distances of less than 40 miles. That is, there seems to exist significantly more spatial concentration for the high risk/complexity industry than for the low risk/complexity industry. For larger distances, the densities are quite similar.

However, as explained in Section 3.3.3.3.1, it is necessary to control for industry size and concentration in order to gauge whether an industry is abnormally spatially concentrated. Consequently, Figure 3.2 further contrasts the two density estimates in the figure above against their respective 95% confidence intervals. Panel A shows that, indeed, the highest risk/complexity industry exhibits significant abnormal clustering (localization) for distances of less than 30 miles. The kernel density estimate exceeds the 95% confidence interval upper limit at those low distance values. In contrast, Panel B shows that this is not the case for the lowest risk/complexity industry. The kernel density estimate lies much closer and is within the confidence interval for most distance values.

3.4.2. Localization and Risk/Complexity

In the previous section, we provided evidence consistent with high risk/complexity industries exhibiting higher levels of localization than low risk/complexity industries. However, it is not possible to draw generalizable conclusions by comparing only two industries. In this section, we aim to provide more generalizable evidence for the implications of the model via broader and more systematic analyses.

As mentioned in the Data section, we base our analysis on the Fama and French 48 industry classifications. We exclude the finance and utilities industries, as well as any industry for which there are fewer than 100 firms in our sample.[26] This leaves us with 24 industries, which represent 91.1% of the firms in our initial sample. Table 3.1 lists the 24 industries, along with their annualized volatility, their required worker skill level, and their risk/complexity (RC) index (i.e., the combination of industry volatility and skill). The industries are ordered based on their RC index. Consistent with general intuition, the highest risk/complexity industries are “Electronic equipment,” “Measuring and control equipment,” and “Computers,” while the lowest risk/complexity industries are “Meals, restaurants, and hotels,” “Food,” and “Retail.”

We now turn to analyzing the implications of the model. Recall that Proposition 3.2 suggests that in dense environments, knowledge spillovers generate gains from coordination that are increasing on the uncertainty/riskiness of an industry. Moreover, as a consequence of the assumptions of the model, these coordination gains should be more salient in industries that focus on complex goods or services.

Therefore, in our empirical setting, if knowledge spillovers are an important determinant of a firm’s location decision, then the industry localization index should be positively correlated with the RC index at close distance ranges (i.e., there should be a higher

[26] This cutoff at 100 firms per industry greatly increases the accuracy of constructing the DO counterfactual.

degree of clustering in higher risk/complex industries). Further, this relationship should dissipate at longer distances. We explore this in Figure 3.3.

Consistent with the predictions of our model, Panel A of Figure 3.3 shows that there exists a strong positive relationship between the localization index and the RC index of the 24 industries in our sample for close distances between 0 and 20 miles. This relationship weakens substantially once we increase the distance interval to between 20 and 40 miles (Panel B). For larger distances, the previous relationship disappears completely (Panels C and D).

Overall, the evidence in Figure 3.3 is consistent with knowledge spillovers driving the localization decisions of firms. Furthermore, this is consistent with the findings in the literature that knowledge spillovers operate within close distances (e.g., Arzaghi and Henderson 55). Next, we estimate similar kernel densities just as we have done so far in the context of patents.

3.4.3. Robustness

While headquarters location provides a first approximation for the locality of a firm's activity, a firm may also conduct some of its operations outside its headquarters location. For instance, Honeywell is headquartered in New Jersey, while a significant portion of its patents are produced in Boston and in the San Francisco economic areas. We obtain information on inventors from the Harvard Patent Network Dataverse, which contains the locations of inventors for over 151,000 U.S. patents from 2006 to 2009. We are concerned with the patents that can be assigned to a firm in the Compustat universe at the time of the patent application. We focus on patent applications for the period from 2006 to 2009.[27]

We repeat the localization test using inventor patent locations instead of headquarters locations. Figure 3.4 displays the results. Interestingly, the degree of localization for inventors is even stronger than for headquarters locations for close distances. This result is reassuring, since one would expect a higher degree of localization for industries in which knowledge spillovers play a prominent role. In addition, as with the headquarters locations, there is a positive relation between the localization index and the RC index for the 17 industries in our sample. This relationship is most pronounced for close distances between 0 and 20 miles (Panel A) and disappears for larger distances (Panels B to D). Overall, our robustness analysis strengthens our confidence in the predictions of our model and the main empirical results in this paper.

[27] We restrict the sample to a 4-year period for computational reasons. Also, ending the sample in 2009 provides a 6-year window for a patent application to be granted, which mitigates truncation problems. Our sample consists of 5,670 patents from with 12,769 inventor locations, operating in 17 different industries.

3.4.4. Research and Development Expenses

Thus far, the results are consistent with our model, which suggests that the knowledge spillover channel is an important determinant of industry clustering. However, the results are also consistent with a labor-matching channel. For example, dense urban environments can support thicker labor markets, which in turn can facilitate the matching of workers and employers. It is natural to expect that the gains to matching may be greater for industries in which talent is scarce (complex industries) or in industries where the risk of getting laid off is higher (risky industries). Thus, clustering should generate greater gains from labor matching for riskier and more complex industries.

We implement two approaches to show that the knowledge spillover channel has an increasing partial effect for riskier and more complex industries, above and beyond any effect driven by labor matching. First, we investigate the industry clustering of R&D expenses, which are commonly argued to be particularly sensitive to and indicative of knowledge spillovers (CITES). Second, we implement an approach to isolate the similarity and timing of investment opportunities. While the labor-matching channel can partially explain the cross-sectional relationship of industry clustering, it should have little to say about the timing and similarity of investment between firms within a location. In contrast, our model suggests that firm's investment decisions are interdependent.

As a first step, we repeat our analysis of headquarter location in Section 3.4.3.4.2, with the only difference being that we weight each pair of headquarters locations by their aggregated R&D expenses when computing the kernel densities. Consequently, the resulting kernel densities indicate the probability of an additional dollar of R&D agglomerating within a certain distance for a given industry. Figure 3.5 displays the results of the localization test. Overall, R&D in riskier or more complex industries clusters even more than headquarters locations, and the relation among them is strongly positive for close distances between 0 and 20 miles (Panel A). As in our previous tests, the relation between localization and the RC index dissipates at longer distances (Panels B to D).

Next, we modify the approach developed in [69] to capture similarity in the timing of investment decisions. In particular, we examine whether the R&D expenses of firms located close by (i.e., within 20 miles) exhibit greater similarity within riskier and more complex industries. We follow two steps:

Step 1) For each firm-year, we obtain residuals by estimating the specification

$$R\&D_{it} = \alpha + BX_{i,t-1} + \widetilde{r}_{it}, \tag{3.14}$$

where $X_{i,t-1}$ is a vector of firm characteristics and time dummies. The residual $\widetilde{r}_{it}$ captures the unexplained component of R&D expenses.

We report the results for our estimation of Step 1 in Panel A of Table 3.2. We estimate three specifications. In the first specification, we account for any static differences across headquarters location by implementing Core-Based Statistical Area (CBSA) fixed effects transformations. In the second specification, we additionally control for heterogeneity in the level of R&D across industries by including industry dummies. In the third specification, we exploit within-firm variation by implementing firm fixed effects transformations.[28]

Step 2) For each possible pair of firms in a given industry, we then calculate the absolute value of the differences in residuals from Step 1 and estimate the following specification

$$\begin{aligned} |\widetilde{r}_{it} - \widetilde{r}_{jt}| = \beta_0 + \beta_1 1(d \leq 20miles)_{ij} \times RC\ index_{ij} + \\ \beta_2 1(d \leq 20miles)_{ij} + \beta_3 RC\ index_{ij} + \epsilon_{ijt}, \end{aligned} \tag{3.15}$$

where $1(d \leq 20miles)_{ij}$ is an indicator that takes the value of 1 if the headquarters of firm i and firm j are within 20 miles, and zero otherwise. If $\beta_1 < 0$, then firms headquartered within 20 miles make more similar R&D expenses decisions in relatively riskier and more complex industries, on average. Alternatively, we estimate similarities between changes in R&D expenses by estimating the following specification

$$\begin{aligned} |(\widetilde{r}_{it} - \widetilde{r}_{i,t-1}) - (\widetilde{r}_{jt} - \widetilde{r}_{j,t-1})| = \beta_0 + \beta_1 1(d \leq 20miles)_{ij} \times RC\ index_{ij} + \\ \beta_2 1(d \leq 20miles)_{ij} + \beta_3 RC\ index_{ij} + \epsilon_{ijt}. \end{aligned} \tag{3.16}$$

We report estimates of equation 3.15 in Panel B of Table 3.2. The statistically negative coefficient on $1(d \leq 20mi) \times RC\ index$ in specifications 2 and 3 indicates that firms located within 20 miles exhibit a greater degree of similarity in R&D expenses in riskier and more complex industries.

Finally, we report estimates of equation 3.16 in Panel C of Table 3.2. The coefficient on $1(d \leq 20mi) \times RC\ index$ is negative and statistically significant for all specifications, indicating that not only R&D expenses of firms headquartered close by are more similar in more uncertain and more complex industries, but also changes in R&D expenses.

We recognize that many labor market forces are likely to drive firm clustering decisions and it is not our aim to disprove these channels, or even to measure the relative importance of each channel. Instead, our aim is to convincingly show that knowledge spillovers have a positive partial effect on industry clustering. The clustering of R&D expenses in riskier and more complex industries in the cross-section, and more importantly through time, is strongly supportive of our goal. The similarity of investment decisions, and the clustering of investment through time, are inputs specific to our model.

3.4.5. Customer-Supplier Proximity

Firms in bilateral relationships, such as customers and suppliers, are likely to develop relation-specific investments (e.g., [70]). These relation-specific investments cause the

[28]Note that the firm fixed effects transformations subsume the industry dummies and CBSA level effects, as these are invariant through time within a given firm.

firms' investments to be heavily interdependent, which is consistent with the complementarity assumption of our model. Locating a headquarters near customers can allow suppliers to learn information about the investment opportunities of the customers, which in turn can increase the precision of their expectations regarding their own investment payouts. In turn, customers can learn from suppliers regarding the quality and timing of the production of intermediate goods. Our model predicts that a firm that sells a high proportion of its output to a few important customers should gain more from proximity to customers in risky/complex industries. Thus, we study the relationship between customer–supplier headquarters distance and industry risk/complexity.

We identify suppliers and customers from the Compustat segment files.[29] According to financial accounting standards, publicly listed corporations are required to disclose the amount of revenue raised from each customer that accounts for at least 10% of total revenue. Often, suppliers only list abbreviated or informal customer names.[30] To map the information from the customer–supplier file to a firm's financial and headquarters location information in Compustat, we implement the name-matching algorithm implemented by [71].[31] Our final sample includes 2,323 customer–supplier pairs from 1997 to 2013.

In our sample, we can identify suppliers that are heavily dependent on customers, but we cannot necessarily identify customers that depend heavily on suppliers.[32] The average customer is approximately 14 times as large as the average supplier in our sample, according to total assets. Thus, in our sample, it is more likely that customer locations enter the location decisions of suppliers rather than supplier locations affecting the location decisions of customers. As such, we construct our tests around suppliers' decisions to locate near customers. In particular, we examine the relationship between the risk and complexity of customer industries and the distance that suppliers choose to locate from customers. Nonetheless, we perform a robustness exercise in which we view the colocation of customers and suppliers as a joint decision in relation to the combined risk of the two adjoining industries, and we find similar effects.[33].

Results from estimating the relationship between customer–supplier proximity and customer risk/complexity are presented in Table 3.3. The dependent variable in Columns 1-2 is the natural log of the distance between headquarters locations (in miles) for each customer–supplier pair.[34] The coefficient estimates suggest that when suppliers sell to customers in the highest risk/complexity industry, their headquarters tend to be 172 miles closer to the customer, on average, when compared to suppliers that sell to customers that are in the lowest risk/complexity industry.[35] Furthermore, suppliers are 7.9 percentage

[29] These data have been used in recent studies by [71], [70], and [72], among others.

[30] As highlighted by [72], the SFAS No. 131 does not require firms to list the identity of major customers. It only requires firms to record the presence of a customer that makes up at least 10% of sales. However, most suppliers voluntarily disclose the customer name.

[31] We would like to thank Ted Fee and Shawn Thomas for providing us with this algorithm, which was recently extended to include firms through 2013.

[32] For example, Walmart constitutes at least 10% of sales for six firms in Compustat, but none of the referenced suppliers constitute 10% of Walmart's expenditures.

[33] These results are reported in Appendix B.

[34] We only include one observation for each customer–supplier pair.

[35] The average log distance is 6.317. Consequently, the marginal effect at the mean is exp(6.317)–exp(6.317–0.3721), which yields 172.11.

points more likely to locate within 20 miles of customers in the highest risk/complexity industry relative to customers in the lowest risk/complexity industry (Column 4). The 7.9 percentage point increase is economically large, representing 104.9% of the sample average propensity for suppliers to locate near customers.

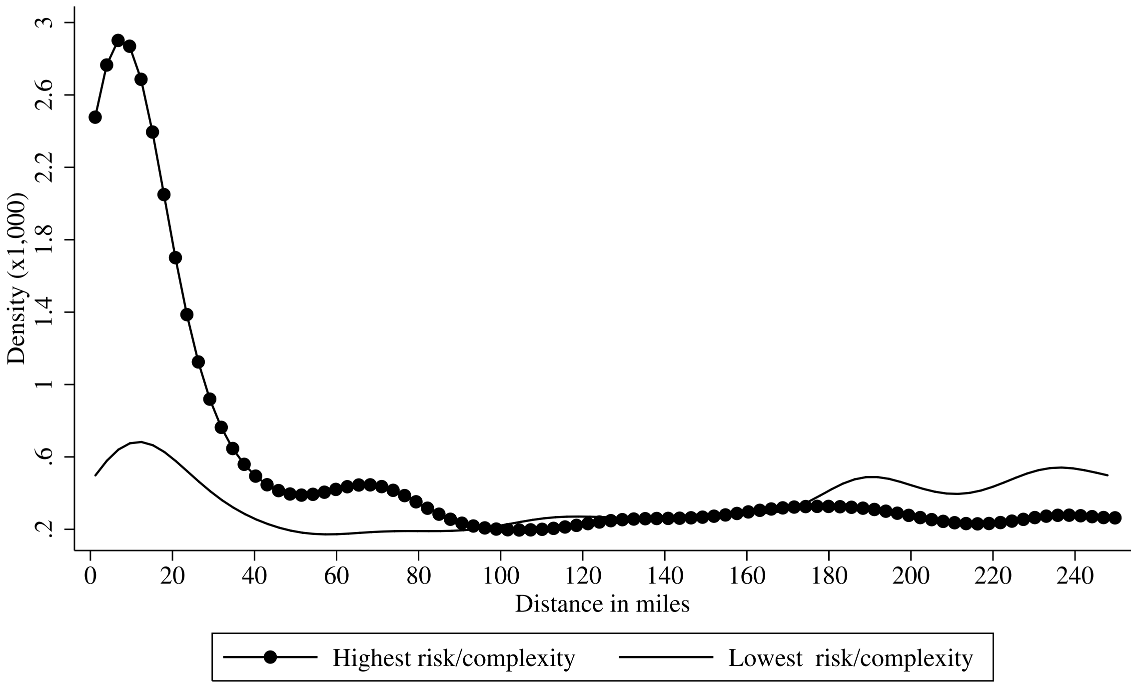
Table 3.4 examines the relationship between customer risk/complexity and the likelihood of suppliers to locate within 20 miles of customers, between 20 miles and 40 miles, between 40 miles and 60 miles, and between 60 miles and 80 miles. Consistent with the results presented in previous sections, the estimates indicate that the relationship between customer risk/complexity and customer–supplier proximity dissipates at larger distances. The coefficient estimate associated with customer risk/complexity is almost 6 times larger for distances within 20 miles than for distances between 20 and 40 miles. Further, the relationship between customer risk/complexity and customer–supplier proximity is not statistically significant for distances between 20 and 60 miles (Columns 2 and 3), and it changes sign for distances between 60 and 80 miles (Column 4).

3.5. Conclusion

We build on the global games literature by developing a theory in which investment decisions can be thought of as games of incomplete information in which payouts depend on the decisions of related firms. We propose that when firms observe noisy private signals about investment opportunities, co-locating with related firms facilitates communication and the sharing of private information. By sharing private information, firms gain precision in project valuations. In this context, dense urban centers can be thought of as a technology that facilitates face-to-face interaction and knowledge sharing, reducing project uncertainty and therefore improving project selection.

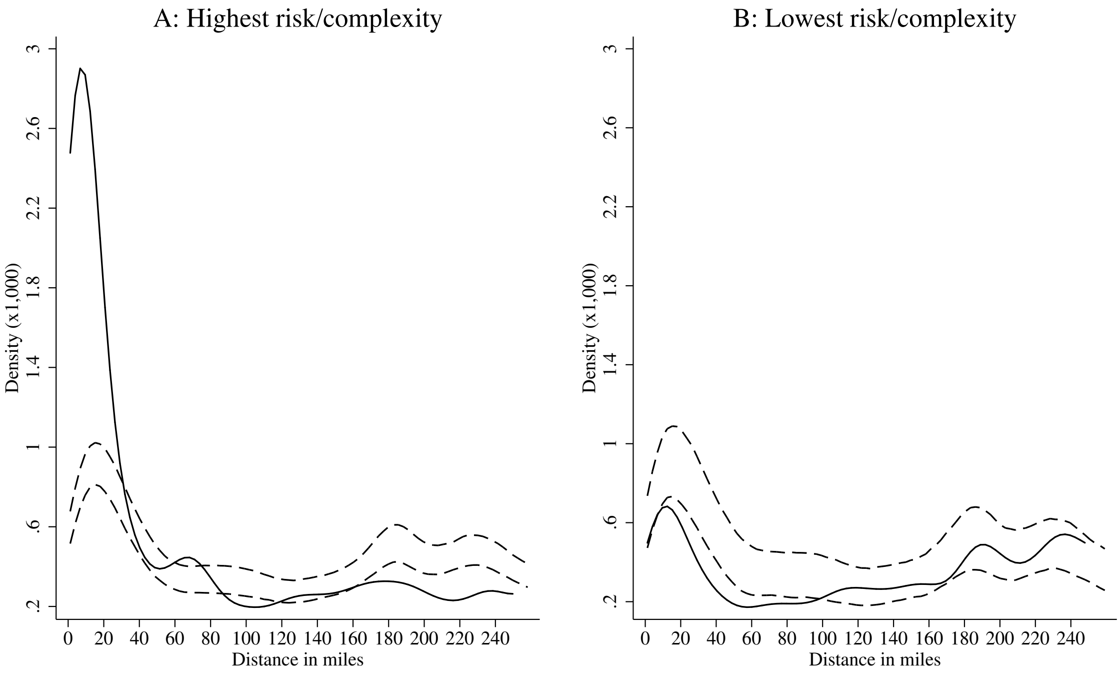
Our model shows that the benefits from this process are greater for firms in relatively more uncertain industries and more complex industries. Consistent with this proposition, we show that agglomeration patterns are significantly more pronounced for relatively more complex industries in relatively more uncertain environments, using both firm headquarters and patent inventor locations as a proxy for business activity. Further, we show that this pattern also holds for R&D expenses and in customer–supplier locations. Finally, we show that customer–supplier proximity is strongly and positively related to the risk and complexity of the customer industry. Overall, our results link knowledge sharing with intrinsic industry characteristics, and they contribute to the understanding of the important phenomenon of firm clustering.

FIGURE 3.1: Kernel densities



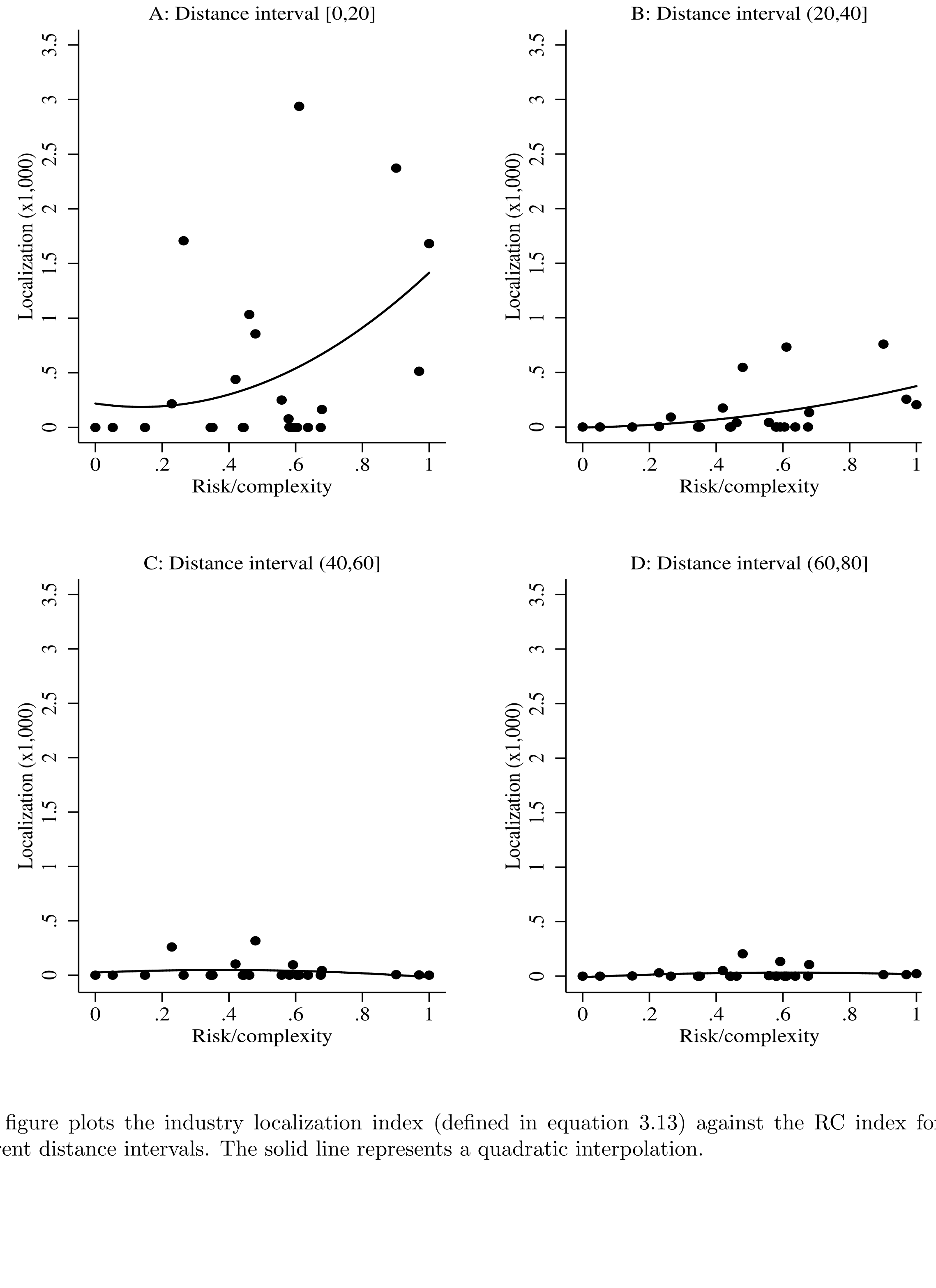
Kernel densities of the highest and lowest risk/complexity industries (×1,000 for scale). Industry groups are based on the Fama and French 48 industries classification (industries related to Finance and Utilities are not considered). Industries are ranked by risk/complexity based on stock index volatility and skill requirements.

Figure 3.2: Kernel densities with confidence intervals



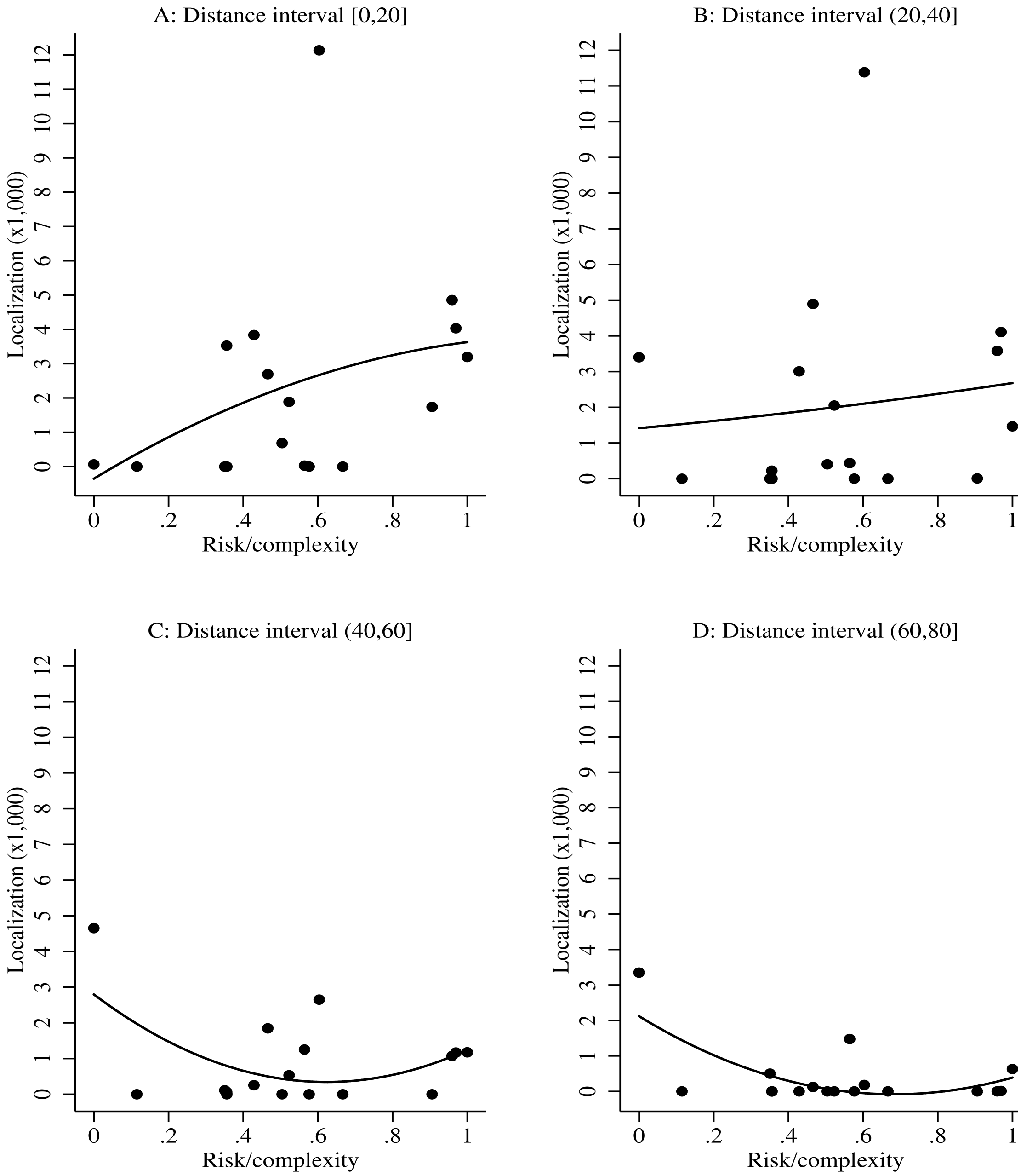
Kernel densities of the highest and lowest risk/complexity industries (×1,000 for scale) with 95% confidence intervals. Industry groups are based on the Fama and French 48 industries classification (industries related to Finance and Utilities are not considered). Industries are ranked by risk/complexity based on stock index volatility and skill requirements.

FIGURE 3.3: Industry localization index and risk/complexity



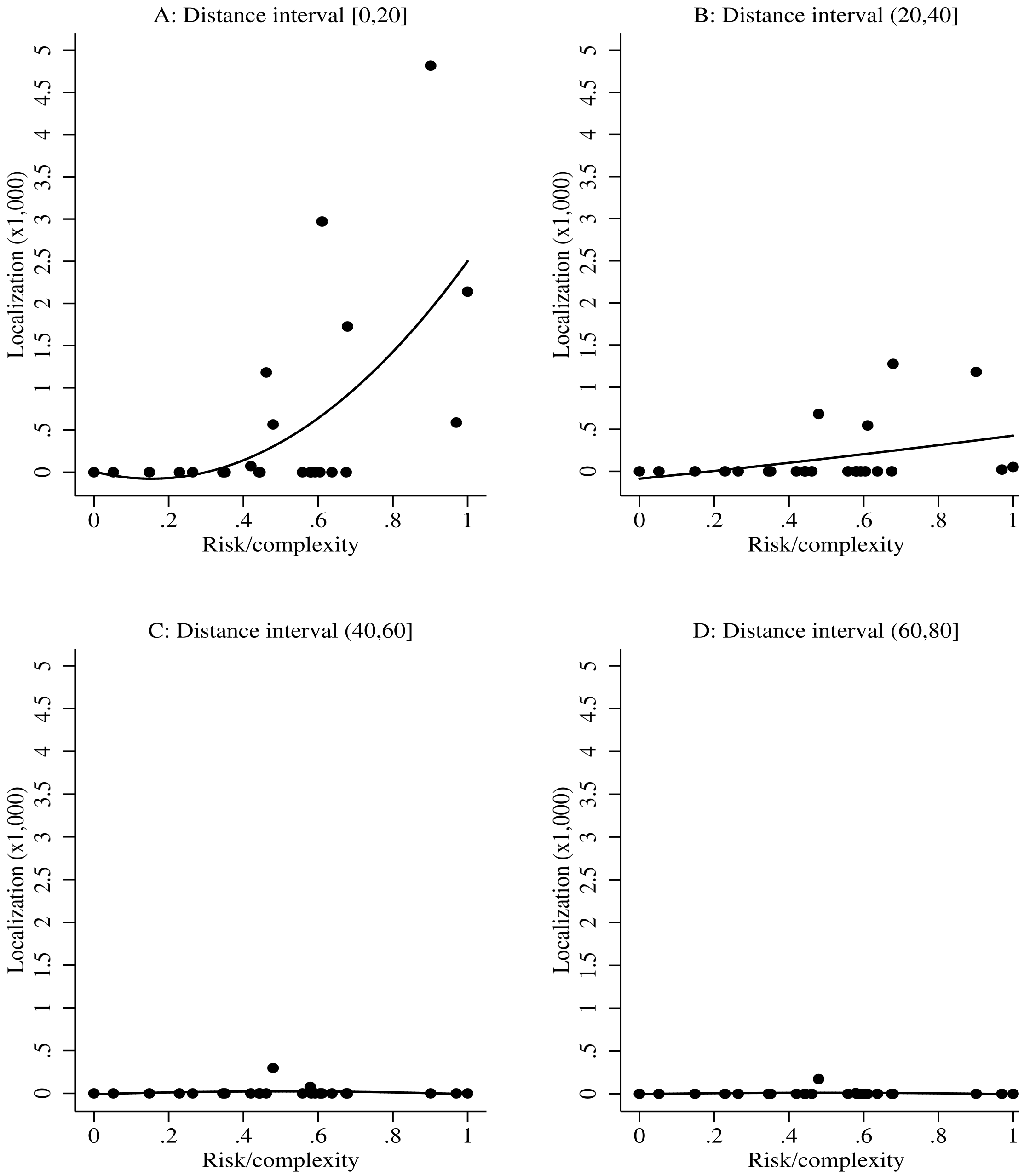
This figure plots the industry localization index (defined in equation 3.13) against the RC index for different distance intervals. The solid line represents a quadratic interpolation.

Figure 3.4: Industry localization index of patent activity and risk/complexity



This figure plots the industry localization index of patent activity (defined in equation 3.13) against the RC index for different distance intervals. The solid line represents a quadratic interpolation.

Figure 3.5: Industry localization index and risk/complexity when weighting by R&D



This figure plots the industry localization index when weighting headquarters by their R&D expense against the RC index for different distance intervals. The solid line represents a quadratic interpolation.

Table 3.1: Summary

Industry	Annualized volatility	Skill	Rank (volatility)	Rank (skill)	RC index
Electronic equipment	0.339	3.508	1	4	1
Measuring and control equipment	0.322	3.556	2	2	0.970
Computers	0.271	3.763	6	1	0.901
Automobiles	0.320	2.734	4	17	0.679
Steel	0.320	2.718	3	18	0.675
Machinery	0.262	3.071	7	10	0.637
Oil	0.215	3.368	13	6	0.611
Personal Services	0.230	3.229	11	9	0.605
Electrical equipment	0.255	2.997	9	11	0.592
Construction	0.271	2.836	5	16	0.581
Healthcare	0.202	3.372	15	5	0.579
Telecommunications	0.200	3.328	16	7	0.558
Pharmaceuticals	0.147	3.524	22	3	0.479
Entertainment	0.260	2.580	8	20	0.461
Chemicals	0.211	2.920	14	13	0.444
Construction materials	0.239	2.688	10	19	0.442
Medical equipment	0.154	3.299	21	8	0.420
Transportation	0.185	2.853	18	14	0.351
Wholesale	0.168	2.969	20	12	0.345
Clothing	0.229	2.255	12	23	0.264
Household consumer goods	0.141	2.844	24	15	0.229
Retail	0.173	2.367	19	21	0.149
Food	0.142	2.329	23	22	0.052
Meals, restaurants, and hotels	0.188	1.812	17	24	0

This table ranks the different industries based on stock index volatility and skill requirements. To construct the risk/complexity index (RC index), the volatility and skill metrics are standardized and averaged. Then, the resulting values are normalized so that the RC index ranges from 0 to 1.

Table 3.2: Clustering of R&D Expenses

Panel A: First Stage

	(1)	(2)	(3)
Leverage	0.0029	0.0261***	0.0186***
	(0.0026)	(0.0024)	(0.0026)
log(Sales)	-0.0302***	-0.0144***	0.0013*
	(0.0006)	(0.0006)	(0.0007)
log(Assets)	0.0099***	-0.0013**	-0.0392***
	(0.0006)	(0.0006)	(0.0008)
Market-to-book	0.0000**	0.0000	0.0000**
	(5.79e-06)	(5.30e-06)	(3.78e-06)
Z-score	0.0000	0.0000	0.0000***
	(5.63e-07)	(5.15e-07)	(3.77e-07)
ROA	0.0000**	0.0000	-0.0000***
	(7.87e-06)	(7.19e-06)	(7.88e-06)
Year FE	yes	yes	yes
CBSA FE	yes	yes	no
Industry FE	no	yes	no
Firm FE	no	no	yes
N	45,640	45,640	45,640
R^2	0.251	0.376	0.803

Panel B: Residual Distance

	(1)	(2)	(3)
1(d≤20mi)×RC index	-0.0103	-0.0198**	-0.0232***
	(0.0085)	(0.0090)	(0.0063)
1(d≤20mi)	-0.0068	0.0012	0.0066**
	(0.0048)	(0.0050)	(0.0030)
RC index	0.0409***	0.0452***	0.0223**
	(0.0158)	(0.0166)	(0.0091)

Panel C: Residual Changes

	(1)	(2)	(3)
1(d≤20mi)×RC index	-0.0298***	-0.0301***	-0.0260***
	(0.0080)	(0.0080)	(0.0077)
1(d≤20mi)	0.0114***	0.0121***	0.0091**
	(0.0041)	(0.0041)	(0.0038)
RC index	0.0242	0.0217	0.0229*
	(0.0147)	(0.0143)	(0.0132)

This table reports the results of the two-step test of residual distances. Panel A reports the results of the first step, where R&D is regressed on lagged firm controls to obtain the residuals for step 2. Panel B reports the results of the estimation of equation 3.15, where the dependent variable is the absolute value of the residual distances for all possible pairs of firms in an industry. The explanatory variable of interest is $1(d \leq 20mi) \times RC\ index$, the interaction between a dummy variable that takes the value of 1 if the pair of firms are headquartered within 20 miles and zero otherwise, and the RC index. Panel C reports the results of the estimation of equation 3.16, where the dependent variable is the absolute pair difference in changes in the first-stage residuals. In panels B and C, standard errors are clustered at the Fama–French 48 industry-year level and are reported in parentheses, below the coefficient estimates. ***p<0.01, **p<0.05, *p<0.1.

Table 3.3: Supplier Locations and Customer Risk

	log(distance)		Within 20 mi	
	(1)	(2)	(3)	(4)
RC index$_{customer}$	-0.4288***	-0.3721***	0.0942***	0.0785**
	(0.1274)	(0.1361)	(0.0321)	(0.0323)
log(Sales)$_{customer}$		0.0767		-0.0121*
		(0.0469)		(0.0063)
log(Sales)$_{supplier}$		-0.0465**		0.0037*
		(0.0215)		(0.0019)
Number of firms$_{customer}$		0.0288		-0.0060
		(0.0349)		(0.0063)
Number of firms$_{supplier}$		0.0202		0.0005
		(0.0269)		(0.0051)
R^2	0.0445	0.0514	0.0694	0.0747
N	2,323	2,323	2,323	2,323

This table reports estimates for the relation between customer and supplier locations and customer risk/complexity. Customers are identified from the Compustat segment files. The dependent variable in Columns 1-2 is the natural log of the distance (in miles) between a customer and a supplier headquarters location. The binary dependent variable in Columns 3-4 is equal to 1 if a customer and a supplier are located within 20 miles of each other. Distances are calculated from geographic coordinates for corporate headquarters ZIP codes. Standard errors are clustered at the Fama–French 48 industry level and are reported in parentheses, below the coefficient estimates. ***p<0.01, **p<0.05, *p<0.1.

Table 3.4: Supplier Locations and Customer Risk at Larger Distances

	Within 20 mi (1)	Within 20-40mi (2)	Within 40-60mi (3)	Within 60-80mi (4)
RC index$_{customer}$	0.0785**	0.0132	0.0141	-0.0148*
	(0.0323)	(0.0085)	(0.0119)	(0.0084)
log(Sales)$_{customer}$	-0.0121*	0.0004	-0.0050	0.0015
	(0.0063)	(0.0021)	(0.0036)	(0.0018)
log(Sales)$_{supplier}$	0.0037*	-0.0029*	0.0008	-0.0026
	(0.0019)	(0.0015)	(0.0017)	(0.0030)
Number of firms$_{customer}$	-0.0060	0.0044	-0.0024	0.0000
	(0.0063)	(0.0029)	(0.0040)	(0.0014)
Number of firms$_{supplier}$	0.0005	0.0004	0.0016	-0.0023
	(0.0051)	(0.0025)	(0.0025)	(0.0029)
R^2	0.0747	0.0241	0.0260	0.0167
N	2,323	2,148	2,096	2,068

This table reports estimates for the relation between customer and supplier locations and customer risk/complexity at larger distances. Customers are identified from the Compustat segment files. The dependent variables in Columns 1-4 are indicators for whether a customer and supplier are located within 20 miles, between 20 miles and 40 miles, between 40 miles and 60 miles, and between 60 miles and 80 miles, respectively. Distances are calculated from geographic coordinates for corporate headquarters ZIP codes. Standard errors are clustered at the Fama–French 48 industry level and are reported in parentheses, below the coefficient estimates. ***p<0.01, **p<0.05, *p<0.1.

Bibliography

[1] Mayer Christopher Gyourko, Joseph and Todd Sinai. Superstar Cities. *American Economic Journal: Economic Policy*, 5:167–99, 2013.

[2] Chang-tai Hsieh and Enrico Moretti. Why Do Cities Matter? *NBER Working Paper*, 2015.

[3] Gilles Duranton Combes, Pierre-Philippe, Gilles Duranton Laurent Gobillon, Pierre-Philippe, and Laurent Gobillon. Spatial wage disparities: Sorting matters! *Journal of Urban Economics*, 2008.

[4] Rafael LaPorta Florencio Lopez-de-Silanes Gennaioli, Nicola and Andrei Shleifer. Human Capital and Regional Development. *Quarterly Journal of Economics*, 2013.

[5] Daron Acemoglu and Melissa Dell. Productivity Differences Between and Within Countries. *American Economic Journal: Macroeconomics*, 2010.

[6] Stijn Van Nieuwerburgh and Pierre-Olivier Weill. Why Has House Price Dispersion Gone Up? *Review of Economic Studies*, 77(4):1567–1606, 2010.

[7] Gautier Pieter Albrecht, James and Susan Vroman. Directed Search in the Housing Market. *Review of Economic Dynamics*, 2016.

[8] Albert Saiz. The geographic determinants of housing supply. *Quarterly Journal of Economics*, 2010.

[9] Peter Ganong and Daniel Shoag. Why Has Regional Income Convergence in the U.S. Declined? *HKS Working Paper No. RWP12-028*, 2013.

[10] Richard Hornbeck and Enrico Moretti. Who Benefits From Productivity Growth? The Local and Aggregate Impacts of Local TFP Shocks on Wages, Rents, and Inequality. *NBER Working Paper*, 2015.

[11] Edward L. Glaeser and David C Mare. Cities and Skills. *Journal of Labor Economics*, 2001.

[12] J.M. Shapiro. Smart Cities: Quality of Life, Productivity, and the Growth Effects of Human Capital. *The Review of Economics and Statistics*, 2006.

[13] Victor Couture. Knowledge spillovers in cities: An auction approach. *Journal of Economic Theory*, 157:668–698, 2015.

[14] David Albouy. What are Cities Worth? Land Rents, Local Productivity, and the Total Value of Amenities. *The Review of Economics and Statistics*, 2016.

[15] David Albouy and Nathan Seegert. The Optimal Population Distribution Across Cities and The Private-Social Wedge. *Working Paper*, 2016.

[16] Cecile Gaubert. Firm Sorting and Agglomeration. *Working Paper*, 2016.

[17] Klaus Desmet and Esteban Rossi-Hansberg. Spatial Development. *American Economic Review*, 2014.

[18] Gilles Duranton Behrens, Kristian and Frederic Robert-Nicoud. Productive cities: Agglomeration, selection, and sorting. *Journal of Political Economy*, 2014.

[19] Juan Carlos Serrato and Owen Zidar. Who benefits from state corporate tax cuts? A local labor markets approach with heterogeneous firms. *American Economic Review*, 2016.

[20] Pinheiro Roberto Eeckhout, Jan and Kurt Schmidheiny. Spatial sorting: Why New York, Los Angeles and Detroit attract the greatest minds as well as the unskilled. *Working Paper*, 2010.

[21] Jan Eeckhout and Philipp Kircher. Assortative matching with large firms: Span of control over more versus better workers. *Working Paper*, 2012.

[22] Kristian Behrens and Frederic Robert-Nicoud. Agglomeration theory with heterogeneous agents. *Handbook of Regional and Urban Economics*, 2015.

[23] De la Roca Jorge Puga, Diego and Gianmarco Ottaviano. City of dreams. *Working Paper*, 2015.

[24] Freedman Matthew Pavan-Ronni Baum-Snow, Nathaniel. Why has urban inequality increase. *Working Paper*, 2014.

[25] Diamond Rebecca. The Determinants and Welfare Implications of US Workers' Diverging Location Choices by Skill: 1980-2000. *American Economic Review*, 2016.

[26] Donald R. Davis and Jonathan I. Dingle. The Comparative Advantage of Cities. *NBER Working Paper*, 2015.

[27] Lu Han and William Strange. Bidding Wars for Houses. *Real Estate Economics*, 2014.

[28] Sherwin Rosen. Wages-based Indexes of Urban Quality of Life. *In Current Issues in Urban Economics*, 1982.

[29] Jennifer Roback. Wages, Rents, and the Quality of Life. *Journal of Political Economy*, 1982.

[30] Nathaniel Baum-Snow and Ronni Pavan. Inequality and City Size. *The Review of Economics and Statistics*, 2013.

[31] Patrick Kline and Enrico Moretti. Local Economic Development, Agglomeration Economies and the Big Push: 100 Years of Evidence from the Tennessee Valley Authority. *Quarterly Journal of Economics*, 2014.

[32] Arnaud Costinot and Jonathan Vogel. Matching and Inequality in the World Economy. *Journal of Political Economy*, 2010.

[33] Lawrence Katz and Kevin Murphy. 'Changes in Relative Wages, 1963-1987: Supply and Demand Factors. *The Quarterly Journal of Economics*, 1992.

[34] Antonio Ciccone and Giovanni Peri. 'Identifying human-capital externalities: Theory with applications. *The Review of Economic Studies*, 2005.

[35] Klenow Peter Hsieh, Chang-tai. Misallocation and Manufacturing TFP in China and India. *The Quarterly Journal of Economics*, 2009.

[36] Blum Bernardo Bacolod, Marigee and William Strange. Skills in the city. *Journal of Urban Economics*, 2009.

[37] Enrico Moretti. Local Labor Markets. *Handbook of Labor Economics, Chapter 14*, 2011.

[38] Enrico Moretti. *The New Geography of Jobs*. Houghton Mifflin Harcourt, 2012.

[39] Enrico Moretti. Human capital externalities in cities. *Handbook of Regional and Urban Economics*, 4:2243–2291, 2004.

[40] Edward Glaeser and David Maré. Cities and skills. *Journal of Labor Economics*, 19 (2):316–42, 2001.

[41] Luis Garicano Willie Fuchs and Luis Rayo. Optimal Contracting and the Organization of Knowledge. *The Review of Economic Studies*, 2015.

[42] Alfred Marshall. *Principles of Economics*. Macmillan, London, 1920.

[43] Hans Carlsson and Eric Van Damme. Global games and equilibrium selection. *Econometrica*, 61(5):989–1018, 1993.

[44] Stephen Morris and Hyun S. Shin. Global games: Theory and applications. In Mathias Dewatripont, Lars P. Hansen, and Stephen J. Turnovsky, editors, *Advances in Economics and Econometrics*, pages 56–114. 2003.

[45] E. Ozdenoren and K. Yuan. Feedback effects and asset prices. *Journal of Finance*, 63(4):1939–1975, 2008.

[46] Gilles Duranton and Henry G. Overman. Testing for localization using micro-geographic data. *Review of Economic Studies*, 72(4):1077–1106, 2005.

[47] Gilles Duranton and Diego Puga. Micro-foundations of urban agglomeration economies. *Handbook of Regional and Urban Economics*, 4:2063–2117, 2004.

[48] Pierre-Philippe Combes and Gilles Duranton. Labour pooling, labour poaching, and spatial clustering. *Regional Science and Urban Economics*, 36(1):1–28, 2006.

[49] Robert W. Helsley and William C. Strange. Knowledge barter in cities. *Journal of Urban Economics*, 56(2):327–345, 2004.

[50] Donald R. Davis and Jonathan I. Dingel. A spatial knowledge economy. Working paper., 2012.

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

[52] Michael Storper and Anthony J. Venables. Buzz: Face-to-face contact and the urban economy. *Journal of Economic Geography*, 4(4):351–370, 2004.

[53] David B. Audretsch and Paula E. Stephan. Company-scientist locational links: The case of biotechnology. *American Economic Review*, 86(3):641–652, 1996.

[54] Kristy Buzard, Gerald A. Carlino, Robert M. Hunt, Jake K. Carr, and Tony E. Smith. Localized knowledge spillovers: Evidence from the agglomeration of american R&D labs and patent data. Working paper., 2015.

[55] Mohammad Arzaghi and J. Vernon Henderson. Networking off Madison Avenue. *Review of Economic Studies*, 75(4):1011–1038, 2008.

[56] Gerald Carlino and William R. Kerr. Agglomeration and innovation. Working paper., 2014.

[57] Glenn Ellison, Edward L. Glaeser, and William R. Kerr. What causes industry agglomeration? Evidence from coagglomeration patterns. *American Economic Review*, 100(3):1195–1213, 2010.

[58] Giulia Faggio, Olmo Silva, and William C. Strange. Heterogeneous agglomeration. 2014. Working paper.

[59] Stuart S. Rosenthal and William C. Strange. Evidence on the nature and sources of agglomeration economies. *Handbook of Regional and Urban Economics*, 4:2119–2171, 2004.

[60] Avinash K. Dixit and Robert S. Pindyck. *Investment Under Uncertainty*. Princeton University Press, Princeton, NJ, 1994.

[61] Casey Dougal, Christopher A. Parsons, and Sheridan Titman. Urban vibrancy and corporate growth. *Journal of Finance*, 70(1):163–210, 2015.

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

[63] Paul M. Romer. Increasing returns and long-run growth. *Journal of Political Economy*, pages 1002–1037, 1986.

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

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

[66] George-Marios Angeletos and Chen Lian. Incomplete information in macroeconomics: Accomodating frictions in coordination. *Handbook of Macroeconomics, Vol. 2.*, 2016. Forthcoming.

[67] Berthold Herrendorf, Richard Rogerson, and Akos Valentinyi. Growth and structural transformation. In Philippe Aghion and Steven N. Durlauf, editors, *Handbook of Economic Growth*, volume 2, pages 855–941. 2014.

[68] Christo Pirinsky and Qinghai Wang. Does corporate headquarters location matter for stock returns? *Journal of Finance*, 61(4):1991–2015, 2006.

[69] Kelly Shue. Executive networks and firm policies: Evidence from the random assignment of MBA peers. *Review of Financial Studies*, 26(6):1401–1442, 2013.

[70] Shantanu Banerjee, Sudipto Dasgupta, and Yungsan Kim. Buyer–supplier relationships and the stakeholder theory of capital structure. *Journal of Finance*, 63(5): 2507–2552, 2008.

[71] C. Edward Fee and Shawn Thomas. Sources of gains in horizontal mergers: Evidence from customer, supplier, and rival firms. *Journal of Financial Economics*, 74(3):423–460, 2004.

[72] Michael G. Hertzel, Zhi Li, Micah S. Officer, and Kimberly J. Rodgers. Inter-firm linkages and the wealth effects of financial distress along the supply chain. *Journal of Financial Economics*, 87(2):374–387, 2008.

Appendix A

Appendix Chapter 1

A1. Housing Prices

I estimate equilibrium bids which will be representative of equilibrium housing prices. Here I closely follow [Albrecht, Gautier and Vroman 2013].

Write the expected bid as

$$E(b(w(s)) = \frac{\int_r^{w(s)} b(w(s))h(w(s))dw(s)}{\int_r^{w(s)} h(w(s))dw(s)}$$

where $b(w(s))$is the optimal bid in a first price auction and $h(w(s))$ is the density of the highest valuation drawn by the buyers visiting a particular seller, conditional on the seller having at least one visitor.

Assume that the reservation value will be the lowest wage available $w(\underline{s})$, so we can write:

$$E(b(w(s)) = \frac{\int_{w(\underline{s})}^{w(s)} b(w(s))h(w(s))dw(s)}{\int_{w(\underline{s})}^{w(s)} h(w(s))dw(s)} = \frac{\int_{\underline{s}}^{s} b(w(s))h(s)ds}{\int_{\underline{s}}^{s} h(s)ds}$$

Notice that

$$h(s) = f(s|H) = \frac{P(H|s)v^i(s)}{P(H)}$$

Given the properties of Poisson distributions, the probability that a buyer who has $w(s)$ has the highest valuation is

$$P(H|s) = e^{-\theta(1-V^i(s))}$$

and the unconditional probability that any buyer has the highest valuation is

$$P(H) = \int e^{-\theta(1-V^i(s)} v^i(s)ds = \frac{1-e^{-\theta}}{\theta}$$

So

$$h(s) = \frac{\theta e^{-\theta(1-V^i(s))} v^i(s)}{1-e^{-\theta}}$$

Given the optimal bidding in a first price auction,

$$b(w(s)) = w(s) - \frac{V(s)}{v(s)}$$

Basically, each bidder who faces an uncertain number of buyers would offer his expected value minus its virtual surplus.

Finally, calculate the expected bid:

$$E(b(w(s)) = \frac{\int_{\underline{s}}^{s} b(w(s))h(s)ds}{\int_{\underline{s}}^{s} h(s)ds} = w(s) - \frac{\int_{\underline{s}}^{s} \frac{(V(s))}{v(s)} h(s)ds}{\int_{\underline{s}}^{s} h(s)ds}$$

(A2. Full Characterization of the Two Cities Case) Lets prove the existence and uniqueness for two cities. Since there is free mobility, the utility of a worker of ability s must be equal across space. Workers must be indifferent between the two cities. Then, for every worker s, it must be the case that

$$EU^1(s) = EU^2(s) \quad \forall s \epsilon [\underline{s}, \overline{s}]$$

The second condition is that I have a fixed amount of talent within the country (closed economy), which means that

$$V(s) = V^1(s) + V^2(s) \quad \forall s \epsilon [\underline{s}, \overline{s}]$$

Thus I can re-write the above condition as:

$$\frac{1 - V^1(s)}{v^1(s)} e^{-\theta^1(1-V^1(s))} = \frac{1 - V(s) + V^1(s)}{v(s) - v^1(s)} e^{-\theta^2(1-V(s)+V^1(s))} \ \forall s \epsilon [\underline{s}, \overline{s}]$$

There are two important observations to make here: This is a difficult problem, since I must search for a function $V^1()$ that can hold this condition for every s. Now, it is crucial to note that, since this must hold for every point and that $v^1 = \frac{\partial V^1}{\partial s}$, this is equivalent to solving an ordinary differential equation (ODE) for every point s. Also notice that $< V(s), v(s), \theta^1, \theta^2 >$ are exogenous parameters.

Thus by rearranging the terms, I arrive at the following expression for the equilibrium ODE:

$$\frac{\partial V^1(s)}{\partial s} = v(s) \frac{[1 - V^1(s)] e^{-\theta^1(1-V^1(s))}}{[[1 - V^1(s)] e^{-\theta^1(1-V^1(s))} + [1 - V(s) + V^1(s)] e^{-\theta^2(1-V(s)+V^1(s))}]}$$

Given that I have a first-order differential equation with an initial condition, I must show that$f(s, V1(s))$ is Lipschitz–continuous in V^1 and continuous in s. Lipschitz continuity (i.e., bounded variation) is easy to observe. Notice that $f(s, V1(s))$ is bounded by one (since it is a weighted average), thus it is Lipschitz. Second, since$V$ is continuous by definition, the composition of continuous functions is also continuous, thus $f(s, V1(s))$ is continuous in s. Note that $< j^1, j^2 >$are endogenous parameters, since they depend on the number of workers that come to produce to the city. So, it is necessary to find a fixed point. I fix j^i, solve the system of differential equations, and compute a new tightness parameter. I do this until j^i converges to a stable parameter. I can now recall the Picard–Lindeloff Theorem, which states that there exists a unique solution to this contraction. □

(A2. Monotone likelihood property proof)

Given that our analytical results are comparisons of different levels of housing tightness, as this is an endogenous parameter, I would first like to know whether there is a monotonic relation between supply and tightness (since we will be observing supply and comparing different housing supplies). I must show that the elasticity of new buyers to new houses is smaller than one. The condition I must look for is

$$\frac{B_1 + \triangle B}{S_1 + \triangle S} < \frac{B_2 - \triangle B}{S_2}$$

This will always hold as long as $\frac{\triangle B}{\triangle S} < 1/2$

To prove that this is true, notice that

$$B_1 = \int (v(s) \frac{[1 - V^1(s)]e^{-\frac{B_1}{S_1}(1-V^1(s))}}{\left[[1 - V^1(s)]e^{-\frac{B_1}{S_1}(1-V^1(s))} + [1 - V(s) + V^1(s)]e^{-\frac{(1-B_1)}{S_2}(1-V(s)+V^1(s))}\right]})ds$$

After some algebraic manipulation, we can show that

$$\partial B_1 / \partial S_1 < 1$$

This is always lower than 1, Thus, tightness is monotonic in housing supply. What about relative tightness? Given that extra (or less) supply has an effect on the tightness of both cities, the last thing we must check is whether the change in tightness in City 1 is larger than the change in tightness in City 2:

$$\frac{B_1}{S_1} - \frac{B_1 + \triangle B}{S_1 + \triangle S} > \frac{B_2}{S_2} - \frac{B_2 - \triangle B}{S_2} \Longleftrightarrow \frac{B_1 \triangle S - S_1 \triangle B}{S_1(S_1 + \triangle S)} > \frac{\triangle B}{S_2}$$

This condition will depend on the level of the original elasticities. For example, if City 2 originally had a extremely restricted housing supply, then the number of people that leave the city will create a very strong change in its local housing market. In general, the following condition must hold:

$$\frac{\triangle B}{\triangle S} < \frac{1}{2}$$

Given this expression, I derive for this elasticity this will be the case, except for extreme values of S_2:

$$\vartheta \equiv \theta_1 / \theta_2 = \frac{S_2}{S_1} \frac{B_1}{B_2} = \frac{S_2}{S_1} \int (\frac{[1 - V^1(s)]e^{-\frac{B_1}{S_1}(1-V^1(s))}}{\left[[1 - V(s) + V^1(s)]e^{-\frac{(1-B_1)}{S_2}(1-V(s)+V^1(s))}\right]})ds$$

$$= \frac{S_2}{S_1} \int (\frac{[1 - V^1(s)] e^{\frac{1}{\vartheta} \frac{(1-V^1(s))}{[1-V(s)+V^1(s)]}}}{[[1 - V(s) + V^1(s)]]}) ds$$

I now wish to analyze how this relative tightness measure would change if, for example, the supply of housing in City 2 were to change. This will have a mechanical effect in terms of relaxing the tightness of the housing market in City 2, but as people move from City 1 to city two, this will have the same effect in City 1. The system of cities receives a "positive" shock; the question now is: In which city will the tightness condition relax more? If the monotonicity condition holds, I should expect that a higher supply of houses in City 2 makes the City 2 housing market less tight relative to City 1, thus relative tightness should increase:

$$\frac{\partial \vartheta}{\partial S_2} =$$

$$\frac{1}{S_1} \int (\frac{[1 - V^1(s)] e^{\frac{1}{\vartheta} \frac{(1-V^1(s))}{[1-V(s)+V^1(s)]}}}{[[1 - V(s) + V^1(s)]]}) ds - \frac{S_2}{S_1} \frac{[1 - V(s) + V^1(s)]}{[1 - V^1(s)]} \frac{1}{\vartheta^2} \int (\frac{[1 - V^1(s)] e^{\frac{1}{\vartheta} \frac{(1-V^1(s))}{[1-V(s)+V^1(s)]}}}{[[1 - V(s) + V^1(s)]]}) ds \frac{\partial \vartheta}{\partial S_2} \tag{A.1}$$

$$\iff \frac{\partial \vartheta}{\partial S_2} = \frac{\vartheta}{S_2} \frac{1}{(1 + \frac{S_2}{S_1} \frac{[1-V(s)+V^1(s)]}{[1-V^1(s)]} \frac{1}{\vartheta})} > 0$$

Finally, I must show that $v^1/v^2(s)$ is an increasing function. Write

$$v^1/v^2(s) = \frac{(1 - V^1(s)) e^{-\theta^1(1-V^1(s))}}{(1 - V^2(s)) e^{-\theta^2(1-V^2(s))}} = \frac{(1 - V^1(s))}{(1 - V^2(s))} exp(\frac{\theta^2}{\theta^1} \frac{(1 - V^2(s))}{(1 - V^1(s))})$$

Let us define $g(s) = \frac{(1-V^1(s))}{(1-V^2(s))}$, then

$$v^1/v^2(s) = g(s) exp(\frac{\theta^2}{\theta^1} \frac{1}{g(s)})$$

Thus,

$$\frac{\partial (v^1/v^2)(s)}{\partial s} = g'(s) exp(\frac{\theta^2}{\theta^1} \frac{1}{g(s)}) [1 - \frac{\theta^2}{\theta^1} \frac{1}{g(s)}]$$

Now, I must show that $[1 - \frac{\theta^2}{\theta^1} \frac{1}{g(s)}] > 0 \Leftrightarrow \frac{\theta^2}{\theta^1} \frac{1}{g(s)} < 1 \Leftrightarrow \frac{\theta^2}{\theta^1} < g(s)$

Since $\theta^1 > \theta^2$, $\frac{\theta^2}{\theta^1} < 1$

Finally, I must show that $g(s)$ is an increasing function (since $g(0) = 1$), if $g(s)$ in an increasing function then $\frac{\theta^2}{\theta^1} < 1 \leq g(s)$ will hold for every $1 > s > 0$

Notice that
$$g'(s) = \frac{-v^1(s)(1 - V^2(s)) + v^2(s)(1 - V^1(s))}{(1 - V^2(s))^2}$$

$$= \frac{v(s)g(s)}{[1 - V^1(s)]e^{-\theta^1(1-V^1(s))} + [1 - V^2(s)]e^{-\theta^2(1-V^2(s))}}[e^{-\theta^2(1-V^2(s))} - e^{-\theta^1(1-V^1(s))}]$$

Therefore, I must show again that

$$e^{-\theta^2(1-V^1(s))} - e^{-\theta^1(1-V^1(s))} > 0 \Leftrightarrow \frac{\theta^2}{\theta^1} < 1 < \frac{(1 - V^1(s))}{(1 - V^2(s))}$$

This is equivalent to showing that

$$V^1(s) < V^2(s) \quad \forall s$$

If I can show that
$$v^1(s) < v^2(s) \quad \forall s$$

then the above will also hold.

Now, let us show that $v^1(0) < v^2(0) \Leftrightarrow v^1(0) - v^2(0) < 0 \Leftrightarrow e^{-\theta^1} < e^{-\theta^2}$ and since $\theta^1 > \theta^2$ this will be true.

Second, I must also show that

$$\partial v^1 / \partial v^2 \leq 1$$

(The intuition is that if v^1 does not grow faster than v^2, then it will always be the case that $v^1(s) < v^2(s) \quad \forall s$)

Now, after some algebraic manipulation, I can show that $\partial v^1 / \partial v^2 = 1$ Thus the result will hold. □

Finally, from Proposition 2, I have shown that the skill distribution in City 1 is more skill-abundant than in City 2 (i.e., the monotone likelihood ratio property holds). The rest of the proof for Proposition 3 directly follows from Costinot Voguel 2009 Lemma 3.

A4. Lumosity Data: robustness check

As a robustness check, I employ a novel data set that can provide explicit measurements of talent for a large sample of workers in different occupations and in different cities across the US. I use data from a large online brain training and neuroscience research company which offers a brain training service that consists of different games in the areas of memory, attention, flexibility, speed of processing, numerical problem solving, and verbal fluency. For a sample of almost 90,000 users across 54 MSAs, the company provides information of performance in 10 different games that could be used as a measurement of talent or ability in different cognitive dimensions. For each user, I also have their occupation detail, along with some baseline characteristics including age, gender, years of completed education, and MSA. I aggregate the performance in each separate task into an overall cognitive ability index, whose distribution I will use to alternative target different moments of the talent distribution. It is important to note that I must move away from other alternative measures of talent (such as the AFQT in NLSY79), because of sample size restrictions, which are binding once we aim to characterize the distribution of skills at the MSA level[1].

One important drawback of using data from an online app is that it is not necessarily a representative sample of the population. In order to deal with the nonrandom nature of our sample, I use a weighting method so that, for each MSA, I match the first and second moments of the observables in our sample to the equivalent moments observed for each MSA in the CPS[2]. By doing this, I generate weighted moments for the skill distribution that are representative of the underlying population[3].

The estimates for the taste for amenities parameter, using either years of completed education from the CPS or Lumosity data, are very consistent among each other. For consistency, I present the results using the CPS.

[1]Nevertheless, I check that our alternative talent measurements are consistent with other sources of ability measures which have been previously used in the literature, such as AFQT. I do this by occupation (given sample restrictions). For each occupation, we compare residuals of a regression that uses either the AFQT score or the game score as the dependent variable and includes education, gender, and age as explanatory variables. At an aggregate occupation level, both measures of game scores and AFQT seem to provide similar distributions of unobserved talent, which is captured by the residuals in the regression.

[2]I target the mean and variance for gender, age and years of completed education for each city in our sample.

[3]I use a method called *stable weights* that balances covariates for estimation with incomplete outcome data. This new methodology provides a new weighting approach that finds the weights of minimum variance that adjust or balance the empirical distribution of the observed covariates up to levels prospected by us. This method allows me to balance very precisely the means of the observed covariates and other features of their marginal and joint distributions, such as variances. I apply this procedure for each MSA, generating the minimum variance weights for each city. For more details, see Zubizarreta (2015).

Appendix B

Appendix Chapter 3

Appendix A

Proof of Lemma 3.1

Let x' be the firm's belief about the other firm's private signal. This belief is a random variable distributed $N(\bar{\theta}, \sqrt{\frac{2\sigma^2\tau^2+\sigma^4}{\sigma^2+\tau^2}})$. Consequently, both firms will invest if

$$\bar{\theta} = \frac{\sigma^2 y + \tau^2 x'}{\sigma^2 + \tau^2} \geq \kappa,$$

or equivalently

$$x' \geq \kappa + (\sigma^2/\tau^2)(\kappa - y). \tag{A.1}$$

The probability of investment is given by the cumulative probability of x' being higher than $\kappa + (\sigma^2/\tau^2)(\kappa - y)$, that is

$$P(I) = 1 - \Phi\left\{\frac{\kappa + (\sigma^2/\tau^2)(\kappa - y) - \bar{\theta}}{\sqrt{\frac{2\sigma^2\tau^2+\sigma^4}{\sigma^2+\tau^2}}}\right\}. \tag{A.2}$$

Given that firms have linear payoffs, their expected utilities are equal to

$$v(\bar{\theta},\kappa) = \bar{\theta} + 1\left[1 - \Phi\left\{\frac{\kappa + (\sigma^2/\tau^2)(\kappa - y) - \bar{\theta}}{\sqrt{\frac{2\sigma^2\tau^2+\sigma^4}{\sigma^2+\tau^2}}}\right\}\right] - 1$$
$$= \bar{\theta} + -\Phi\left\{\frac{\kappa + (\sigma^2/\tau^2)(\kappa - y) - \bar{\theta}}{\sqrt{\frac{2\sigma^2\tau^2+\sigma^4}{\sigma^2+\tau^2}}}\right\}, \tag{A.3}$$

and since the payoffs of *Not invest* are 0, the threshold κ is such that equation A.3 is equal to zero:

$$v(\kappa,\kappa) = \kappa + -\Phi\left\{\frac{(\sigma^2/\tau^2)(\kappa - y)}{\sqrt{\frac{2\sigma^2\tau^2+\sigma^4}{\sigma^2+\tau^2}}}\right\} = 0. \tag{A.4}$$

Finally, defining $\gamma \equiv \frac{(\sigma^2/\tau^2)}{\sqrt{\frac{2\sigma^2\tau^2+\sigma^4}{\sigma^2+\tau^2}}}$, we obtain the following implicit equation that characterizes the equilibrium:

$$\kappa + -\Phi\left\{\gamma(\kappa - y)\right\} = 0. \tag{A.5}$$

Proof of Proposition 3.2

When $\sigma \to 0$, the information structure is similar to the one in the nonlimit case, but now

$$\bar{\theta} = \frac{\sigma^2 y + \tau^2 x}{\sigma^2 + \tau^2} \to x. \tag{A.6}$$

As in the nonlimit case, solving the general game for a continuum of players is equivalent to solving the two-player game. Since each firm observes an independent private signal x, they will infer the other firm's private signal $x' \backsim N(x, 2\sigma^2)$. The *switching strategies* are

$$s(x) = \begin{cases} I & if \quad x > \kappa \\ NI & if \quad x \leq \kappa, \end{cases} \tag{A.7}$$

and therefore the probability that the other firm observes a signal less than κ will be given by

$$\Phi\left(\frac{k-x}{\sqrt{2}\sigma}\right), \tag{A.8}$$

so each firm's expected pay-off is

$$x-\Phi\left(\frac{k-x}{\sqrt{2}\sigma}\right). \tag{A.9}$$

By Iterated Elimination of Strictly Dominated Strategies (see Carlsson and Van Damme 43) the unique equilibrium of this game is one where both firms will invest only if they observe a private signal equal or greater than $1/2$, and thus $dI(y)=0$. Consequently, the gain in coordination in this case is the totality of the miscoordination from the nonlimit case. Thus, the gain in coordination is an increasing function of τ.

Appendix B

Table B.B.1: Firm summary statistics

	Mean	SD	p25	p50	p75
Log(Assets)	3.23	2.76	1.60	3.44	5.09
ROA	-0.13	0.30	-0.20	0.01	0.07
Log(Size)	3.28	2.57	1.50	3.38	5.08
Market leverage	0.16	0.19	0.00	0.09	0.25
Investment	0.34	22.70	0.01	0.03	0.07
R&D	7.4	13.1	0.0	0.0	7.4

This table describes the 9,167 firms in the main sample. The variables Assets, Size (Market Cap), and R&D are denominated in Millions.

Table B.B.2: Robustness for Table 3.3

	log(distance)		Within 20 mi	
RC index$_{combined}$	-0.2921*	-0.2760	0.0753***	0.0642**
	(0.1691)	(0.1885)	(0.0275)	(0.0311)
log(Sales)$_{customer}$		0.0785***		-0.0125**
		(0.0287)		(0.0051)
log(Sales)$_{supplier}$		-0.0458**		0.0036
		(0.0194)		(0.0029)
Number of firms$_{customer}$		0.0375		-0.0075
		(0.0344)		(0.0059)
Number of firms$_{supplier}$		0.0251		-0.0009
		(0.0343)		(0.0054)
R^2	0.0426	0.0503	0.0671	0.0734
N	2,323	2,323	2,323	2,323

This table repeats the estimations in Table 3.3, with the only difference being that the RC index of the customer and the RC index of the supplier are averaged. Standard errors clustered at the Fama–French 48 industry level and are reported in parentheses, below the coefficient estimates.