

# Spillovers, Absorptive Capacity and Agglomeration\*

Sergey Lychagin<sup>†</sup>

Central European University

Department of Economics

March 16, 2012

## Abstract

I study knowledge spillovers in an industry where firms are mobile and heterogeneous in their ability to adopt outside knowledge (*absorptive capacity*). I develop a static model of industry agglomeration where, in equilibrium, the force of attraction induced by spillovers is counteracted by the force of repulsion created by local competition. The model is applied to a sample of the US software firms. I estimate the structural parameters of the model and obtain the following results: (a) The data are consistent with highly localized knowledge spillovers; (b) The attraction force induced by spillovers creates a significant sorting pattern placing firms with higher absorptive capacity in more agglomerated counties; (c) Ignoring firm heterogeneity in absorptive capacity leads to substantially biased estimates of gains from spillovers in policy experiments.

---

\*This paper is based upon a chapter of my Penn State University Ph.D. thesis. I am grateful to Kala Krishna and Joris Pinkse for their invaluable advice and support. I also would like to thank Saroj Bhattarai, Edward Coulson, Jonathan Eaton, Mark Roberts, Andrés Rodríguez-Clare, Lily Samkharadze, Alexander Tarasov, Neil Wallace and Stephen Yeaple whose comments contributed to this paper. All remaining errors are mine.

<sup>†</sup>Email: lychagins@ceu.hu

# 1 Introduction

Knowledge spillovers lie at the heart of many economic theories. In models of endogenous growth, cross-firm spillovers are essential in creating increasing returns to scale (Romer (1986)). In urban and regional economic theories, geographically localized spillovers are used to explain why economic activity tends to be densely concentrated in space (Glaeser (1999)). In Ricardian models of international trade (e.g., Eaton and Kortum (2001)), the lack of perfect cross-country spillovers is instrumental in generating trade flows. In the development literature, localized spillovers are the source of persistent gaps in productivity across countries (e.g., Feenstra (1996)). In the above theories, it is crucial to know the scope of knowledge spillovers and the magnitude of their economic impact.

Quantifying knowledge spillovers is a difficult task as they are almost never directly observed. A usual approach is to correlate each firm's knowledge-generating activity to the performance of its neighbors, assuming that the former causes the latter. However, this approach may be problematic if spillovers affect firms differentially, i.e., if firms are heterogeneous in *absorptive capacity*. An advanced technology firm built in a rural area will not bring any benefits to its geographic neighbors, as they are likely to produce very basic and simple products and not rely on the frontier technology. The same firm located in a megapolis with hundreds of advanced competitors who are eager to find and adopt latest inventions is likely to create a lot of positive externalities. To understand how a spillover from location A affects firms at location B one has to answer two questions: What is the spillover's impact on a firm at B, given this firm's absorptive capacity? What are the absorptive capacities of firms attracted to B?

To create a framework for addressing these questions, I construct a structural model of location choice in the presence of R&D spillovers, assuming unobserved heterogeneity of firms in absorptive capacity. I demonstrate that, in equilibrium, firms that are more responsive to spillovers tend to be over-represented in agglomerated locations. I show that this spatial sorting pattern can be used to identify the model. Then, I apply the model to data on production and locations of firms in the US software industry. I find evidence that spillovers within this industry exist and are highly localized in space. I demonstrate that the spatial sorting of firms by absorptive capacity produces substantial differences in the economic impact of spillovers across geographic locations.

Location choice is modeled as a static two-period game. There is a fixed mass of firms, which simultaneously choose locations in the first period. In the second period, which can

be thought of as consisting of several years, the firms produce varieties of a final good; the firms cannot relocate or exit during this time. The firms have innate differences in absorptive capacity and R&D stocks, which they are endowed with at the time of birth. R&D stocks generate spillovers that decay with distance, which is consistent with a large literature including Jaffe, Trajtenberg and Henderson (1993), Greenstone, Hornbeck and Moretti (2010), Lychagin et al. (2010), to name just a few examples. Locations differ in wages, capital rental rates and consumer demand; these location characteristics are exogenous.

In equilibrium, the location choice and hence the spatial distribution of firms is shaped by three forces. First, there is an agglomerative force induced by spillovers. Since firms vary in absorptive capacity, they respond differentially to this force: a firm who benefits more from spillovers is drawn more strongly to agglomerated regions, even if these regions are expensive and have tough competition. Second, there is a counteracting centrifugal force induced by local competition in the final goods market. Finally, there is a dispersion force caused by idiosyncratic location preferences of firm owners and managers.

Since the force induced by spillovers acts differentially on firms, it creates sorting. Identification of the model's main parameters relies on detecting the magnitude of this sorting pattern in the joint distribution of firm locations and absorptive capacities. While absorptive capacities are not directly observed, their distribution can be inferred from a firm-level panel dataset on the production occurring in the second period of the game. The temporal dimension of the data permits the estimation of each firm's absorptive capacity by correlating the variations in the firm's total factor productivity to the variations in the R&D stocks of its geographic neighbors. Although these firm-level estimates would be inaccurate in the short panel setting, they can be used to precisely identify the joint density of firm absorptive capacities and locations, if the number of firms in the sample grows to infinity. This density is then fitted to the predictions of the location choice model in order to identify the model's key parameters.

In the empirical application, I focus on the US software industry. Software firms are highly agglomerated around Silicon Valley in California and Boston, Massachusetts. It is commonly believed that knowledge spillovers are partly responsible for this extreme agglomeration. There is a body of anecdotal evidence that spillovers play an important part in software development and that they are highly localized in space (e.g., Saxenian (1996) provides a number of supporting stories). There is also evidence that the software industry features significant firm heterogeneity in absorptive capacity (Matusik and Heely (2005)). The aim of my empirical application is twofold: first, to determine how the impact and the geographic

scope of spillovers are affected by the firm heterogeneity, and second, to quantify an overall agglomerative force induced by spillovers.

The estimation results suggest that spillovers in the software industry are highly localized: if the receiving firm's absorptive capacity is kept fixed, the spillover's productivity effect declines by half at a distance of 59 kilometers from the spillover's origin. I also find that the pattern of spatial sorting by absorptive capacity is statistically significant. Spatial sorting visibly distorts the scope of spillovers. For instance, simulations of the model show that a spillover from San Jose, the center of Silicon Valley, has an impact in Alameda county almost two times larger than in the less populated Santa Cruz county, although both counties are approximately at the same distance from San Jose. Such difference in the effects of spillovers takes place because Alameda firms have higher absorptive capacity.

By modeling knowledge spillovers and location choice in one setting, this paper brings together the literature on knowledge production function, which finds its roots in the work of Zvi Griliches, and the new economic geography literature pioneered by Krugman (1991). This paper is also in line with an emerging interest within the urban and regional economic literature in studying the consequences of the firm- and individual-level heterogeneity on the spatial distribution of economic activity. Finally, my empirical approach is not only applicable in the context of knowledge spillovers. It can be used to identify any other agglomerative force affecting firm locations, for example, forces associated with labor market pooling, or input sharing.

The literature on the knowledge production function relates each firm's productivity to the firm's own R&D and spillovers from other firms. The effect and the properties of spillovers are inferred by estimating a firm-level production function that includes R&D stocks of geographic neighbors in the firm's total factor productivity term. The most recent examples of papers that employ this technique include Orlando (2004), who studies spillovers within a broad US industry defined by SIC 35, and Lychagin et al. (2010), who estimate semi-parametrically how quickly spillovers decline with distance from the origin. Another related paper (Keller (2002)) uses the same approach to study country-level knowledge diffusion. All these works find evidence that spillovers exist, have positive impact on productivity, and are geographically localized. They all assume that firms (countries, in the context of Keller) are identical in their absorption of spillovers, and ignore the endogeneity of location choice. My paper contributes to this literature by relaxing these two assumptions.

The new economic geography literature provides me with an analytically tractable framework for studying agglomeration: a "footloose capital" model by Martin and Rogers (1995).

I modify their model by introducing distance-mediated knowledge spillovers, which create an agglomeration force, idiosyncratic location preferences, which ensure that this force does not generate “black hole” cities, and, most importantly, firm heterogeneity in absorptive capacity. I contribute to this literature by showing that the *unobserved* heterogeneity in absorptive capacity can be used to identify the agglomeration force induced by R&D spillovers, if one has suitable panel data (in my case, balance sheet items) to obtain the joint distribution of firm locations and absorptive capacity.

This paper is also related to a new strand of literature on urban and regional economics that focuses on spatial sorting of heterogeneous firms and individuals across cities, and aims at distinguishing the effect of sorting from the benefits of agglomeration. For example, Baldwin and Okubo (2006) construct a model of reallocation that introduces costly migration into the Melitz framework and shows that the more productive firms are the first to leave their home regions to serve the biggest market. Combes et al. (2009) show that agglomeration economies and firm selection have different effects on the distribution of firm productivity within cities, which helps to identify the relative contributions of these factors into the cross-city differences in average productivity. Behrens, Duranton and Robert-Nicoud (2010) study a model of location decisions by heterogeneous workers, which predicts that the more productive workers choose to be in the bigger cities. Unlike these theories, which focus on sorting by innate productivity, I construct the model to focus on sorting by absorptive capacity. My identification technique provides a new way of distinguishing the effect of agglomeration from the effect of sorting.

Finally, the empirical approach developed in my paper is not restricted to spillovers; it can be applied to infer the strength of any other agglomeration force. To date, the literature that aimed at decomposing the contributions of various forces into the spatial distribution of economic activity relied on reduced form equations where observations represent industries. A usual approach was to relate some measure of geographic concentration in a given industry to a set of proxy variables that indicate the importance of different agglomeration forces. Examples of such work include Audretsch and Feldman (1996), Rosenthal and Strange (2001), Ellison, Glaeser and Kerr (2010).

The next section of the paper lays out a model of location choice in the presence of spillovers that can be taken to the data. Section 3 links the model tightly to estimation and shows that its key elements are identified from firm-level data on locations and balance sheet items. Section 4 then proposes an estimation method. This method is applied to data on the US software industry in Section 5. Finally, Section 6 summarizes main findings.

## 2 A Model of Location Choice

There are  $L$  locations (cities) indexed by either  $l$  or  $m$ ;  $l, m \in \{1, \dots, L\}$ . Cities are populated by immobile consumers with CES preferences defined on all available varieties of a final good. Consumers in city  $m$  spend an exogenous amount  $E_m$  on the final good, which captures the size of local demand. There is a unit mass of infinitely small firms indexed by  $i \in \Omega$  that are free to choose their locations. Each firm produces one distinct variety of the final good; the varieties can be shipped to any city.

When the firm is born, it has a stock of R&D ( $R_i$ ). To simplify the analysis,  $R_i$  is assumed to be exogenous. A part of this R&D stock may spill over to other firms, increasing their profitability. Firms differ in their innate ability to adopt and commercialize spillovers; this is captured by a firm specific parameter  $\alpha_i$  (absorptive capacity).<sup>1</sup> As shown later in this section, different locations provide different advantages to the firm: they may vary in the level of wages, the degree of local competition, or proximity to firms generating high spillovers. Aside from these factors, the firm's location preferences are also affected by an idiosyncratic component  $\varepsilon_{il}$ . This component captures monetary payoffs (e.g., the firm may be able to secure location incentives from officials at certain places), as well as non-monetary benefits (e.g., a CEO of the firm may prefer the city where he was born) received by the firm at location  $l$ . Finally, the firm is characterized by an idiosyncratic productivity shock  $\varphi_i$ . In total, firms are heterogeneous in  $L + 3$  dimensions:  $R_i$ ,  $\alpha_i$ ,  $\varphi_i$ , and  $\{\varepsilon_{il}\}_{l=1}^L$ .

### 2.1 Timing

Location choice and production are modeled as a Nash equilibrium in a static game with the following sequence of events:

**Period 0** A unit mass of firms is born. At the time of birth, each firm is endowed with a vector of characteristics:  $[\alpha_i, R_i, \varphi_i, \varepsilon_{i1}, \dots, \varepsilon_{iL}]$ .

**Period 1** Firms simultaneously choose locations. A firm is allowed to have only one location.

---

<sup>1</sup>An influential strand of literature originating from Cohen and Levinthal (1989) suggests that the firm's R&D effort and absorptive capacity are endogenous and closely interrelated. By investing into R&D, the firm improves its awareness of useful knowledge available for adoption, hence increasing its absorptive capacity. Although in principle this mechanism could be incorporated in the model, it is shut down to keep the identification argument as simple as possible.

**Period 2** Given locations, firms set prices in each city, produce output, and ship it to the consumers.

## 2.2 Demand

Each city is inhabited by consumers with CES preferences:

$$U_m = \left( \int_{i \in \Omega} Q_{im}^{(\sigma-1)/\sigma} di \right)^{\sigma/(\sigma-1)}, \quad \sigma > 1$$

where index  $i$  is used to denote a variety produced by firm  $i$ .

Given that consumers in city  $m$  spend an exogenous amount  $E_m$  on the final good, one can easily find the demand for each variety:

$$Q_{im}(p_{im}) = \frac{E_m}{P_m} \left( \frac{p_{im}}{P_m} \right)^{-\sigma} \quad (1)$$

The demand depends on  $P_m$ , a location-specific price index defined in a standard way:

$$P_m = \left( \int_{i \in \Omega} p_{im}^{1-\sigma} di \right)^{1/(1-\sigma)}$$

## 2.3 Supply

There is a unit mass of infinitely small single-variety single-location firms. The firms' profits are location specific; they are derived below from equilibrium conditions for the final goods market in each city.

Firms produce output using capital and labor in a constant returns to scale Cobb-Douglas technology with a Hicks-neutral total factor productivity (TFP) term  $A_i$ :

$$Q_i = A_i K_i^{\beta_k} L_i^{\beta_l}$$

Total factor productivity depends on three components: own stock of R&D,  $R_i$ , spillovers from other firms' R&D received at location  $l$ ,  $S_l$ , and the firm-specific productivity shock,  $\varphi_i$ :

$$A_i = R_i^{\beta_r} S_l^{\alpha_i} \varphi_i$$

where the parameter  $\alpha_i$  captures the absorptive capacity of firm  $i$ .<sup>2</sup>

---

<sup>2</sup>Firms in a single narrowly defined industry can differ in absorptive capacity for a number of reasons.

Spillovers originate from stocks of R&D; their intensity declines with distance from their origin<sup>3</sup>:

$$S_l = \sum_{m=1}^L e^{-\lambda\rho(l,m)}\mathcal{R}_m, \quad \mathcal{R}_m = \int_{\substack{i \in \Omega \\ l_i=m}} R_i di \quad (2)$$

where  $m$  indexes cities,  $l_i$  is the location of firm  $i$ ,  $\rho(\cdot, \cdot)$  is the geographic distance.

The constant returns to scale production technology implies that the marginal cost function is constant and has the following form:

$$mc_{il} = \frac{c(w_l, r_l)}{R_i^{\beta_r} S_l^{\alpha_i} \varphi_i}$$

Apart from the productivity term in the denominator, the marginal cost depends on  $w_l$  and  $r_l$ , the location-specific factor prices. Both  $w_l$  and  $r_l$  are assumed to be exogenous.

In addition to the cost of production, firms face a cost of transportation<sup>4</sup>. The value of a good produced in city  $l$  declines with distance to the city where the good is consumed. This is captured by the iceberg cost  $\tau_{lm}$ . Producer has to ship  $\tau_{lm}$  units of the good from the origin city  $l$ , in order to have one unit arrive to destination  $m$ .

---

For example, developing some products entails an inherent uncertainty in the type of knowledge needed to complete the development project and to tailor the product to the needs of consumers. Firms designing such products have to rely on the outside knowledge more frequently. Thus, they receive more benefits from being exposed to spillovers. Other determinants of absorptive capacity are closely studied in the management literature. One example studying specifically the US software industry is Matusik and Heely (2005).

<sup>3</sup>The decay of spillovers with distance reflects the costs of communication. Although modern technologies make phone calls and electronic correspondence very cheap, some part of knowledge cannot be efficiently transmitted by means other than face-to-face contact. In addition, geographic proximity provides more opportunities for networking: new research ideas can be informally exchanged during an occasional lunch with colleagues from other firms.

<sup>4</sup>In the software industry, transportation of the final product is virtually costless. At the same time, software products are often accompanied by support and services that involve interactions between customers and developers. Spatial proximity facilitates these interactions by reducing the costs of face-to-face contact and thus increasing the total value of the product for consumers. In my sample, 70% of firms are involved in activities other than software development; in particular, 41% of firms are active in computer-related services. The average revenues from these activities constitute respectively 22% and 10% of total revenues.



## 2.4 Payoffs to firms

Profit of firm  $i$  located in city  $l$  earned from sales at market  $m$  equals the price-cost markup multiplied by the quantity sold:

$$\Pi_{ilm} = \max_{p_{ilm}} \left( p_{ilm} - \tau_{lm} \frac{c(w_l, r_l)}{R_i^{\beta_r} S_l^{\alpha_i} \varphi_i} \right) Q_{im}(p_{ilm})$$

The firm chooses its profit-maximizing price  $p_{ilm}$ , taking the R&D spillovers and the price index as given, where

$$p_{ilm} = \frac{\sigma}{\sigma - 1} \tau_{lm} \frac{c(w_l, r_l)}{R_i^{\beta_r} S_l^{\alpha_i} \varphi_i} \quad (3)$$

Substituting the price back to the profit equation yields an expression for the maximized value of profits:

$$\begin{aligned} \Pi_{ilm} &= \frac{E_m P_m^{\sigma-1}}{\sigma} \left( \frac{\sigma - 1}{\sigma} \frac{R_i^{\beta_r} S_l^{\alpha_i} \varphi_i}{c(w_l, r_l) \tau_{lm}} \right)^{\sigma-1} \\ &= \frac{1}{\sigma} \left( \frac{\sigma - 1}{\sigma} \frac{R_i^{\beta_r} S_l^{\alpha_i} \varphi_i}{c(w_l, r_l)} \right)^{\sigma-1} E_m \left( \frac{P_m}{\tau_{lm}} \right)^{\sigma-1} \end{aligned}$$

By summing firm  $i$ 's profits across all destinations, one obtains the total profit of the firm, given that the firm is located in city  $l$ :

$$\Pi_{il} = \frac{1}{\sigma} \left( \frac{\sigma - 1}{\sigma} \frac{R_i^{\beta_r} S_l^{\alpha_i} \varphi_i}{c(w_l, r_l)} \right)^{\sigma-1} \sum_m E_m \left( \frac{P_m}{\tau_{lm}} \right)^{\sigma-1} \quad (4)$$

The price index in market  $m$  is found by integrating over the price set by firms:

$$P_m = \frac{\sigma}{\sigma - 1} \left[ \int_{j \in \Omega} \left( \frac{R_j^{\beta_r} S_{l_j}^{\alpha_j} \varphi_j}{c(w_{l_j}, r_{l_j}) \tau_{l_j m}} \right)^{\sigma-1} dj \right]^{-\frac{1}{\sigma-1}} \quad (5)$$

Note that the R&D stock of firm  $j$  affects firm  $i$ 's profit via two channels: spillovers and prices.

First,  $R_j$  affects firm  $i$ 's profits positively by increasing spillovers term  $S_l$  in equation (4). The magnitude of this effect depends on the firm specific absorptive capacity,  $\alpha_i$ , and thus varies across firms.

Second, an increase in  $R_j$  makes firm  $j$  and its neighbors more productive, which drives down the price indices in all cities. Firm  $j$ 's R&D stock has a direct and an indirect effect on price. If  $R_j$  is increased, firm  $j$  becomes more productive (direct effect). Part of  $R_j$  spills over to other competitors of firm  $i$  making them more productive, too (indirect effect). Both effects put a downward pressure on the price indices in all markets and reduce profits of firm  $i$ . The magnitude of the combined effect is the same for all firms at a given location. It declines with distance from the location of firm  $j$ , due to increasing transport costs and the decay of spillovers with distance.

## 2.5 Location choice

Before starting production and earning profits, each firm has to settle at some location. Location preferences of firms have two components: the profit,  $\Pi_{il}$ , and the idiosyncratic preference shock, drawn at the time of the firm's birth,  $\{\varepsilon_{il}\}_{l=1}^L$ . Preferences  $\Pi_{il}^*$  are defined as a product of  $\Pi_{il}$  and  $e^{\varepsilon_{il}}$ . Firms can be thought of as maximizing a monotone transformation of  $\Pi_{il}^*$ ,

$$\pi_{il}^* = \log \Pi_{il} + \varepsilon_{il}$$

After substituting the expression for profit from (4), one obtains

$$\begin{aligned} \pi_{il}^* = \log \left[ \frac{1}{\sigma} \left( \frac{\sigma-1}{\sigma} R_i^{\beta_r} \varphi_i \right)^{\sigma-1} \right] + \log \left[ \sum_m E_m \left( \frac{P_m}{\tau_{lm}} \right)^{\sigma-1} \right] \\ - (\sigma-1) \log c(w_l, r_l) + (\sigma-1) \alpha_i \log S_l + \varepsilon_{il} \end{aligned}$$

To simplify the algebra, I assume that location preference shocks  $\varepsilon_{il}$  have a type I extreme value distribution with a constant but arbitrary scale parameter  $b$  and are independent across firms and locations. Then, I rescale the payoff equation so that the mean and the variance of the rescaled preference shocks satisfy the usual assumptions of the multinomial logit model.

The rescaling parameter is  $1/b$ , which gives

$$\underbrace{\frac{\pi_{il}^*}{b}}_{\tilde{\pi}_{il}^*} = \underbrace{\frac{1}{b} \log \left[ \frac{1}{\sigma} \left( \frac{\sigma-1}{\sigma} R_i^{\beta_r} \varphi_i \right)^{\sigma-1} \right]}_{f_i} + \underbrace{\frac{1}{b} \log \left[ c(w_l, r_l)^{1-\sigma} \sum_m E_m \left( \frac{P_m}{\tau_{lm}} \right)^{\sigma-1} \right]}_{a_l} + \underbrace{\frac{(\sigma-1)}{b} \alpha_i \log S_l}_{\gamma} + \underbrace{\frac{\varepsilon_{il}}{b}}_{\tilde{\varepsilon}_{il}}$$

The payoff variable and the idiosyncratic component are relabeled as  $\tilde{\pi}_{il}^*$  and  $\tilde{\varepsilon}_{il}$ . The constant coefficient of the spillover term is denoted as  $\gamma$ . The remaining parts of the equation are collected under a firm specific term  $f_i$  and a location specific term  $a_l$ :

$$\tilde{\pi}_{il}^* = f_i + a_l + \gamma \alpha_i \ln S_l + \tilde{\varepsilon}_{il}$$

Firm  $i$  chooses a location that maximizes the firm's payoff:

$$l_i = \arg \max_l (f_i + a_l + \gamma \alpha_i \ln S_l + \tilde{\varepsilon}_{il}) \quad (6)$$

Location choice is driven by three forces. First, there are location-specific factors that are common to all firms, captured by the location fixed effect,  $a_l$ . These factors include proximity to consumers, prices in the final good markets, and the level of local input costs. Second, there are knowledge spillovers that the firm expects to receive at location  $l$ . Spillovers affect firms differentially: firms with higher values of  $\alpha_i$  are drawn stronger to high-spillover locations. Parameter  $\gamma$  determines the overall strength of this force. Finally, there is a random shock  $\tilde{\varepsilon}$ , which reflects the idiosyncratic part of the location preferences. The firm specific term  $f_i$  does not affect location choice.

## 2.6 Equilibrium

An equilibrium of the game is characterized by

1. Locations of all firms,  $\{l_i\}_{i \in \Omega}$
2. Pricing decisions of every firm in every market  $m$ ,  $\{p_{ilm}\}_{l,m=1,i \in \Omega}^L$ , given *an arbitrary* own location  $l$  (this includes off-equilibrium locations)

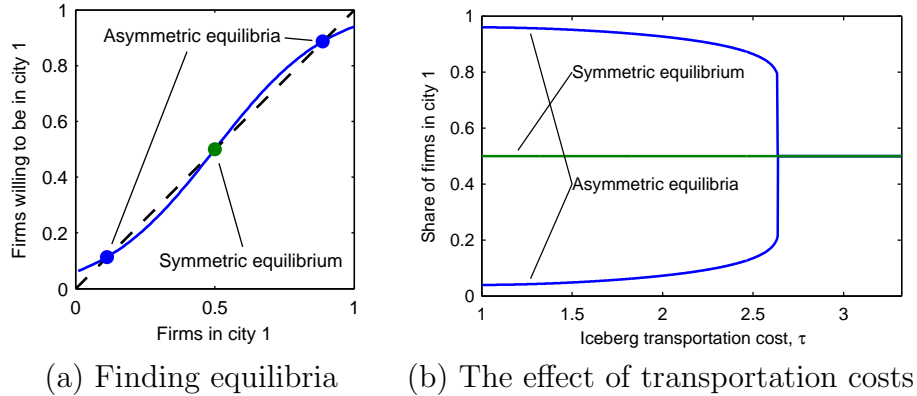


Figure 1: Multiple equilibria in the two-city model

Local wages, rental rates, and consumer expenditures are exogenously given. In principle, labor and capital markets could be explicitly modeled here in exactly the same way as the final goods market. However, I choose not to do so to keep the model simple.

In equilibrium, each firm chooses the location and the prices that maximize the firm's payoffs, given actions of other firms. This defines a best response mapping from the set of location and pricing decisions of all firms into itself. An equilibrium, by definition, is a fixed point of this best response mapping.

The equilibrium is guaranteed to exist; the proof using the Brouwer fixed point theorem can be found in the Appendix. However, it is not always unique. Uniqueness tends to occur if firms are repelled from each other strongly enough. For example, Seim (2006) shows that in a special case of her spatial competition model the equilibrium is unique. However, if the dispersion force induced by competition becomes dominated by the agglomeration force induced by spillovers, multiple equilibria may arise.

To illustrate, I set up and solve a simple version of the model with two symmetric cities, assuming that all firms have the same absorptive capacities and R&D stocks ( $\alpha_i = 1$ ,  $R_i = 1$ ). Given  $\tau$ , the iceberg cost parameter, all equilibria in the game are found as follows. Let  $n_1$  be a share of firms located in city 1. Since R&D stocks are deterministic and uniform, one can find the aggregate R&D stock and the spillovers for both cities,  $R_l$  and  $S_l$ . The spatial distribution of firms,  $\{n_l\}_{l=1}^L$ , and spillovers,  $\{S_l\}_{l=1}^L$ , determine price indices in equation (5), which are used to obtain location fixed effects  $a_l$ . Given  $a_l$  and  $S_l$ , one can solve the location choice problem (6) and find a share of firms that are willing to be located in city 1. This share is depicted in Figure 1a as a solid curve. In equilibrium, it should be equal to  $n_1$ ; that is, the solid curve should intersect the 45 degree line.

The equilibria are found for a number of values of the iceberg cost parameter. The results are depicted in Figure 1b. When the transportation cost is high, competition in the final goods market is highly localized. Hence, firms try to avoid locating in cities with many competitors, and agglomerations do not arise. As the transportation cost falls, it is no longer possible to escape competition in less populated cities. The benefits of receiving more spillovers from geographic neighbors outweigh the negative effect of competition on profits. This gives rise to equilibria with asymmetric distributions of firms across cities. The idiosyncratic shock in the location preferences ensures that every city is always populated by a non-zero share of firms.

## 2.7 The sorting property

All equilibria of the location choice model have one important property that is used later in the empirical exercise to test whether the model’s implications are consistent with data.

Firms are heterogeneous in the benefits they draw from spillovers. Consequently, firms with high absorptive capacity,  $\alpha_i$ , are attracted to high-spillover locations relatively stronger than firms with low  $\alpha_i$ . This creates spatial sorting of firms by their absorptive capacity.

**Proposition 1.** *Consider any equilibrium of the model. Let  $l_1$  and  $l_2$  be two locations such that  $S_{l_2} > S_{l_1}$  in this equilibrium. In equilibrium, the firms are spatially sorted by  $\alpha$ . In fact, the distribution of firms’ absorptive capacities at location  $l_2$  stochastically dominates the distribution at  $l_1$ :*

$$F(\alpha|l_2) < F(\alpha|l_1)$$

*Proof.* See Appendix.

In the empirical exercise, it is found that the above sorting pattern does exist in data: firms with high absorptive capacities tend to be overrepresented in high-spillover cities.

## 3 Identification

This section shows how to identify the key elements of the model that affect location choice and the geographic scope of spillovers. It contains the two main innovations of the paper.

First, it is shown that firm-level data on production can be used to identify the pattern of firm heterogeneity in absorptive capacity and find how this pattern varies across firms in geographic space. Unlike the related studies in the agglomeration literature that use proxies,

this paper infers absorptive capacity of individual firms directly from fluctuations in firm productivity attributed to spillovers.

Second, it is demonstrated that the spatial pattern of absorptive capacities found above can be used to identify the location choice problem. Identification does not require the econometrician to observe all factors affecting firm location, such as local wages or consumer demand.

### 3.1 Production function

First, it is shown that data on output, factor inputs, R&D, and firm locations can be used to identify the determinants of each firm's productivity, including the joint distribution of firms' absorptive capacities and locations,  $f(\alpha_i, l_i)$ , and the spillovers' distance decay rate,  $\lambda$ .

Identification relies on the temporal variation in the production function components. To introduce the time dimension, the production period is subdivided into  $T$  years, indexed by  $t = 1, \dots, T$ . Available data provide a snapshot of every firm's operations in each of these years.

Output of firm  $i$  in year  $t$  is denoted as  $Q_{it}$ . The production function has the same structure as before:

$$Q_{it} = A_{it} K_{it}^{\beta_k} L_{it}^{\beta_l}, \quad A_{it} = R_{it}^{\beta_r} S_{lit}^{\alpha_i} \varphi_{it}$$

Time variation in output comes from a number of sources, working together as follows:

1.  $R_{it}$  is determined exogenously. It is drawn at the time of the firm's birth from some known distribution. This determines the time path of  $R_{it}$ .
2. Variation in spillovers,  $S_{lit}$ , is caused by variation in R&D stocks of individual firms.
3. The productivity shock,  $\varphi_{it}$ , is a catch-all term that accounts for productivity variation not caused by own R&D or spillovers. It is further subdivided into three parts:  $\varphi_{it} = \varphi_i V_t W_{it}$ , where  $\varphi_i$  is the constant component drawn by the firm at the time of birth,  $V_t$  – a common time-varying component,  $W_{it}$  – a residual part.
4. Capital and labor are adjusted by the firm each year, in response to time variation in factor prices, consumer expenditures and own productivity.

After taking logarithms, the production function becomes linear:

$$q_{it} = \beta_l \log L_{it} + \beta_k \log K_{it} + \beta_r \log R_{it} + \alpha_i \log S_{lit} + u_i + v_t + w_{it} \quad (7)$$

Lower-case letters denote logarithms.

**Assumption 1.** *Labor and capital are predetermined variables. That is, firms observe  $w_{it}$  only after choosing  $K_{it}$  and  $L_{it}$ .*

This assumption rules out a possibility that  $w_{it}$  has a time-varying part not observed by the econometrician, but observed by the firm at the time when capital and labor are adjusted. Examples of such variables, observable to firms but not the econometrician, may include the quality of labor, or intangible assets other than the R&D stock.

**Assumption 2.** *The idiosyncratic productivity shock  $w_{it}$  is independent across time, independent of the firm's location, own R&D stock, aggregate R&D stocks of all firms, factor prices and consumer expenditures.*

This is a very restrictive assumption. It fails if the random productivity shock is persistent in time, or if productivity is more unpredictable and dispersed for more R&D-intensive firms.

Taken together, these assumptions ensure that  $w_{it}$ , the error term in the production function equation, is strictly exogenous. In particular, factor inputs are not correlated with the error term, even though they are endogenously chosen.

Equation (7) is used to identify parameters  $\beta$  and  $\lambda$  common to all firms<sup>5</sup>, as well as the joint distribution of firm locations and absorptive capacities,  $f(\alpha_i, l_i)$ .

### 3.1.1 Identifying the common parameters

To identify common parameters  $\beta$  and  $\lambda$ , this paper uses a standard technique from the literature on random-coefficient panel data models (e.g. Arellano and Bonhomme (2009), Wooldridge (2005)). This technique relies on transforming equation (7) to eliminate the firm-specific slope and intercept,  $\alpha_i$  and  $u_i$ .

This transformation is convenient to define, if equation (7) is rewritten in the vector

---

<sup>5</sup>Recall, that the spillover term is parametrized by  $\lambda$  (see equation (2))

form:

$$\mathbf{q}_i = \mathbf{z}_i \beta + \mathbf{x}_i(\lambda) \begin{bmatrix} \alpha_i \\ u_i \end{bmatrix} + \mathbf{w}_i, \quad \text{where} \quad (8)$$

$$\mathbf{x}_{it}(\lambda) = [\log S_{lit}(\lambda), 1]$$

$$\mathbf{z}_{it} = [\log K_{it}, \log L_{it}, \log R_{it}, \delta_{1t}, \dots, \delta_{Tt}]$$

$$\delta_1, \dots, \delta_T - \text{year dummies}$$

Define transformation matrix  $\mathbf{M}_i$  as

$$\mathbf{M}_i(\lambda) = \mathbf{I}_T - \mathbf{x}_i(\mathbf{x}'_i \mathbf{x}_i)^{-1} \mathbf{x}'_i$$

A vector multiplied by this matrix gets projected on a linear space orthogonal to  $\mathbf{x}$ . It is easy to see from the definition of  $\mathbf{M}_i$  that  $\mathbf{M}_i \mathbf{x}_i = 0$ . Multiplying both parts of equation (8) by  $\mathbf{M}_i$  removes the firm-specific components:

$$\mathbf{M}_i \mathbf{q}_i = \mathbf{M}_i \mathbf{z}_i \beta + \mathbf{M}_i \mathbf{w}_i \quad (9)$$

Strict exogeneity of the error term, ensured by Assumptions 1 and 2, gives a moment condition that is used to identify the common parameters:

$$E[\mathbf{M}_i(\lambda)(\mathbf{q}_i - \mathbf{z}_i \beta) | \mathbf{z}_i, \{\mathcal{R}_m\}_{m=1}^L, l_i] = 0 \quad (10)$$

The list of conditioning variables includes  $\{\mathcal{R}_m\}_{m=1}^L$ , a vector of aggregate R&D stocks for all years and locations, and  $l_i$ , firm  $i$ 's location. These two variables are required for computing  $\log \mathbf{S}_{l_i}$ ; conditioning on  $\log \mathbf{S}_{l_i}$  itself is impossible, since it depends on the unknown parameter  $\lambda$ .

### 3.1.2 Identifying the absorptive capacities

Knowing common parameters  $\beta$  and  $\lambda$ , one can use equation (8) to obtain information about firm specific absorptive capacities,  $\alpha_i$ .

Identification of  $\alpha_i$  relies on the dataset's time dimension. In a short panel setting, which is implicitly assumed throughout the paper, it is only feasible to obtain a noisy estimate of  $\alpha_i$ .



For example, consider a naive least squares estimator for  $\alpha_i$  and  $u_i$ :

$$\begin{bmatrix} \hat{\alpha}_i \\ \hat{u}_i \end{bmatrix} = (\mathbf{x}'_i \mathbf{x}_i)^{-1} \mathbf{x}'_i (\mathbf{q}_i - \mathbf{z}_i \beta), \quad (11)$$

The estimate of  $\alpha_i$  is contaminated by a finite error term,  $\nu_i$ , which does not vanish as the number of firms in the sample increases to infinity, while the number of years stays fixed:

$$\hat{\alpha}_i = \alpha_i + \nu_i = \alpha_i + \frac{\sum_t (\log S_{it} - \overline{\log S_i})(w_{it} - \bar{w}_i)}{\sum_t (\log S_{it} - \overline{\log S_i})^2} \quad (12)$$

Note, that the magnitude of  $\nu_i$  is inversely proportional to the variance of spillovers received by firm  $i$ . If this variance is small,  $\hat{\alpha}_i$  is a very noisy estimate of  $\alpha_i$ .

Even though the absorptive capacities cannot be identified separately for each firm, equation (8) still provides enough information to identify the *distribution* of absorptive capacities. Arellano and Bonhomme (2009) show that under the following independence assumption, one should be able to identify the distribution of absorptive capacities for each location,  $f_{\alpha|l}(\alpha_i|l_i)$ .

**Assumption 3.** *The vector of idiosyncratic productivity shocks,  $\mathbf{w}_i$ , and the absorptive capacity of firm  $i$ ,  $\alpha_i$ , are independent given  $\mathbf{z}_i$ ,  $\{\mathcal{R}_m\}_{m=1}^L$ , and  $l_i$ .*

This assumption would fail to hold if, for example, low absorptive capacity is determined by a rigid organizational structure of the firm that at the same time isolates the firm from high productivity shocks. This will create a negative dependence between  $\alpha_i$  and the covariance matrix of  $\mathbf{w}_i$ , which violates the latter assumption.

Intuitively, since  $\mathbf{w}_i$  and  $\alpha_i$  are conditionally independent, the error term  $\nu_i$  in equation (12) is also conditionally independent of  $\alpha_i$ . Hence, if the distribution of  $\mathbf{w}_i$  is known or can be somehow identified, one can find the conditional density of  $\nu_i$  and separate it from the conditional density of  $\alpha_i$  using a non-parametric deconvolution. Arellano and Bonhomme (2009) provide a detailed treatment of this argument.

Knowing  $f_{\alpha|l}(\alpha_i|l_i)$ , one can obtain the joint density of firm absorptive capacities and locations by taking a product of  $f_{\alpha|l}(\alpha_i|l_i)$  and the share of firms at location  $l_i$ ,  $n_{l_i}$ . Thus, the joint density  $f(\alpha_i, l_i)$  is identified.

## 3.2 Location choice problem

The spatial pattern of the firms' absorptive capacities found above is used to identify the parameters that shape the geographic distribution of firms in equation (6). These parameters include location fixed effects  $\{a_l\}_{l=1}^L$ , which capture forces induced by local competition and factor prices, and the attraction parameter  $\gamma$ , which reflects the overall influence of spillovers on location choice.

The identification argument goes in two steps. First, it is shown that  $\gamma$  is identified from the pattern of spatial sorting by absorptive capacity. Then, it is demonstrated that one can identify  $\{a_l\}_{l=1}^L$  given  $\gamma$ .

### 3.2.1 Identifying spillover-induced attraction force

To find  $\gamma$ , the location choice problem (6) is solved conditional on absorptive capacity,  $\alpha$ . Denote as  $n_l(\alpha)$  the conditional share of firms choosing location  $l$ . Due to the multinomial logit assumption, the location choice problem has a simple closed form solution:

$$n_l(\alpha) = \frac{\exp(a_l + \gamma\alpha \log S_l)}{\sum_m \exp(a_m + \gamma\alpha \log S_m)} \quad (13)$$

Let  $l_1$  and  $l_2$  be two locations such that  $S_{l_2} \neq S_{l_1}$ ; let  $\alpha_1$  and  $\alpha_2$  be arbitrary values of absorptive capacity. The latter equation can be used to obtain a relationship between shares of firms with absorptive capacities  $\alpha_1$  and  $\alpha_2$  at locations  $l_1$  and  $l_2$ :

$$\frac{n_{l_1}(\alpha_1)}{n_{l_2}(\alpha_1)} = \exp \left[ \gamma(\alpha_1 - \alpha_2) \log \frac{S_{l_1}}{S_{l_2}} \right] \frac{n_{l_1}(\alpha_2)}{n_{l_2}(\alpha_2)}$$

This equation can be solved for  $\gamma$ :

$$\gamma = \log \left[ \frac{n_{l_1}(\alpha_1) n_{l_2}(\alpha_2)}{n_{l_2}(\alpha_1) n_{l_1}(\alpha_2)} \right] \left[ (\alpha_1 - \alpha_2) \log \frac{S_{l_1}}{S_{l_2}} \right]^{-1}$$

Note that  $n_l(\alpha)$  is a conditional distribution, defined as  $n_l(\alpha) = f(\alpha, l)/f_\alpha(\alpha)$ , which implies

$$\gamma = \left[ \log \frac{f(\alpha_1, l_1)}{f(\alpha_2, l_1)} - \log \frac{f(\alpha_1, l_2)}{f(\alpha_2, l_2)} \right] [(\alpha_1 - \alpha_2)(\log S_{l_1} - \log S_{l_2})]^{-1} \quad (14)$$

Intuitively, the attraction parameter  $\gamma$  relates the variation of firms' absorptive capacities across locations to the variation in spillovers these locations offer. The equation is well defined if all four combinations of  $\alpha$  and  $l$  lie in the support of joint density  $f(\cdot, \cdot)$ . The right

hand side of this equation is identified from the production function for any given pair of  $\alpha_1$  and  $\alpha_2$ . Hence,  $\gamma$  is also identified.

### 3.2.2 Identifying location fixed effects.

Given  $\gamma$ , one can identify the location fixed effects from equation (6). Consider the same pair of locations  $l_1$  and  $l_2$  as above. Use equation (13) and the definition of conditional probability to obtain

$$f(\alpha, l_1) = e^{a_{l_1} - a_{l_2}} \left( \frac{S_{l_1}}{S_{l_2}} \right)^{\gamma\alpha} f(\alpha, l_2)$$

Integrate  $\alpha$  out to determine the overall popularity of  $l_1$ :

$$\int f(\alpha, l_1) d\alpha = e^{a_{l_1} - a_{l_2}} \int \left( \frac{S_{l_1}}{S_{l_2}} \right)^{\gamma\alpha} f(\alpha, l_2) d\alpha$$

This equation can be solved for  $a_{l_2}$ :

$$a_{l_2} = a_{l_1} + \log \left[ \int \left( \frac{S_{l_1}}{S_{l_2}} \right)^{\gamma\alpha} f(\alpha, l_2) d\alpha \right] - \log \left[ \int f(\alpha, l_1) d\alpha \right]$$

Note that the location fixed effects are identified up to an arbitrary additive constant. If one normalizes  $a_{l_1}$  to zero, the latter equation can be used to identify location fixed effect  $a_{l_2}$  for an arbitrary location  $l_2$ .

## 4 Estimation

This section discusses an estimation procedure used in the empirical exercise. The procedure consists of two main steps:

1. The production function (7) is estimated in a GMM framework using the moment condition given in (10). This yields common production function parameters,  $\beta$  and  $\lambda$ , and naive OLS estimates of firm-specific absorptive capacities,  $\{\hat{\alpha}_i\}_{i \in \Omega}$ .
2. The location choice problem (6) is solved and estimated by a method of maximum likelihood using  $\lambda$  and  $\{\hat{\alpha}_i\}_{i \in \Theta}$  obtained above. Estimated in this step are the attraction parameter,  $\gamma$ , and the location fixed effects,  $\{a_l\}_{l=1}^L$ .

The details on these two steps are outlined in what follows.

## 4.1 The production function

Using GMM to estimate equation (7) is relatively straightforward; the asymptotic properties of this estimator are studied in a great detail in Arellano and Bonhomme (2009).

However, there is one important issue that arises due to data limitations. Instead of using a direct measure of output, this paper has to rely on revenue data. Replacing output with revenue in equation (7) makes the estimation results more difficult to interpret.

This issue is generic for a big part of literature studying firm productivity. Firm-level prices or physical output are almost never available in a typical dataset. The consequences of using revenue in place of output are discussed at length in Griliches and Mairesse (1995) and Katayama, Lu and Tybout (2009).

In the current setting, replacing output with revenue has two effects on the estimates. First,  $\beta$  and  $\alpha$  become biased by a factor of  $(\sigma - 1)/\sigma$ . As demonstrated by Griliches and Mairesse (1995), this may result in a spuriously observed decreasing returns to scale even if the underlying technology exhibits constant returns, i.e.  $(\beta_l + \beta_k)(\sigma - 1)/\sigma < 1$ , although  $\beta_l + \beta_k = 1$ .

Second, the estimation mixes knowledge spillovers (a positive effect of firm  $j$ 's R&D on firm  $i$ 's productivity) with pecuniary spillovers (a negative effect of firm  $j$ 's R&D on the price index faced by firm  $i$ ). Since knowledge and pecuniary spillovers both affect the revenue and decline with distance from firm  $j$ , estimated  $\alpha_i$  picks up their combined effect and therefore should be interpreted accordingly.

## 4.2 Location choice problem

The production function provides an estimate of the spillover decay parameter,  $\lambda$ , and naive OLS estimates of the firm-specific absorptive capacities,  $\hat{\alpha}_i$ . These are the two missing pieces that are needed to estimate the location choice problem (6).

The estimation is implemented as a maximum likelihood procedure. The likelihood function is obtained from the predictions of the location choice problem, stated in (13). Equation (13), though, cannot be used directly; it makes predictions about the true absorptive capacities,  $\alpha_i$ , which are neither observed nor identified. Hence, it has to be transformed to find the model's predictions about  $\hat{\alpha}_i$ , the naive OLS estimate of  $\alpha_i$ .

Recall, that  $\hat{\alpha}_i$ , defined in equation (12), is a noisy estimate of  $\alpha_i$ . The noise, denoted as  $\nu_i$ , depends on the residual from the production function equation,  $w_{it}$ . To obtain the distribution of  $\hat{\alpha}_i$  from the distribution of  $\alpha_i$ , and vice versa, one has to know the density of

$w_{it}$ .

**Assumption 4.** *Idiosyncratic productivity shocks  $w_{it}$  have normal distribution with zero mean and an unknown variance  $\sigma_w^2$ .*

In principle, this assumption (together with the ones made earlier) is enough to identify and estimate  $f_\alpha(\cdot)$ , the marginal density of  $\alpha_i$ , non-parametrically. However, to make the estimation procedure more straightforward, the latter density is parametrized as well.

**Assumption 5.** *The marginal distribution of the firms' absorptive capacities is normal, with an unknown mean  $\mu_\alpha$ , and an unknown variance  $\sigma_\alpha^2$ .*

The objective is to find a likelihood function for measured absorptive capacity and firm location,  $f(\hat{\alpha}_i, l_i; \theta)$ . The unknown parameters include location fixed effects, attraction parameter  $\gamma$ , and the parameters of the distributions introduced in the latter two assumptions:  $\theta = [\{a_l\}_{l=1}^L, \gamma, \sigma_w^2, \mu_\alpha, \sigma_\alpha^2]'$ . This vector is likely to have very high dimensionality (if geographic space is discretized at a level of county,  $\{a_l\}_{l=1}^L$  alone has more than 3,000 elements). High dimensionality may cause problems during numerical maximization of the likelihood function. Therefore,  $\{a_l\}_{l=1}^L$  and  $\sigma_w^2$  are estimated separately. The estimation of  $\sigma_w^2$  is discussed in the Appendix. Details on  $\{a_l\}_{l=1}^L$  are presented below.

An iteration of the maximum likelihood algorithm is organized as follows:

1. Choose some initial guess for  $\theta_0 = [\mu_\alpha, \sigma_\alpha^2, \gamma]$ .
2. Integrate  $\alpha$  out from equation (13) and take logarithms:

$$\log n_l = a_l + \log \int \frac{\exp(\alpha \gamma \log S_l) f_\alpha(\alpha; \mu_\alpha, \sigma_\alpha^2)}{\sum_{m=1}^L \exp(a_m + \alpha \gamma \log S_m)} d\alpha$$

Spillover decay parameter  $\lambda$  is known from the production function estimates, hence  $S_l$  is also known at this point. The spatial distribution of firms,  $\{n_l\}_{l=1}^L$  is estimated non-parametrically in advance. Let  $H$  be a mapping defined below that takes location fixed effects  $\mathbf{a} = \{a_l\}_{l=1}^L$  as an argument

$$H(\mathbf{a}; \theta_0) = \log n_l - \log \int \frac{\exp(\alpha \gamma \log S_l) f_\alpha(\alpha; \mu_\alpha, \sigma_\alpha^2)}{\sum_{m=1}^L \exp(a_m + \alpha \gamma \log S_m)} d\alpha$$

Berry, Levinsohn and Pakes (1995) show that  $H(\mathbf{a}; \theta_0)$  is a contraction mapping, and that a solution of  $\mathbf{a} = H(\mathbf{a}; \theta_0)$  is unique up to an additive constant. Starting from

an arbitrary initial guess,  $\mathbf{a}_0$ , one can use this mapping recursively to find  $\widehat{\mathbf{a}}(\theta_0)$ , an approximate solution.

3. The estimated location fixed effects are substituted to equation (13), which is then transformed to yield the density of absorptive capacities conditional on location:

$$f_{\alpha|l}(\alpha; \theta_0) = \frac{\exp(\widehat{a}_l + \alpha\gamma \log S_l)}{\sum_{m=1}^L \exp(\widehat{a}_m + \alpha\gamma \log S_m)} \frac{f_{\alpha}(\alpha; \mu_{\alpha}, \sigma_{\alpha}^2)}{n_l}$$

4. The latter density has to be convolved with the density of  $\nu_i$  to obtain the distribution of  $\widehat{\alpha}_i$ . In the process, everything is conditioned on location  $l$ ; since spillovers are determined by location, one can use the density of  $w_{it}$  to derive the conditional density of  $\nu_i$  from equation (12) (see Appendix for details). After convolving  $f_{\alpha|l}(\alpha; \theta_0)$  and  $f_{\nu|l}(\nu; \theta_0)$ , one obtains the joint distribution of locations and the naive OLS estimates of absorptive capacities:

$$f(l, \widehat{\alpha}; \theta_0) = n_l \int f_{\alpha|l}(\widehat{\alpha} - \nu; \theta_0) f_{\nu|l}(\nu; \theta_0) d\nu$$

5. Compute the log likelihood function  $\mathcal{L}(\theta_0) = \sum_i \log f(l_i, \widehat{\alpha}_i; \theta_0)$ . If  $\theta_0$  does not attain the maximum of  $\mathcal{L}(\cdot)$ , start over.

## 5 Data and Results

### 5.1 Data description

The dataset used in the empirical exercise is a subsample of the Standard and Poor's Compustat, focusing on a single 4-digit industry. The sample includes domestic firms whose primary activity is in the development of software products (SIC code 7372, NAICS 4-digit code 5112). The type of primary activity is inferred for each firm from the main Compustat file, the Segments data, and, whenever the first two sources are conflicting or silent, from the companies' 10-K reports. Sometimes a firm may switch or slowly drift away from software development to another product or service. In such cases, those time periods are excluded from the firm's data.

The choice of industry for the empirical application was dictated by three requirements: high R&D, young age, and large sample size. First, R&D should play an important role

Table 1: Data coverage, NAICS 5112, year 2002

Data set		Compustat	Economic Census 2002	NSF Survey of Industrial R&D
Sales	average	293	14	
	total	95,554	103,505	
Employees	average	1,096	48	
	total	357,305	356,708	
R&D expenditures	average	48		3.7
	total	15,760		12,927
Number of firms		326	7,370	3,457

**Notes:**

All monetary values are in 2002 prices, \$million.

Compustat data set is based on the companies' 10-K reports; reported numbers may include the balance sheet items of foreign affiliates

Census 2002 covers business establishments located within the U.S.

NSF Survey covers R&D performed at domestic facilities. Data source: National Science Foundation (2006).

in the industry's production process; judging by the ratio of R&D expenditures to sales, software development is one of the most R&D intensive industries in the U.S. economy. As commonly assumed in the literature, R&D to sales ratio reflects the overall importance of spillovers in the industry (Audretsch and Feldman (1996), Rosenthal and Strange (2001)).

Second, the software industry is relatively young. In an old industry locations of existing firms may reflect distant history rather than current economic reality. The U.S. software industry as we know it emerged in 1970-1980s. Since these years, the main geographic centers of activity in this industry have not changed much. The forces that attracted firms to Silicon Valley in early years of this industry seem to stay potent over time.

Finally, software development is one of the most populous 4-digit industries in the sample of Compustat firms. The requirement that the sample size should be large enough ruled out such candidates as biotechnology and semiconductors.

The sample of Compustat firms is by no means representative of the entire industry. Table 1 compares the sample used in this paper to the total universe of software companies in terms of firm size. Compustat firms tend to be an order of magnitude larger both in sales and employment.

Table 1 also shows that the total R&D expenditures of the Compustat firms are very close to the total national R&D reported in the NSF Survey. Accounting methods used by

Table 2: Summary statistics

	Mean	Std. dev.	Min	Max
Sales, $Y$	243	1,428	0.001	39,508
Employees, $L$	955	3,382	1	71,000
Capital, $K$	30	133	0.001	2,715
R&D expenditures, $XR$	38	230	0	7,407
R&D stock, $R$	131	770	0	25,607
Observations per firm	7.4	4.5	3	26
Number of firms	688			
Years covered	1978–2006			
Observations	4853			

**Notes:**

All monetary values are in 2002 prices, \$million.

the NSF are known to produce conservative figures<sup>6</sup>. However, even if the NSF figure is 50% lower than the actual expenditures, the Compustat data still captures a decent share of a research activity in the industry. Summary statistics for all years of the data are presented in Table 2.

The stock of R&D is constructed for each firm from its annual R&D expenditures,  $XR_{it}$ , using the perpetual inventory method with depreciation<sup>7</sup>:

$$R_{it} = (1 - 0.15)R_{it-1} + XR_{it-1}$$

The Compustat dataset lists one location for each firm; this is an address of corporate headquarters as reported by the firm in its most recent SEC filing. Historical locations are not available in the Compustat, nor it reports locations of facilities other than headquarters. Unable to overcome these data limitations<sup>8</sup>, I assume that R&D and production activity tend to be concentrated around current headquarters. The location data is depicted in Figure 2.

<sup>6</sup>For instance, this fact is documented and investigated for the pharmaceutical industry (Congressional Budget Office (2006)). Independent estimates of R&D in this industry are twice as large as the NSF numbers for certain years.

<sup>7</sup>Following the rest of the literature (Hall, Jaffe and Trajtenberg (2005) and Bloom, Schankerman and Van Reenen (2007), to name just a few examples), the depreciation rate is set at 15% per year.

<sup>8</sup>Historical locations are available in raw SEC filings. However, I could not obtain any of them in electronic form for years prior to 1994. To check whether headquarters are typically co-located with R&D labs, Orlando (2004) used the Directory of American Research and Technology. He found that co-location occurs in more than 90% of his sample. I leave these issues for future work.



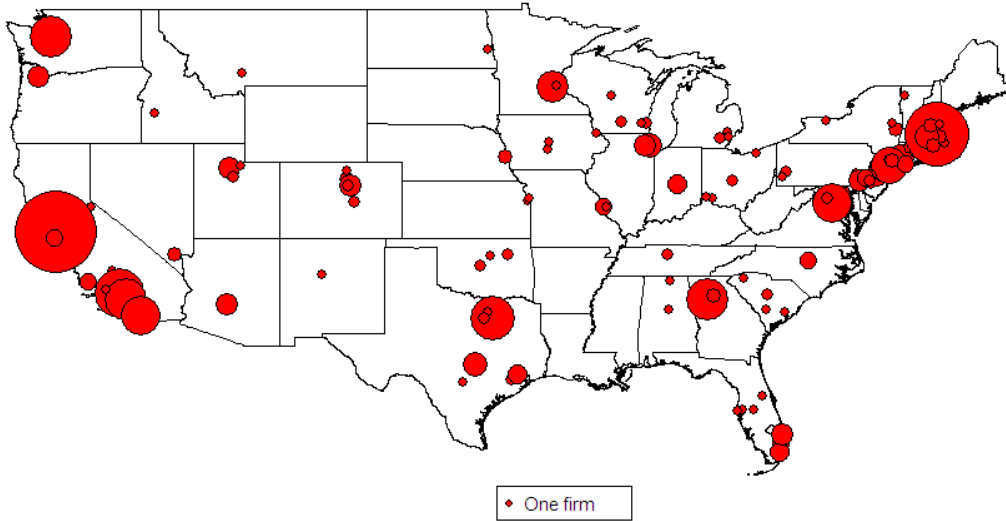


Figure 2: Number of firms, by county

## 5.2 Production function estimates

Table 3 reports the estimates of the common production function parameters. Two specifications of equation (7) are reported. Specification (1) restricts absorptive capacities to be the same for all firms. This is the usual assumption in the spillovers literature. Ignoring heterogeneity in  $\alpha_i$  may bias the estimate of  $\lambda$ ; the Appendix provides a detailed discussion of why this bias may occur. Specification (2) is the preferred one; it does not place any restrictions on absorptive capacities and is estimated using Arellano-Bonhomme GMM, as discussed above.

To avoid numerical stability problems at high values of  $\lambda^9$ , the logarithms of spillovers are replaced with the levels. The estimates of common parameters in both specifications suggests that software development is not a capital intensive industry, which is in line with the common sense. The technology shows decreasing returns to scale; this contradicts the CRS assumption of the location model. However, as explained before, the returns to scale estimate may be biased downwards, due to using sales as a proxy for physical output. To produce the bias of the necessary size, the CES demand elasticity has to be around  $\sigma = 5$ .

Unlike the rest of the common parameters, the estimate of the spillover decay,  $\lambda$ , varies greatly across the specifications. The first specification where absorptive capacities are re-

---

<sup>9</sup>When  $\lambda$  is high, spillovers decline very fast, which causes  $S_i$  to be close to zero for remote locations. Taking a logarithm of a near-zero value produces a lot of noise in the objective function maximized during estimation.

Table 3: Production function estimates

	(1)		(2)	
$\log K_{it}$	0.09	(4.7)	0.11	(4.6)
$\log L_{it}$	0.74	(25)	0.66	(17)
$\log R_{it}$	0.11	(6.4)	0.09	(2.7)
$S_{it}$	0.01	(2.4)		
Spillover half-life ( $\log(2)/\lambda$ ), km	1,931	(1.0)	59	(2.2)
# of observations	4,852			

**Notes:**

Robust t-statistics are reported in the parentheses.

Specifications:

(1): Firms are homogeneous in  $\alpha_i$ . NLS FE estimator.

(2): Firms are heterogeneous in  $\alpha_i$ . A-B GMM estimator.

stricted to be the same for all firms, does not give a tight estimate of  $\lambda$ . The nonlinear least squares objective function that is used to obtain the estimates has two minima: a local minimum corresponding to a half-life distance of 30 kilometers, and a global minimum at 1,931 kilometers. Neither of these two candidate estimates is significantly different from zero.

The preferred specification (2) gives a unique solution for  $\lambda$ , which corresponds to a half-life distance estimate of 59 kilometers significant at 5% level. To put this value in some perspective, imagine a firm that chooses between locating in San Francisco, and in San Jose, an approximate center of Silicon Valley. Even though these two cities are only a one hour commute apart, spillovers originating in San Jose decline in their intensity by roughly 50% by the time they arrive to San Francisco. If the firm is concerned about spillovers, the San Jose location provides a substantial advantage.

### 5.3 Preliminary evidence on sorting

According to the model of location choice, an opportunity to absorb spillovers attracts firms to high-spillover locations. The force of attraction depends on the firm's absorptive capacity,  $\alpha_i$ . As shown in Proposition 1, this force creates a pattern of spatial sorting by  $\alpha_i$ , with high-type firms over-represented in high-spillover regions.

Before imposing the structure of the location choice model, one has to be certain that this sorting pattern exists in data. First, I check if spatial sorting by  $\alpha_i$  can be visually detected. Figure 3 plots OLS estimates of firms' absorptive capacities,  $\hat{\alpha}_i$ , against spillovers received by these firms. These estimates have many outliers caused by the noise term  $\nu_i$  in equation (12). The variance of this noise is especially large for firms located in low-spillover regions.

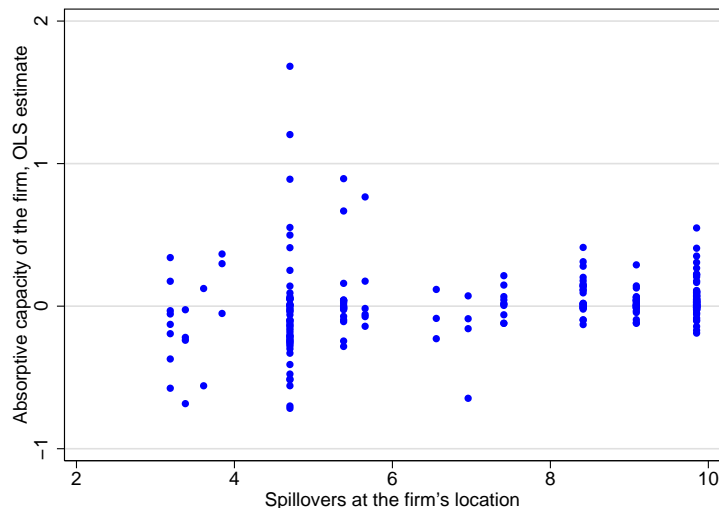


Figure 3: Spatial sorting by  $\alpha_i$ . Firms from high-spillover counties ( $S_l > 3$  \$bil)

To minimize outliers, firms from low-spillover counties are dropped<sup>10</sup>. Consistent with the story of sorting, Figure 3 suggests that firms in counties with higher levels of spillovers have on average higher absorptive capacities.

To test the link between  $\hat{\alpha}_i$  and  $S_{l_i}$  more formally, consider an equation that relates the absorptive capacity of a firm to the spillovers received at the firm's location:

$$\hat{\alpha}_i = \delta_0 + \delta_1 S_{l_i} + \epsilon_i \quad (15)$$

The results of estimating this equation are reported in Table 4. To ensure that the test outcomes are not driven by outliers, two alternative estimators are used: the median and the robust regression. Both methods are known to be more stable with respect to outliers than the OLS estimator. As a further robustness check, the estimation exercise is repeated on two samples: the full sample, and a subsample of firms located in top-20 high-spillover counties. Since the standard median regression estimator is built on the assumption of error homoskedasticity, two t-statistics are reported in each specification – the standard, and the bootstrapped one.

The estimates confirm that the sorting pattern found in Figure 3 is statistically significant. There is a positive relationship between the firm's absorptive capacity and the spillovers

<sup>10</sup>The remaining firms plotted in Figure 3 are predominantly located in Silicon Valley and the area around Boston.

received by the firm in its home county<sup>11</sup>.

Table 4: Preliminary evidence on sorting by  $\alpha_i$

Dependent variable: $\hat{\alpha}_i$	Median, all	Robust, all	Median top-20	Robust, top-20
$S_{i_i}$	0.011	0.019	0.013	0.025
Regular t-statistic	(4.3)	(2.4)	(5.1)	(7.4)
Bootstrapped t-statistic	(1.6)	(2.1)	(2.0)	(3.7)
# of firms	688	688	268	268

**Notes:**

All: sample includes all firms.

Top-20: firms from top-20 high-spillover counties.

Median: median regression. Robust: robust regression.

## 5.4 Geographic scope of spillovers in the presence of sorting

As the main prediction of the location choice model is found to be consistent with data, the structure of the model can be imposed and used to estimate the parameters of the choice equation (6).

The estimates are reported in Table 5. First, note that  $\gamma$ , the parameter capturing the attraction force induced by spillovers, is positive, consistent with the preliminary evidence and Proposition 1.

Second, as suggested by the estimate of  $\mu_\alpha$ , more than a half of firms have negative  $\alpha_i$ . The reason why  $\alpha_i$  may be negative was discussed at length in the estimation section. R&D stocks generate knowledge spillovers and pecuniary spillovers, which have opposite effects on the outside firm's sales. Since output in the production function is proxied by sales, the estimate of  $\alpha_i$  picks up both effects and can potentially be negative depending on which effect dominates.

To get a feel of the strength of spatial sorting and the magnitude of spillovers, consider an average firm located in Santa Clara county, the center of Silicon Valley, with an R&D stock of \$0.22 billion in year 2002<sup>12</sup>. Suppose that the firm enters in 2002. As reported in Table 6, this entry has a sizeable effect on other Santa Clara software firms: their aggregate

<sup>11</sup>In a series of unreported exercises,  $\hat{\alpha}_i$  is replaced with a proxy widely used in the agglomeration literature: the ratio of R&D expenditures to sales (e.g. Audretsch and Feldman (1996) and Rosenthal and Strange (2001)). The results obtained in these exercises are qualitatively the same as here, irrespective of the estimation method or the subsample used.

<sup>12</sup>For comparison, this is approximately a one-fifth of the R&D stock of Adobe Systems.

Table 5: Parameters of the location choice problem

Attraction parameter, $\gamma$	0.053
Distribution of $\alpha_i$ :	
$\mu_\alpha$	-0.05
$\sigma_\alpha^2$	0.41
# of firms	688

Table 6: Spillovers from a \$220 million firm entering in Santa Clara county. Effects on firms from neighbor counties

County	Distance from Santa Clara, km	Increase in sales			Increase in sales, \$million
		Mechanisms active:			
		Dist. decay	Sorting	Both	
Santa Clara	0	1.49%	1.49%	1.49%	114
Santa Cruz	34	0.90%	0.61%	0.32%	0.67
San Mateo	45	0.78%	1.33%	0.69%	36
Alameda	48	0.74%	1.06%	0.50%	13
San Francisco	68	0.56%	0.79%	0.25%	3.75

revenue increases by 1.49 percent. The aggregate revenue earned by software establishments in Santa Clara county in 2002 is \$7,640 million<sup>13</sup>, hence the total gain in sales caused by spillovers from the entering firm amounts to \$114 million in year 2002 alone.

Effects from spillovers experienced by firms in other counties depend on two factors: the distance from Santa Clara and the composition of local firms in terms of absorptive capacity. To compare the contributions of these factors, Table 6 isolates their effects on sales. In column 3, only the distance decay mechanism is allowed to work. Spatial sorting by  $\alpha_i$  is ignored: the gains in sales are calculated in assumption that absorptive capacities of all Silicon Valley firms have the same distribution as in Santa Clara county. In column 4, the predictions of gains are calculated assuming that spillovers do not decline with distance, but at the same time accounting for the cross-county differences in absorptive capacities. Finally, column 5 reports estimates of productivity gains that take into account both sorting and distance decay.

Comparing predictions in columns 3 and 5 reveals the importance of accounting for firm heterogeneity in  $\alpha_i$  and endogenous location choice when studying the scope of R&D spillovers. In this example, ignoring spatial sorting seriously overestimates the effects of spillovers on firms from San Francisco and Santa Cruz counties.

<sup>13</sup>Data source: Economic Census 2002.

## 6 Conclusion

In this paper, I construct a theoretical model of an industry where firms generate spillovers, are heterogeneous in absorptive capacity and mobile in geographic space. This model predicts that firms are spatially sorted by their absorptive capacity. Equilibria of the model feature agglomerations that attract firms with high absorptive capacity, and sparsely populated regions that are more popular among firms indifferent to spillovers. Due to this sorting pattern, the overall effect of spillovers varies across regions.

The model is fitted to data from the US software industry. It is often believed that knowledge spillovers are ubiquitous in this industry, and that they are strong enough to have an agglomerative effect on firm locations. Spillovers are viewed to be partly responsible for high geographic concentration of software firms around few large urban areas. The data provide evidence that knowledge spillovers between software companies indeed exist and are highly localized in geographic space. It is also found that firms which are more responsive to spillovers are more likely to be located in big clusters.

The above findings have direct implications for policy. The model was used to simulate the gains from spillovers that would occur if local officials of Santa Clara county in California managed to attract one additional average-sized firm to their location. As the results suggest, the gains would be shared by the neighbor counties. These gains roughly decline with distance. However, distance is not the only factor that determines the gains from spillovers here. For example, Santa Cruz county is located next to Santa Clara but receives less gains per firm than the more distant Alameda county. The size of the gains from spillovers is almost equally affected by the distance to Santa Clara, and the pattern of spatial sorting of firms. Knowing who are the main beneficiaries of the policy and what are its benefits is important; this knowledge would help to avoid free-riding by the adjacent counties on the efforts of the Santa Clara officials. The model also suggests that the similar policy targeted at attracting new firms is less likely to yield any big percentage gains in output if implemented in some state with a sparse population of software companies.

The model developed in this paper assumes away an endogeneity of R&D and its potential effect on absorptive capacity. Relaxing this assumption is the most important direction for future work. Common sense suggests that doing R&D is cheaper in agglomerated cities, as specialized labor and materials are likely to be in abundant supply there. If R&D positively affects the firm's absorptive capacity (as suggested by Cohen and Levinthal (1989)), the observed pattern of spatial sorting may have nothing to do with sorting at all. The observed

heterogeneity of firms may develop ex-post, after the firms enter and start investing in R&D. Firms in agglomerated cities will become more R&D-intensive and thus will have higher absorptive capacity. Therefore, it is important to make R&D endogenous and allow for its correlation with absorptive capacity. This will also make possible the use of the model for simulation of policies promoting R&D, such as R&D subsidies.

## References

- Arellano, Manuel, and Stephane Bonhomme.** 2009. “Identifying Distributional Characteristics in Random Coefficients Panel Data Models.” mimeo.
- Audretsch, David B., and Maryann P. Feldman.** 1996. “R&D Spillovers and the Geography of Innovation and Production.” *The American Economic Review*, 86(3): 630–640.
- Baldwin, Richard E., and Toshihiro Okubo.** 2006. “Heterogeneous Firms, Agglomeration and Economic Geography: Spatial Selection and Sorting.” *Journal of Economic Geography*, 6(3): 323–346.
- Behrens, Kristian, Gilles Duranton, and Frederic Robert-Nicoud.** 2010. “Productive Cities: Sorting, Selection and Agglomeration.” mimeo.
- Berry, Steven, James Levinsohn, and Ariel Pakes.** 1995. “Automobile Prices in Market Equilibrium.” *Econometrica*, 63(4): 841–890.
- Bloom, Nicholas, Mark Schankerman, and John Van Reenen.** 2007. “Identifying Technology Spillovers and Product Market Rivalry.” National Bureau of Economic Research Working Paper 13060.
- Cohen, Wesley, and Daniel Levinthal.** 1989. “Innovation and Learning: The Two Faces of R&D.” *The Economic Journal*, 99(397): 569–596.
- Combes, Pierre-Philippe, Gilles Duranton, Laurent Gobillon, Diego Puga, and Sebastien Roux.** 2009. “The Productivity Advantages of Large Cities: Distinguishing Agglomeration from Firm Selection.” mimeo.
- Congressional Budget Office.** 2006. “Research and Development in the Pharmaceutical Industry.” Congress of the United States Pub. No. 2589.

- Eaton, Jonathan, and Samuel Kortum.** 2001. “Technology, Trade, and Growth: A Unified Framework.” *European Economic Review*, 45(4-6): 742–755.
- Ellison, Glenn, Edward L. Glaeser, and William R. Kerr.** 2010. “What Causes Industry Agglomeration? Evidence from Coagglomeration Patterns.” *The American Economic Review*, 100(3): 1195–1213.
- Feenstra, Robert C.** 1996. “Trade and Uneven Growth.” *Journal of Development Economics*, 49(1): 229–256.
- Glaeser, Edward L.** 1999. “Learning in Cities.” *Journal of Urban Economics*, 46(2): 254–277.
- Greenstone, Michael, Richard Hornbeck, and Enrico Moretti.** 2010. “Identifying Agglomeration Spillovers: Evidence from Winners and Losers of Large Plant Openings.” *Journal of Political Economy*, 118(3): 536–598.
- Griliches, Zvi, and Jacques Mairesse.** 1995. “Production Functions: The Search for Identification.” National Bureau of Economic Research Working Paper 5067.
- Hall, Bronwyn, Adam Jaffe, and Manuel Trajtenberg.** 2005. “Market Value and Patent Citations.” *The RAND Journal of Economics*, 36(1): 16–38.
- Jaffe, Adam B., Manuel Trajtenberg, and Rebecca Henderson.** 1993. “Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations.” *Quarterly Journal of Economics*, 108(3): 577–598.
- Katayama, Hajime, Shihua Lu, and James R. Tybout.** 2009. “Firm-level Productivity Studies: Illusions and a Solution.” *International Journal of Industrial Organization*, 27(3): 403–413.
- Keller, Wolfgang.** 2002. “Global Localization of International Technology Diffusion.” *The American Economic Review*, 92(1): 120–142.
- Krugman, Paul.** 1991. “Increasing Returns and Economic Geography.” *The Journal of Political Economy*, 99(3): 483–499.
- Lychagin, Sergey, Joris Pinkse, Margaret Slade, and John Van Reenen.** 2010. “Spillovers in Space: Does Geography Matter?” National Bureau of Economic Research Working Paper 16188.



- Martin, Philippe, and Carol Ann Rogers.** 1995. “Industrial Location and Public Infrastructure.” *Journal of International Economics*, 39(3-4): 335–351.
- Matusik, Sharon F., and Michael B. Heely.** 2005. “Absorptive Capacity in the Software Industry: Identifying Dimensions That Affect Knowledge and Knowledge Creation Activities.” *Journal of Management*, 31(4): 549–572.
- National Science Foundation.** 2006. “Research and Development in Industry: 2002.” Division of Science Resource Statistics NSF 06-322.
- Orlando, Michael.** 2004. “Measuring Spillovers from Industrial R&D: On the Importance of Geographic and Technological Proximity.” *The RAND Journal of Economics*, 35(4): 777–786.
- Rath, Kali P.** 1992. “A Direct Proof of the Existence of Pure Strategy Equilibria in Games with a Continuum of Players.” *Economic Theory*, 2(3): 427–433.
- Romer, Paul M.** 1986. “Increasing Returns and Long-Run Growth.” *The Journal of Political Economy*, 94(5): 1002–1037.
- Rosenthal, Stuart, and William Strange.** 2001. “The Determinants of Agglomeration.” *Journal of Urban Economics*, 50(2): 191–229.
- Saxenian, AnnaLee.** 1996. *Regional advantage: Culture and competition in Silicon Valley and Route 128*. Harvard University Press.
- Seim, Katja.** 2006. “An Empirical Model of Firm Entry with Endogenous Product-Type Choices.” *The RAND Journal of Economics*, 37(3): 619–640.
- Wooldridge, Jeffrey M.** 2005. “Fixed-Effects and Related Estimators for Correlated Random-Coefficient and Treatment-Effect Panel Data Models.” *The Review of Economics and Statistics*, 87(2): 385–390.

## 7 Appendix

### 7.1 The equilibrium of the location game

Let  $\Omega = [0, 1]$  be a set of firms; a firm is represented by a point  $i \in \Omega$ . All firms have the same set of actions,  $\mathcal{A} = \{1, \dots, L\}$ . Each firm is characterized by an R&D stock  $R_i$ , innate

productivity  $\varphi_i$ , absorptive capacity  $\alpha_i$ , and a vector of idiosyncratic location preference shocks  $[\tilde{\varepsilon}_{i1}, \dots, \tilde{\varepsilon}_{iL}]$ . These are well-defined random variables, in a sense that they are all Borel-measurable functions from  $\Omega$  to  $\mathbb{R}$ . A pure-strategy profile  $\Lambda$  is a measurable function from  $\Omega$  to  $\mathcal{A}$ . Payoffs to firm  $i$  are given by  $\tilde{\pi}_i^*$ , a mapping from the set of all profiles  $\Theta = \{\Lambda\}$  to  $\mathbb{R}$ .

$$\tilde{\pi}_i^*(\Lambda) = \frac{1}{b} \log \left[ \frac{1}{\sigma} \left( \frac{\sigma-1}{\sigma} R_i^{\beta_r} \varphi_i \right)^{\sigma-1} \right] + \frac{1}{b} \log \left[ \sum_m E_m \left( \frac{P_m(\Lambda)}{\tau_{\Lambda(i)m} c_{\Lambda(i)}} \right)^{\sigma-1} \right] + \frac{(\sigma-1)}{b} \alpha_i \log S_{\Lambda(i)}(\Lambda) + \tilde{\varepsilon}_{i\Lambda(i)} \quad (16)$$

where

$$S_l(\Lambda) = \sum_{m=1}^L e^{-\lambda \rho(l,m)} \mathcal{R}_m, \quad \mathcal{R}_m = \int_{\Lambda^{-1}(m)} R_i di,$$

$$P_m(\Lambda) = \frac{\sigma}{\sigma-1} \left[ \int_{\Omega} \left( \frac{R_i^{\beta_r} S_{\Lambda(i)}^{\alpha_i} \varphi_i}{c_{\Lambda(i)} \tau_{\Lambda(i)m}} \right)^{\sigma-1} di \right]^{-\frac{1}{\sigma-1}}$$

$$c_l = c(w_l, r_l)$$

To ensure that all expressions above are well-defined, assume that there is no city with zero costs (i.e.,  $c_l > 0$ ), and that the following two integrals exist:

$$\mathcal{R} = \int_{\Omega} R_i di, \quad \int_{\Omega} \left( R_i^{\beta_r} \mathcal{R}^{\alpha_i} \varphi_i \right)^{\sigma-1} di$$

Given an arbitrary strategy profile  $\Lambda \in \Theta$ , each firm chooses an action that maximizes the firm's payoff. This defines the best response mapping  $h : \Theta \rightarrow \Theta$ . The equilibrium of the game is a fixed point of  $h$ .

In order to prove that the equilibrium exists, some notation has to be introduced. First, denote

$$d_l = \frac{\left[ \sum_m E_m \left( \frac{P_m(\Lambda)}{\tau_{lm} c_l} \right)^{\sigma-1} \right]^{1/b}}{\sum_k \left[ \sum_m E_m \left( \frac{P_m(\Lambda)}{\tau_{km} c_k} \right)^{\sigma-1} \right]^{1/b}}; \quad \mathbf{d} = \{d_l\}_{l=1}^L$$

Intuitively,  $\mathbf{d}$  would be the distribution of firms across cities, if spillovers were not affecting location choice. Vector  $\mathbf{d}$  belongs to an  $(L-1)$ -dimensional simplex  $\Delta^{L-1}$ .

Second, note that the total R&D always sums up to  $\mathcal{R}$ , irrespective of the strategy profile being played:

$$\sum_{m=1}^L \mathcal{R}_m = \mathcal{R}$$

Hence, the distribution of R&D stocks, denoted as  $\mathbf{r} = \{\mathcal{R}_m/\mathcal{R}\}_{m=1}^L$ , also belongs to  $\Delta^{L-1}$ . Denote  $\Psi$  a set of all pairs  $(\mathbf{d}, \mathbf{r})$ ; this set forms a Cartesian product of two  $(L-1)$ -simplexes.

The proof of existence follows the steps of a similar argument in Rath (1992).

**Lemma 1.** *The best response mapping  $h$  can be expressed as a composition of two mappings:*

$$h = h_2 \circ h_1 : \Theta \xrightarrow{h_1} \Psi \xrightarrow{h_2} \Theta$$

*Proof.* Mapping  $h_1$  is given by the definitions of  $\mathbf{d}$  and  $\mathbf{r}$ . To find  $h_2$ , one has to show that knowing  $(\mathbf{d}, \mathbf{r})$  alone is sufficient to determine the best response of every firm without relying on any additional information on other firms' actions contained in  $\Lambda$ . Note, that the payoff function (16) has only two elements that depend on actions of other firms: the spillover term and the term containing the price index. By definition, the spillover term depends on  $\mathbf{r}$  only. One can use the definition of  $\mathbf{d}$  to transform the payoff function and obtain

$$\begin{aligned} \tilde{\pi}_i^*(\Lambda) = \frac{1}{b} \log \left[ \frac{1}{\sigma} \left( \frac{\sigma-1}{\sigma} R_i^{\beta_r} \varphi_i \right)^{\sigma-1} \right] + \log d_{\Lambda(i)} + \frac{1}{b} \log \left[ \sum_k \sum_m E_m \left( \frac{P_m(\Lambda)}{\tau_{km} c_k} \right)^{\sigma-1} \right] \\ + \frac{(\sigma-1)}{b} \alpha_i \log S_{\Lambda(i)}(\mathbf{r}) + \tilde{\varepsilon}_{i\Lambda(i)} \end{aligned}$$

The third term still depends on  $\Lambda$ , but it does not depend on the firm's own actions, since the firm is infinitely small and cannot affect the price index  $P(\Lambda)$  by changing  $\Lambda(i)$ . Hence, one can drop this term from the location choice problem without affecting the solution. The rest of the expression on the right hand side depends on other firms' actions only via  $\mathbf{d}$  and  $\mathbf{r}$ . Therefore, to find the best response, it is enough to know the vector  $(\mathbf{d}, \mathbf{r})$ , which proves the lemma.

**Lemma 2.** *If  $h_1 \circ h_2$  has a fixed point, then  $h$  also has one.*

*Proof.* Let  $(\mathbf{d}^0, \mathbf{r}^0)$  be such fixed point, that is,

$$(\mathbf{d}^0, \mathbf{r}^0) = h_1(h_2(\mathbf{d}^0, \mathbf{r}^0))$$

Take  $h_2$  of both sides of this equation:

$$h_2(\mathbf{d}^0, \mathbf{r}^0) = h_2(h_1(h_2(\mathbf{d}^0, \mathbf{r}^0))) = h(h_2(\mathbf{d}^0, \mathbf{r}^0))$$

Define a strategy profile  $\Lambda^0 = h_2(\mathbf{d}^0, \mathbf{r}^0)$ . As evident from the latter equation,  $\Lambda^0$  is a fixed point of  $h$ .

**Lemma 3.**  $h_1 \circ h_2$  has a fixed point.

*Proof.*  $h_1 \circ h_2$  is a continuous mapping (continuity is not strictly proven here) from  $\Psi$  to  $\Psi$ .  $\Psi$  is a convex compact set (as a Cartesian product of two multidimensional simplexes). By the Brouwer fixed point theorem,  $h_1 \circ h_2$  has a fixed point.

**Theorem 1.** *The location game has an equilibrium.*

*Proof.* Apply sequentially lemmas 3 and 2. Best response mapping  $h$  has a fixed point, which is by definition an equilibrium.

## 7.2 Proof of Proposition 1

Location choice problem (6) can be solved to obtain the cities' shares of firms conditional on the firms' absorptive capacity,  $n_l(\alpha)$

$$n_l(\alpha) = \frac{\exp(a_l + \gamma\alpha \log S_l)}{\sum_m \exp(a_m + \gamma\alpha \log S_m)}$$

Using Bayes' theorem, one can transform this equation to yield the density of absorptive capacities conditional on location

$$f_{\alpha|l}(\alpha|l) = \frac{\exp(a_l + \gamma\alpha \log S_l)}{\sum_m \exp(a_m + \gamma\alpha \log S_m)} \frac{f_\alpha(\alpha)}{n_l}$$

To prove that  $f_{\alpha|l}(\alpha|l_2)$  stochastically dominates  $f_{\alpha|l}(\alpha|l_1)$ , it suffices to show that these two distributions satisfy the increasing likelihood ratio property. The likelihood ratio equals

$$\frac{f_{\alpha|l}(\alpha|l_2)}{f_{\alpha|l}(\alpha|l_1)} = \exp((a_{l_2} - a_{l_1}) + \gamma\alpha(\log S_{l_2} - \log S_{l_1})) \frac{n_{l_1}}{n_{l_2}}$$

Let  $\alpha_2 > \alpha_1$  be two arbitrary values of absorptive capacity. It follows from the latter equation that

$$\frac{f_{\alpha|l}(\alpha_2|l_2)}{f_{\alpha|l}(\alpha_2|l_1)} = \exp(\gamma(\alpha_2 - \alpha_1)(\log S_{l_2} - \log S_{l_1})) \frac{f_{\alpha|l}(\alpha_1|l_2)}{f_{\alpha|l}(\alpha_1|l_1)}$$

Since  $S_{l_2} > S_{l_1}$ , and  $\gamma > 0$ , the likelihood ratio increases in  $\alpha$ :

$$\frac{f_{\alpha|l}(\alpha_2|l_2)}{f_{\alpha|l}(\alpha_2|l_1)} > \frac{f_{\alpha|l}(\alpha_1|l_2)}{f_{\alpha|l}(\alpha_1|l_1)}$$

Increasing likelihood ratio implies first-order stochastic dominance.

### 7.3 Inferring the geographic scope of spillovers when firms are mobile and heterogeneous in absorptive capacity

In contrast to this paper, the spillovers literature usually assumes that firms are homogeneous in  $\alpha_i$  and that firm locations are exogenous. This allows applying a much simpler standard fixed effects estimator to obtain the production function parameters,  $\beta$  and  $\lambda$ . However, if firms differ in  $\alpha_i$  and do choose locations, these estimates are in general biased either upwards or downwards, depending on the true parameters of the model and the spatial distribution of firms in equilibrium.

To see how the bias may occur, consider a sample of pairwise distances between spillover-generating and spillover-receiving firms. Suppose that the firms are of two types: the ones that uniformly benefit from spillovers, and the ones that do not benefit at all. The productivity effects of a spillover on these two types of firms is depicted by two blue-colored curves in Figure 4, as a function of distance from the spillover's origin.

The equilibrium in the location choice model produces a mixture of high-type and low-type firms in each city. One could use the same axes in Figure 4 to plot an average productivity effect of spillovers for firms in this equilibrium. A point on the horizontal axis,  $d$ , represents a subsample of all firms that are located at distance  $d$  from some spillover-generating firm. A point on the vertical axis represents an average effect of spillovers traveling from distance  $d$  on this subsample.

Intuitively, whenever one assumes that  $\alpha_i$  are homogeneous and tries to fit equation (7) to data, he finds an average effect of spillovers in all such subsamples, corresponding to every  $d$ , and then approximates the resulting curve by an exponential function. This curve may take any shape that lies between the solid and the dashed line in Figure 4, depending on which type of firms dominates at different distance ranges.

For instance, if there is one agglomeration in the country, the high-type firms are on average located closer to the sources of spillovers, than the low-type firms. Hence the high-type firms are over-represented in the subsamples corresponding to small  $d$ 's. Consequently,

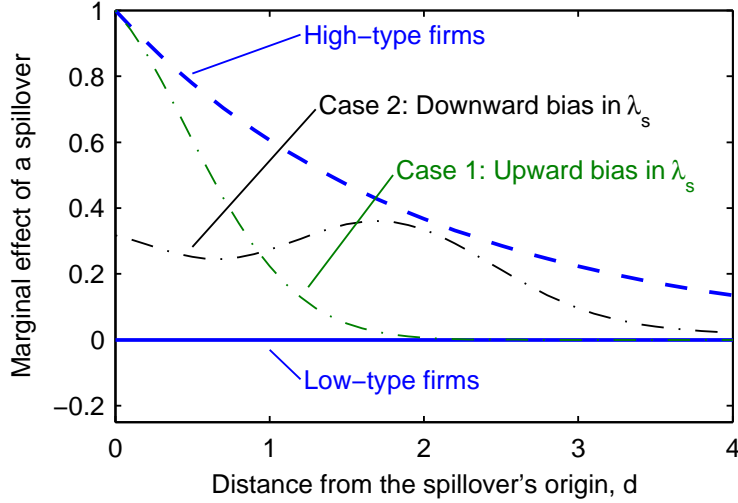


Figure 4: Potential bias in the estimates of spillover decay,  $\lambda$

the curve representing the average effect of a spillover will be closer to the dashed line (the effect of spillovers on high-type firms) around zero, and to the solid line (the effect of spillovers on low-type firms) at high distances. Therefore, the estimate of  $\lambda$ , the rate of spillover's decline, will be biased upwards, which corresponds to case 1 in Figure 4. Simulations show that downward bias, depicted as case 2, is also possible under some spatial distributions of firms.

## 7.4 The distribution of noise in estimated absorptive capacities

Assume that the error in the production function equation,  $w_{it}$ , is an i.i.d. random variable with constant variance  $\sigma_w^2$ . Let  $\widehat{W}_i = M_i \widehat{w}_i$  be the residual from the within-transformed equation (9). Find firm  $i$ 's sum of squared residuals:

$$\widehat{W}_i' \widehat{W}_i = \widehat{w}_i' M_i' M_i \widehat{w}_i = \widehat{w}_i' M_i \widehat{w}_i = \widehat{w}_i' A_i' A_i \widehat{w}_i$$

where  $A_i$  is a  $(T-2) \times T$  matrix whose rows form an orthonormal basis in the subspace orthogonal to  $x_i$ . By construction,  $A_i A_i' = I_{T-2}$ ; it can be shown that  $M_i = A_i' A_i$ .

To see what is approximated by the sum of squared residuals, consider the covariance matrix of  $A_i w_i$ :

$$\text{cov}(A_i w_i) = E[A_i w_i w_i' A_i'] = E[A_i E[w_i w_i' | A_i] A_i'] = \sigma_w^2 E[A_i A_i'] = \sigma_w^2 I_{T-2},$$

Firm  $i$ 's sum of squared residuals approximates the trace of this matrix,  $(T - 2)\sigma_w^2$ . Taking a sum across firms, obtain an estimator for  $\sigma_w^2$ :

$$\hat{\sigma}_w^2 = \frac{1}{n(T-2)} \sum_i \widehat{W}_i' \widehat{W}_i$$

Assume that  $w_{it}$  has normal distribution. According to equation (12),  $\nu_i$  is a linear function of  $w_{it}$  with coefficients dependent on location,  $l$ . Hence, it is also normally distributed conditional on  $l$ . Its variance is easily found from (12):

$$\sigma_{\nu|l}^2 = \frac{\sigma_w^2}{\sum_t (s_{lt} - \bar{s}_l)^2}$$