Firm Location and the Creation and Utilization of Human Capital

Andres Almazan
University of Texas

Adolfo de Motta
McGill University

Sheridan Titman*
University of Texas and the NBER

1 December 2003

Abstract

This paper presents a theory of location choice that draws on insights from the incomplete contracts and investment flexibility (real option) literatures. We provide conditions under which human capital is more efficiently created and better utilized within industrial clusters that contain similar firms. Our analysis indicates that location choices are influenced by the extent to which training costs are borne by firms versus employees as well as by the uncertainty about future productivity shocks and the ability of firms to modify the scale of their operations. Extensions of our model consider, among other things, endogenous technological choices by firms in clusters and how behavioral biases (i.e., managerial overconfidence about their firms' prospects) can affect firms' location choices.

*Almazan and Titman are from the McCombs School of Business, University of Texas at Austin, and de Motta is from McGill University. E-mail addresses: andres.almazan@bus.utexas.edu, adolfo.demotta@mcgill.ca, and titman@mail.utexas.edu. We would like to thank seminar participants at Carnegie Mellon University, CEMFI, Columbia University, UBC, USC, and University of Texas and to Aydogan Alti, Ty Callahan and Frederic Robert-Nicoud for helpful comments.

1 Introduction

One of the fundamental issues in economics relates to the location of production. Where firms and industries locate is a primary determinant of the economic growth of both regional and national economies. These choices affect the design of our cities as well as the pattern of trade between nations.

This paper examines the location choice of firms within knowledge-based industries (e.g., software and pharmaceutical development). Specifically, we consider the incentives of these firms to locate either together, within geographical clusters, or in a number of geographically separate regions. The issue of industrial clustering dates back at least to Marshall (1890) and has received substantial attention in the recent literature.[1] By focusing on transportation costs and exogenous natural advantages, the early literature explains why firms in some industries tend to locate in a number of geographically separate regions.[2] This literature, however, is much less applicable to knowledge-based firms whose products are almost costless to transport and which employ very little in the way of resources other than human capital.[3] More applicable are the arguments that focus on the advantages of clustering that arise because of the benefits of a more active market for skilled labor and the potential for knowledge spillovers.

The discussion in most of the recent literature, which points to Silicon Valley

[1] There is an extensive urban economics literature that addresses location issues for generic industries. For excellent reviews see Fujita, Krugman and Venables (1999), Fujita and Thisse (2002) and Duranton and Puga (2003).

[2] Ellison and Glaeser (1999) find that proxies related to natural advantages can explain roughly 20% of their empirical measures of agglomeration.

[3] Abstracting from transportation costs seems particularly suitable to explain location in knowledge-based industries. Moreover, Glaeser and Kohlhase (2003) have reported that transportation costs for manufacturing goods have fallen by over 90% in the last century, and argue that, to a large extent, the world is better characterized as a place where "it is essentially free to move goods, but expensive to move people." This suggests that the issues that we discuss here may be more broadly applicable.

as the quintessential example, is that strong economic forces lead knowledge-based industries to cluster.[4] However, this literature generally ignores those cases of successful knowledge-based firms that locate away from industrial clusters. The most notable case is Microsoft, which became the industry leader after locating in Seattle, which at the time was not a center for software development. Another notable case is Nations-Bank, a North Carolina Bank which became one of the largest banks in the U.S. after taking over Bank of America.[5]

The model developed in this paper is consistent with the Silicon Valley phenomena as well as with the observation that some knowledge-based firms choose to locate on their own. The model is based on the idea that a key distinction between locating within a cluster rather than in isolation has to do with the competitiveness of the market for skilled labor. Specifically, since we assume that it is costly for workers to change locations, an isolated firm can become a monopsonist in the market for the specialized labor while, within a cluster, workers with industry specific skills can sell their labor in a competitive labor market. As Manes and Andrews (1994, p. 120) describe it, the structure of labor markets played a central role in Microsoft's location decision:

> "Paul Allen increasingly argued for a move back to familiar Seattle turf. Hiring might be simpler in Silicon Valley, but keeping employees would clearly be harder, a major consideration in a business where the primary assets walk out the door every night (...) The tremendous demand for their services had made Bay Area engineers notoriously fickle; at the first sign of dissatisfaction, they would find a position across the street or check out a 'job fair' brimming with offers."

[4] In addition to the above cited papers in the economics literature, there is also a discussion of these issues in the management literature. In particular, see Porter (1990). See also Saxenian (1994) for a forceful discussion of these issues in the case of Silicon Valley.

[5] Ellison and Glaeser (1997) document that while a slight degree of concentration is widespread, the more extreme concentration of industries such as automobile and computer exists only in a smaller subset of industries.

As the preceding quote illustrates, a competitive labor market can be a two-edged sword. It can help firms hire labor when they are expanding, but it can also make it difficult to retain labor. Moreover, as we illustrate in our model, labor is more efficiently utilized within clusters since they can be redeployed to the most productive firms. Specifically, within clusters, firms that realize favorable firm-specific productivity shocks benefit from hiring workers that leave firms that suffer unfavorable firm-specific shocks.[6] This aspect of our model extends the analysis in Krugman (1991) that considers the advantage of labor market pooling.[7]

The case for clustering becomes less straightforward when we consider how workers acquire their specialized skills. Following Rotemberg and Saloner (2000), we show that there is an added advantage associated with clustering if developing human capital requires that the worker expends effort. However, if the development of these skills requires an investment (e.g., training) by the firm, then there is an offsetting cost associated with clustering.[8] In other words, within a cluster, employees appropriate the value of the skills (and technology) acquired on the job because they can sell their skills at a competitive price to their employer's competitors. Hence, they have an incentive to put in the effort required to acquire such skills. However, anticipating this, firms within a cluster have less incentive to invest in their employees' human capital, and thus provide less training than their more isolated counterparts.

Our model captures the interaction between these forces in a parsimonious way

[6] There is a second line of research that examines the advantages of thick labor markets that arise from better matching workers with firms. Papers that address the role of the market in improving the quality of matching include Helsley and Strange (1990, 1991) and Combes and Duranton (2001). Mortesen and Pissarides (1999) and Pissarides (2000) review the search literature which addresses the role of market in improving the chances of matching.

[7] Dumais, Ellison and Glaeser (1997) examine this issue empirically. Specifically, they provide evidence that plants locate near other industries when they share the same type of labor, and conclude that "labor market pooling is a dominant force in explaining the agglomeration of industry."

[8] Matouschek and Robert-Nicoud (2003) provide a related analysis of the effect of human capital investments on firms' location decisions. See also Grossman and Hart (1986) for a similar trade-off in their analysis of vertical integration.

that explicitly illustrates that the creation and allocation of human capital are two sides of the same coin: the way that human capital is allocated determines how it is created. Moreover, the model identifies several characteristics that predict which knowledge-based industries are likely to exhibit clustering. For example, when the potential for industry-wide growth is not excessive and when firm-level uncertainty is high, then industries are likely to exhibit clustering. There is also likely to be more clustering in industries where the workers must exert effort to acquire their skills but less clustering in growing industries where firms must provide significant training for their workers.

The model also provides implications about how differences *between* firms within an industry affect their location choices. Specifically, firms with better growth prospects are likely to be better positioned to benefit from their workers' contribution to their own training and from hiring workers that are trained by their competitors. This result provides an alternative interpretation to the empirical findings by Henderson (1986) and Ciccone and Hall (1996) that productivity increases with the density of the economic activity and by Holmes and Stevens (2002) that plant sizes are higher within industry clusters.[9] The conventional interpretation of these findings is that because of various externalities, productivity is higher in clusters. In contrast, our results raise the possibility that clusters tend to attract the most efficient firms, rather than make existing firms more efficient.

We consider three extensions of the main analysis. In the first extension, we introduce uncertainty about aggregate productivity (i.e., systematic shocks) and analyze the relative advantages of clusters versus isolation. We find that a greater degree of aggregate uncertainty reduces firms' incentives to cluster. This is because higher

[9]There are a number of empirical studies that examine issues that relate to productivity and agglomeration. For an excellent review, see Rosenthal and Strange (2003).

aggregate uncertainty in clusters limits firms' abilities to reallocate human capital among themselves and also because, as we show, firms in clusters are not as well positioned to incorporate information about changes in productivity.

The second extension, which allows firms to design their production processes in ways that make them more or less compatible with other firms, explores the possibility that technological choices differ in clusters versus isolation. Specifically, we consider the incentives of firms to deviate from industry norms in clusters. The analysis identifies two opposing effects. By deviating from industry norms, firms increase firm specific risk, which in turn increases the redeployment benefits of clustering. However, if the labor employed by firms with very different technologies are less compatible, a countervailing effect emerges. The relative importance of these effects determines whether clusters prevail in industries in which experimentation and the introduction of new technologies is central.

In the third and final extension we consider how behavioral biases affect firms' location decisions. Specifically, we show that overconfident entrepreneurs are more likely to be attracted to clusters because they overvalue the benefits associated with the ability to hire workers that are trained by their competitors. Within our setting, overconfidence can have social benefits as well as costs. In isolation, overconfidence is costly because it leads to too much training. However, within a cluster, since rational entrepreneurs train too few workers, it is possible that social efficiency and firms' profits can be improved when entrepreneurs are overconfident.

The rest of the paper is organized as follows. Section 2 describes the model and section 3 analyzes it. Section 4 considers the issue of location for heterogenous firms and section 5 presents the analysis of location when workers can also invest. Section 6 considers location when firms also can choose their technologies and section 7 analyzes how overconfidence may affect firms' location choices. Section 8 presents some conclusions of the analysis. Proofs and other technical derivations are relegated to the appendix.

2 The model

We consider a risk neutral economy populated by a continuum of ex-ante identical entrepreneurs (i.e., firms), and an unlimited supply of unskilled workers with reservation wage w^R. Firms have access to perfect capital markets and are endowed with an investment project described below.

As described in Figure 1, there are three relevant dates in the economy, $t = 0, 1$, and 2. At $t = 0$, firms *permanently* locate. Firms choose whether to locate in a regional economy that includes other firms that employ and train similar workers (i.e., within a cluster), or alternatively, to locate away from the cluster (i.e., in isolation).

At $t = 1$, the production process starts with an *initial* stage during which firm i hires a certain number of unskilled workers H_{1i} and then trains h_{1i} of those hired.[10] The actual training is ex post observable but is not verifiable and hence, the workers cannot sign a contract with the firm guaranteeing that they will be trained.

At $t = 2$, the *growth* stage, the firm receives a productivity shock. After observing the shock, firms can contract or expand their operations by either laying off some of its existing workers or, within clusters, by hiring new workers who have obtained training with one of the firm's competitors. The main difference between locating within a cluster rather than in isolation is the access that firms have to trained workers in the growth stage. While in a cluster skilled workers are hired in a competitive market, in isolation, firms have exclusive access to the workers they train at $t = 1$, and behave as monopsonists in the labor market. For simplicity, we assume that firms cannot

[10] For most of the analysis, we assume that training is costly to the firm but requires no effort from the worker. In section 5, we relax this assumption and examine the case in which workers' effort affect the effectiveness of the firm's training.

train new workers at $t = 2$. The number of trained workers employed by firm i at this date is denoted as h_{2i}.

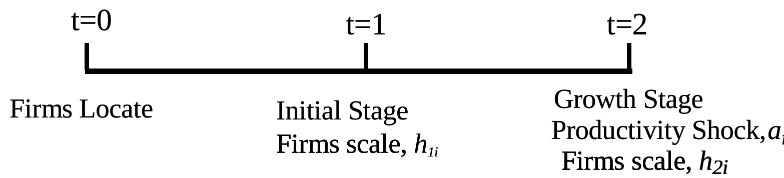

Figure 1: Sequence of Events

Firms have the following production functions at $t = 1$ and $t = 2$:

$$Q_1(h_{1i}) = \alpha h_{1i} - \tau \frac{h_{1i}^2}{2} \quad \text{and} \quad Q_2(h_{2i}) = a_i h_{2i} - \beta \frac{(h_{2i} - h_{1i})^2}{2}.$$

$Q_1(h_{1i})$ and $Q_2(h_{2i})$ correspond to the production functions during the initial stage and the growth stage respectively. In each stage, each firm determines the scale of its operations: h_{1i} during the initial stage, (i.e., the amount of workers trained at $t = 1$) and h_{2i} during the growth stage (i.e., the amount of trained workers employed at $t = 2$). We refer to $\alpha > 0$ as firm productivity in the initial stage and to $\tau > 0$ as the importance of the firm's training costs. In addition, we refer to a_i as the firm productivity in the growth stage and to $\beta > 0$ as the intensity of the firm's adjustment costs, which make the firm production at $t = 2$ depend on the initial scale h_{1i}.[11]

Parameters α, τ and β are deterministic, identical for all firms, and known at $t = 0$ before production starts. In contrast, a_i is the realization of a random variable $\tilde{a}_i$, a firm-idiosyncratic productivity shock that occurs at $t = 2$ prior to production. The shock $\tilde{a}_i$ is distributed according to the c.d.f. $F(\tilde{a}_i)$ with density $f(\tilde{a}_i)$, and bounded support $[a_L, a_H]$. We assume that $a_L > 0$ and $(a_H - a_L)\left(\frac{\tau}{\beta} + 1\right) < \alpha$, which simplifies the analysis by avoiding non-negativity constraints, and we denote $E(\tilde{a}_i) \equiv \bar{a}$ and $E(\tilde{a}_i - \bar{a})^2 \equiv \sigma^2$. Shocks are independent across firms, specifically, we assume that if

[11] The presence of symmetric adjustment costs and the technological linkage between the periods simplifies the analysis but are not necessary for the trade-off between human capital development and allocation that emerges from the model.

a continuum of firms populates a cluster then the empirical distribution of realized shocks, $F(a_i)$, is identical to the ex-ante c.d.f., $F(\tilde{a}_i)$, i.e., no aggregate uncertainty exists.[12]

Notice that we are implicity assuming that the marginal productivity of untrained unskilled workers is the same inside and outside the firm. That is, unless worker training is provided, the firm has no special advantage in employing unskilled workers. This assumption implies that firms benefit from employing unskilled workers only when they can compensate them at a salary below their reservation wage.[13]

We finish the presentation of the model by specifying three important assumptions of our model. First, we assume short term labor contracts that cannot be contingent on training. Hence, our analysis of the contracting issues draws on the literature on incomplete contracts, i.e., Grossman and Hart (1986), and on the effects of the inalienability of human capital, i.e., Hart and Moore (1994).[14] Second, we assume that trained workers must stay in their respective locations (i.e., regions) after they are trained. This assumption captures the idea that individuals initially locate in the region offering the best employment opportunities, but after establishing roots in the community find it costly to relocate. Finally, since we are primarily interested in the interaction between location and the development and utilization of human capital, we abstract from the effects that location may have on product market competition. Specifically, we assume a constant price for a firm's output (that we normalize to one)

[12] In section 6, we relax this assumption and examine location choices in the presence of aggregate uncertainty on productivity shocks.

[13] To save on notation, we have omitted the effect that untrained workers can have on a firm's production. Because, as stated in the main text, we assume that the productivity of unskilled workers outside (i.e., w^R) and inside the firm is the same, our results are unchanged if we specify $Q_j(h_{ji}, u) = Q_j(h_{ji}) + (\theta - w^R) \cdot u = Q_j(h_{ji})$ (for $j = 1, 2$) where u is the amount of untrained workers employed and θ is their marginal productivity inside the firm (which is equal to w^R).

[14] In section 5, we revisit this issue and analyze the location problem when both the firm and workers can make non-contractible relation specific investments. There, we discuss why long-term contracts themselves can create misincentives in firm-worker relationships.

and that firms' products can be transported costlessly within a competitive market.

3 Analysis of the model

We first consider the production and training decisions in isolation and then within an industrial cluster. In each case, we proceed backwards; we start with the scale decision at the growth stage, h_{2i}, and then consider the scale decision at the initial stage, h_{1i}.

3.1 Isolation

In isolation, the analysis of the growth stage is straightforward. At $t = 2$, the firm acts as a monopsonist in the market for skilled workers, and thus pays them the reservation wage for their services, w^R which we normalize to zero.[15] Since the supply of skilled workers is limited by the amount of workers that the firm itself trains at $t = 1$ (i.e., $h_{2i} \leq h_{1i}$), after $\tilde{a}_i$ is realized, the firm solves:

$$\max_{h_{2i} \in [0, h_{1i}]} a_i h_{2i} - \beta \frac{(h_{2i} - h_{1i})^2}{2}. \tag{1}$$

Solving (1) the demand for skilled labor is $h_{2i}^* = h_{1i}$ (i.e., the firm retains all the workers it trains at $t = 1$).

At $t = 1$, the firm decides how many workers to hire and to train. On-the-job training is valuable to workers but is costly to firms and, more importantly, is non-contractible among parties. This means that firms will provide training according to their internal trade-offs without fully incorporating the positive effect of training on workers, a fact that, as we show, will play a crucial role in clusters. Formally, let H_{1i} be the number of workers hired, and then, among those hired, let h_{1i} be the number

[15] This is without loss of generality as long as $a_L \geq w^R = 0$, that is the productivity of a skilled worker inside the firm is always higher than outside the firm.

of them assigned to positions that provide on-the-job training.[16] Hence, firm i solves

$$\max_{h_{1i},H_{1i}} \alpha h_{1i} - \tau\frac{h_{1i}^2}{2} + E(\tilde{a}_i h_{1i}), \tag{2}$$

subject to:

$$h_{1i} \leq H_{1i}. \tag{3}$$

Since unskilled workers are equally productive inside and outside the firm, H_{1i}^* remains indeterminate in equilibrium (other than $H_{i1}^* \geq h_{1i}^*$). Therefore, we solve to obtain h_{1i}^* and then simply set $H_{1i}^* = h_{1i}^*$:

$$\max_{h_{1i}} \alpha h_{1i} - \tau\frac{h_{1i}^2}{2} + \bar{a}h_{1i}. \tag{4}$$

From (4), we derive the first order condition (f.o.c.) to obtain:

$$h_{1i}^* = \frac{\alpha + \bar{a}}{\tau}, \tag{5}$$

which, as showed before, also equals h_{2i}^*. Substituting in (2) yields the firm's value in isolation at $t = 0$, V_i^I:

$$V_i^I = \frac{\alpha^2}{2\tau} + \frac{\alpha\bar{a}}{\tau} + \frac{\bar{a}^2}{2\tau} = \frac{(\alpha + \bar{a})^2}{2\tau}. \tag{6}$$

3.2 Clustering

3.2.1 The growth decision

At $t = 2$, the growth decision by firm i is the solution to the following problem:

$$\max_{h_{2i}} a_i h_{2i} - \beta\frac{(h_{2i} - h_{1i})^2}{2} - wh_{2i} \tag{7}$$

[16]The explicit distinction between hired, H_{1i}, and trained workers, h_{1i}, is consistent with but not essential for the analysis of the firm's decision in isolation. However, we choose to keep the distinction in the isolation analysis to maintain parallelism with the cluster analysis below, where such a distinction plays a crucial role. Also notice that although workers hired at $t = 1$ must receive the reservation wage the expressions are simplified due to the normalization $w^R = 0$.

where w is the wage paid to the skilled workers at $t = 2$. From the f.o.c., we obtain firm i's demand for skilled workers, h_{2i}^*:

$$h_{2i}^* = h_{1i} + \frac{a_i - w}{\beta}. \tag{8}$$

According to (8), firm i hires (*fires*) additional workers if its realized productivity, a_i, is greater (*smaller*) than the wage at $t = 2$, w. The importance of the adjustment costs (measured by β) determines the sensitivity of the firm's demand for skilled workers to a_i.

To determine the wage that clears the market in the cluster at $t = 2$, i.e., w, we need to consider (i) the aggregate demand for skilled workers, D_2^H:[17]

$$D_2^H \equiv \int_{a_L}^{a_H} (h_{1i} + \frac{a_i - w}{\beta}) f(a_i) da_i = \frac{\bar{a} - w}{\beta} + \int_{a_L}^{a_H} h_{1i} f(a_i) da_i \tag{9}$$

and (ii) the aggregate supply of skilled workers:

$$S_2^H \equiv \int_{a_L}^{a_H} h_{1i} f(a_i) da_i. \tag{10}$$

Market clearing, i.e., $D_2^H = S_2^H$, yields the equilibrium wage which is equal to the average productivity of the firms in the cluster, i.e., $w = \bar{a}$.

3.2.2 The initial scale decision

At $t = 1$, because on-the-job training is non-contractible among parties, firms will provide such training according to their internal trade-offs without fully incorporating the positive effect of training on workers. Consequently, workers will be wary of taking lower wages against promises of future skills that will not necessarily be provided. We model this time-inconsistency by considering that, first, a firm hires a certain number of workers H_{1i} at $t = 1$ and then, among those hired, the firm allocates

[17] Notice that, by virtue of the independence of technology shocks, $\int_{a_L}^{a_H} a_i f(a_i) da_i = \bar{a}$.

h_{1i} of them to positions that provide on-the-job training. This optimal training decision is anticipated (i.e., rationally expected) by workers who condition their initial salary demands at $t = 0$, w_{0i}, on the total number of workers hired by the firm. Specifically, workers consider the probability of being trained as the ratio of the anticipated number of workers trained, h_{1i}^e, to the number hired, H_{1i}, and reduce their salary accordingly. Formally, firm i maximizes expected profits by solving:

$$\max_{h_{1i},H_{1i},w_{0i}} \left(\alpha h_{1i} - \tau \frac{h_{1i}^2}{2} + E\left[(\tilde{a}_i - w)\, h_{2i}^* - \beta \frac{(h_{2i}^* - h_{1i})^2}{2}\right]\right) - w_{0i} H_{1i}, \tag{11}$$

subject to:

$$\begin{aligned} h_{1i} &\leq H_{1i} \\ w_{0i} &= -\frac{h_{1i}^e}{H_{1i}} w. \end{aligned} \tag{12}$$

Constraint (12) captures the fact that the firm may choose not to train some of the hired workers while constraint (12) considers the salary reduction from the reservation wage, which is normalized to zero, that workers will accept when hired by the firm as compensation for their expected human capital acquisition (i.e., a participation constraint for workers). Notice that with the initial salary reduction, $-w_{0i}$, workers "pay" at $t = 1$ for their (anticipated) training. Specifically, a worker is trained with probability $\frac{h_{1i}^e}{H_{1i}}$, and, if trained, her salary increases at $t = 2$ by w.[18] Substituting constraint (12) into (11) reveals that H_{1i}^* is indeterminate in equilibrium (other than $H_{1i}^* \geq h_{1i}^*$). This is due to our assumption that unskilled workers are equally productive inside and outside the firm. Consequently, without loss of generality, we assume that the firm hires only the workers that it can credibly claim to train, which implies that $H_{1i}^* = h_{1i}^*$. Given this, the following problem can be solved to obtain firm i's

[18] This contrasts with the case of isolation in which workers do not reduce their wages, i.e., $w_{0i} = w^R = 0$, because workers realize they will not capture any of the value of their developed human capital (i.e., in isolation the firm behaves as a monopsonist at $t = 2$).

optimal scale h_{1i}^*:

$$\max_{h_{1i}} \quad \alpha h_{1i} - \tau \frac{h_{1i}^2}{2} + E\left[(\tilde{a}_i - w) h_{2i}^* - \beta \frac{(h_{2i}^* - h_{1i})^2}{2} \right] + h_{1i}^e w. \tag{13}$$

Substituting $w = \bar{a}$ and $h_{2i}^* = h_{1i} + \frac{a_i - \bar{a}}{\beta}$, and considering that the anticipated level of training h_{1i}^e is not a choice variable for the firm,[19] problem (13) can be reduced to:

$$\max_{h_{1i}} \ \alpha h_{1i} - \tau \frac{h_{1i}^2}{2}, \tag{14}$$

which, when solved, implies

$$h_{1i}^* = \frac{\alpha}{\tau}. \tag{15}$$

Expression (15) shows that, in clusters, a firm's initial scale decision is "myopic," i.e., it is not affected by its expected productivity $\bar{a}$. While, all else equal, a higher expected productivity increases the firms' incentive to invest in human capital, it also increases the wage at $t = 2$, and hence, reduces the firms' incentive to create human capital. In an economy of identical firms, these two effects offset each other leading to the firms' myopia on their initial scale decisions.

Finally, in (11), we can compute the firm value at $t = 0$ in the cluster, V_i^C:

$$V_i^C = \frac{\alpha^2}{2\tau} + \frac{\alpha \bar{a}}{\tau} + \frac{\sigma^2}{2\beta}, \tag{16}$$

which can be decomposed into three terms: (i) the value created from production in the first period (i.e., $Q(h_{i1}^*) = \frac{\alpha^2}{2\tau}$), (ii) the value of the human capital created by the firm (as measured by the wages obtained by the workers trained by the firm, i.e., $h_{1i}^e w = \frac{\alpha}{\tau}\bar{a}$) and (iii) the value of the option to adapt the scale of production in the cluster after the shock is realized (i.e., $E[(\tilde{a}_i - w) h_{2i}^* - \beta \frac{(h_{2i}^* - h_{1i})^2}{2}] = \frac{\sigma^2}{2\beta}$).

[19] To be sure, even though in a (rational expectations) equilibrium the actual level of training equals the workers' conjecture, $h_{1i}^e = h_{1i}^*$, the conjectured level of training cannot be affected by the firm.

3.3 The choice of location

When deciding their locations, firms face a trade-off between the advantages of isolation on the creation of human capital and the advantages of clusters on the utilization of human capital. This trade-off, which is apparent by comparing (6) and (16), is described in the following proposition:

Proposition 1 *The difference in firm value in the cluster versus in isolation is*

$$V_i^C - V_i^I = \frac{\sigma^2}{2\beta} - \frac{\bar{a}^2}{2\tau}. \tag{17}$$

Proposition 1 summarizes the main implications of the analysis so far. In industries where trained workers are more productive, i.e., larger $\bar{a}$, the relative value of isolation increases. In contrast, when there is more uncertainty about which firms will be most productive, i.e., when σ^2 is larger, the relative value of clustering is higher. In addition, the relative value of clusters vis-à-vis isolation is also related to the importance of the firm's adjustment costs and the cost of creating human capital. Specifically, large adjustment costs (i.e., high β) reduce the value of flexibility in the cluster, and hence, of clustering, while large costs of creating human capital (i.e., high τ), make the acquisition of human capital from clusters relatively more attractive, and hence, promotes clusters.

More intuition about the location trade-off can be gained by examining how a social planner would allocate resources in this economy. The social planner must consider two issues: the optimal creation of human capital at $t = 1$, and its optimal utilization at $t = 2$. The competitive market in the cluster allocates skilled workers (once trained) optimally. Hence, the social planner would simply replicate the worker allocation that occurs in the cluster: $h_{2i}^* = h_{1i} + \frac{a_i - \bar{a}}{\beta}$. Therefore, the planner's problem is reduced to finding the optimal firm scale at $t = 1$ in the presence of a competitive labor market, but without the time inconsistency problem in the creation of human capital by firms (e.g., by assuming that firms internalize the future salary gains that workers obtain from firm training). Formally, the problem would be identical to the clustering program but where the term $h_{1i}^{e}w$ is replaced by $h_{1i}w$:

$$\max_{h_{1i}} \quad \alpha h_{1i} - \tau \frac{h_{1i}^2}{2} + E\left[\left(\tilde{a}_i - w\right) h_{2i}^* - \beta \frac{\left(h_{2i}^* - h_{1i}\right)^2}{2}\right] + h_{1i}w. \tag{18}$$

From the f.o.c., we get $h_{1i}^{FB} = \frac{\alpha+\bar{a}}{\tau}$, and substituting in (18) the "first best" firm value is obtained:

$$V_i^{FB} = \frac{\alpha^2}{2\tau} + \frac{\alpha\bar{a}}{\tau} + \frac{\bar{a}^2}{2\tau} + \frac{\sigma^2}{2\beta}. \tag{19}$$

Notice that in the social planner's solution, firms would utilize human capital as they do in the cluster, but would create human capital as they do in isolation. Formally, this is reflected in an additional positive component (with respect to the value in isolation) due to the optimal reallocation of human capital, i.e., $V_i^{FB} = V_i^I + \frac{\sigma^2}{2\beta}$, and a positive additional component (with respect to the value in clusters) due to the optimal investment in human capital, i.e., $V_i^{FB} = V_i^C + \frac{\bar{a}^2}{2\tau}$.

3.4 An alternative specification

We conclude this section by briefly discussing an alternative specification that produces a trade-off that is similar to the one obtained here. Specifically, rather than assuming that training is not contractible, we could have assumed that the workers lack the resources to pay-up front for their training, i.e., there is a minimum wage on the first date that exceeds the equilibrium wage that includes the discount workers are willing to take to receive their training. Of course, both assumptions require, additionally, the inalienability of workers' human capital once acquired (i.e., that firms become residual claimants of their workers' human capital).

Intuitively, with either friction, firms within clusters do not fully internalize the improvements in their workers' human capital, and thus underinvest in their worker's training. In the case that we consider, the fact that training is not contractible creates a time inconsistency problem (namely, once workers agree to reduce their initial salaries, firms feel tempted not to honor their training commitment). Similarly, when there is a minimum wage at $t = 0$, firms in clusters may not be able to benefit from training their workers, and may thus undertrain their workers when they cannot make binding commitments. In contrast, an isolated firm can capture the benefit of training their workers in the first period because they can underpay them relative to their productivity in the last period. As was the case in the previous model, this implies that firms are more likely to isolate when the gains associated with human capital creation are the highest.

We chose to present the model with the non-contractibility assumption in order to simplify the analysis and to facilitate welfare comparisons. However, we feel the limited liability (i.e., minimum wage) alternative can be a compelling assumption in some circumstances and constitutes, in any case, an additional foundation for our analysis that reinforces our results.

4 The location choice of heterogeneous firms

Up to this point, we have considered an economy of (ex-ante) identical firms. In this section, we introduce firm heterogeneity to the analysis. We proceed as follows: First, we present a partial equilibrium analysis, where we take as given the presence of a cluster with salary w and examine the location choice of firm i, which is assumed to be too small to affect w. Once we characterize the individual firm's incentives to cluster, we consider the general equilibrium analysis where all firms simultaneously choose where to locate, and examine the endogenous formation of clusters.

4.1 The location decision of an individual firm

Consider firm i with productivity $\tilde{a}_i$ which is distributed with c.d.f. $F_i(\tilde{a}_i)$ and density $f_i(\tilde{a}_i)$, and where $\tilde{a}_i \in [a_{L_i}, a_{H_i}]$ with $a_{L_i} > 0$. Let $\bar{a}_i \equiv E(\tilde{a}_i)$ and $\sigma^2 \equiv E(\tilde{a}_i - \bar{a}_i)$. Firm's i value in isolation immediately follows from (6) in the previous section, i.e., $V_i^I = \frac{\alpha^2}{2\tau} + \frac{\alpha \bar{a}_i}{\tau} + \frac{\bar{a}_i^2}{2\tau}$. However, to derive its value in the cluster, we must modify the analysis to take into account that, in general, $\bar{a}_i \neq w$. Following similar steps as before (see the appendix for details), we find the human capital created at $t = 1$,

$$h_{1i}^* = \frac{\alpha + (\bar{a}_i - w)}{\tau} \tag{20}$$

and the value of a clustered firm,

$$V_i^C = \frac{\alpha^2}{2\tau} + \frac{\alpha \bar{a}_i}{\tau} + \frac{E\left(\tilde{a}_i - w\right)^2}{2\beta} + \frac{\bar{a}_i^2 - w^2}{2\tau}. \tag{21}$$

Notice that if a firm's expected productivity is equal to the cluster wage, i.e., $\bar{a}_i = w$, the expressions obtained in the case with homogenous firms hold. That is, (20) converges to (15), i.e., $h_{1i}^* = \frac{\alpha}{\tau}$ and (21) converges to (16), i.e., $V_i^C = \frac{\alpha^2}{2\tau} + \frac{\alpha \bar{a}_i}{\tau} + \frac{\sigma^2}{2\beta}$.

To examine firm i's incentives to join a cluster with wage w, we subtract (6) from (21) and rearrange terms to obtain

$$V_i^C - V_i^I = \frac{E\left(\tilde{a}_i - w\right)^2}{2\beta} - \frac{w^2}{2\tau}. \tag{22}$$

As was the case with homogenous firms, the decision is determined by the trade-off between the benefits of the cluster (i.e., redeployment of human capital, $\frac{E(\tilde{a}_i - w)^2}{2\beta}$) and its costs (i.e., underinvestment in human capital, $\frac{w^2}{\tau}$). Further intuition, however, can be gained by expressing (22) as

$$V_i^C - V_i^I = \frac{\left(\overline{a}_i - w\right)^2 + \sigma^2}{2\beta} - \frac{w^2}{2\tau}, \tag{23}$$

from which we can see that the benefits of redeploying human capital stem from: (i) ex-ante differences between expected productivity and the cluster average productivity, i.e., $\frac{(\overline{a}_i - w)^2}{2\beta}$, and (ii) ex-post differences among firms' realized productivities, i.e., $\frac{\sigma^2}{2\beta}$. In other words, the benefits of joining a cluster come from date $t = 2$ differences in productivity that may or may not be anticipated.

Notice that the first effect (i.e., $\frac{(\overline{a}_i - w)^2}{2\beta}$) induces clustering even in the absence of uncertainty (i.e., $\sigma^2 = 0$), because firms with expected productivity that is greater (*smaller*) than the wage expect to hire (*fire*) workers at $t = 2$ (i.e., $E\left(h^*_{2i}\right) - h_{1i} = \frac{\bar{a}_i - w}{\beta}$). Because of the decreasing returns to scale in the creation of human capital, there is an efficiency gain associated with having the firms with a low need for human capital at $t = 2$ train more workers than they need and indirectly transfer those workers to higher productivity firms, which create less human capital than they want to employ. Although a firm does not directly compensate its competitors for the trained workers that join their firms, indirect compensation accrues to the net providers of human capital who, when located in clusters, can offer a lower wage rate at $t = 1$.[20]

An examination of (23) allows us to consider the factors that influence a firm's decision to join the cluster:

Proposition 2 *Clusters are more attractive for firms with: (i) higher variance productivity, σ^2, (ii) lower adjustment costs, β, (iii) higher costs of creating human capital, τ, and (iv) expected productivity that differs more from the cluster wage, i.e., higher $|\overline{a}_i - w|$.*

Proposition 2 shows that clusters are relatively more valuable when either an-

[20] Consulting firms generally need substantially more associates (i.e., the more junior consultants) than partners, and as a result, a number of associates eventually go to work for their clients. This observation can be viewed from the perspective of our model if we view the consulting firms as expecting date $t = 2$ productivity that induce them to shed employees. As such, our model provides a rationale for why consulting firms and their clients may benefit from locating near each other.

ticipated (i.e., $|\overline{a}_i - w|$) or unanticipated (i.e., σ^2) differences in productivity give a reason to redeploy human capital and also when such a redeployment is not expensive (i.e., low β).[21] Furthermore, if firms are not efficient at producing human capital (i.e., low τ) the underinvestment problem in clusters is ameliorated (the difference between the production of human capital in clusters and in isolation is $\frac{w}{\tau}$).

Finally, we can check that an increase in the wage reduces the incentives to cluster (i.e., $\frac{d(V_i^C - V_i^I)}{dw} < 0$) unless $\bar{a}_i < (1 - \frac{\beta}{\tau})w$. While, in general, a higher wage exacerbates the underinvestment problem and reduces the incentive to cluster, firms whose productivities are sufficiently below the wage ($\bar{a}_i < (1 - \frac{\beta}{\tau})w$) can benefit from higher cluster wages. For these firms (which are on average "big sellers" of human capital in the cluster), the higher wage produces an increase in the expected revenue from the human capital sold that more than offsets the additional revenue lost due to an exacerbated underinvestment problem.[22]

4.2 Endogenous clusters with heterogeneous firms

In the previous section, we examined the incentives of firms to join an existing cluster with an exogenous wage, w. This section examines the endogenous formation of clusters, i.e., the simultaneous decision of firms that can isolate themselves or can locate within an endogenously formed cluster.

For simplicity, we assume that there is a continuum of firms with random expected productivities $\tilde{a}_i = \overline{a}_i + \tilde{\varepsilon}_i$ where $\tilde{\varepsilon}_i$ are zero-mean i.i.d. random variables, and $g(\overline{a}_i)$

[21]Notice that for firms with expected productivity above the wage, $\bar{a}_i > w$, higher productivity increases the firms' incentive to cluster. However, the opposite holds for firms with expected productivity below the cluster wage. Taken together, this implies that the incentives to cluster are the lowest for firms with expected productivity that equals the wage in the cluster.

[22]Formally, $\frac{d(V_i^C - V_i^I)}{dw} = \frac{dV_i^C}{dw} = w\frac{dh_{1i}^*}{dw} + E\left[h_{2i}^* - h_{1i}^*\right] = -\frac{w}{\tau} - \frac{\bar{a}_i - w}{\beta}$. An increase in the wage reduces the creation of human capital at $t = 1$ (first term) and increases the benefits of "selling" human capital at $t = 2$ (second term). Therefore, if a firm sells enough human capital (i.e., if $\frac{w - \bar{a}_i}{\beta}$ is large enough), a firm can benefit from a higher wage.

is the frequency of $\bar{a}_i$ in the population. We assume that $\overline{a}_i$ is positive and with bounded support, i.e., $\overline{a}_i \in [\overline{a}_L, \overline{a}_H]$. We consider equilibria characterized by the formation of a unique cluster C which may contain a positive mass of firms, with the rest of the firms choosing isolated locations.[23] Under these conditions, because shocks are independent across firms, the average productivity in the cluster (i.e., the cluster wage) is deterministic and given by $w = \bar{a}_C = \int_{i \in C} \bar{a}_i g(\bar{a}_i)$.

The following proposition, which is proved in the appendix, describes the various equilibria that are possible in this setting:

Proposition 3 *The following types of equilibria can emerge in the above described setting: (i) an equilibrium where all firms are isolated, (ii) an equilibrium where all firms locate within a single cluster, and (iii) an equilibrium where firms with both the highest and lowest expected productivities join a cluster and where the middle firms locate in isolation. The equilibrium that emerges depends on the population of firm characteristics, however, for certain firm characteristics, multiple equilibria arise.*

Proposition 3 states that there may be multiple equilibria where only those firms with the largest and the smallest future expected productivities cluster. In the appendix, we provide a simple example in which two alternative clusters of different sizes and wages can emerge in equilibrium. Specifically, if firms expect a low wage at $t = 2$ in the cluster, most firms join the cluster and the cluster wage is indeed low. Alternatively, if firms expect a high $t = 2$ cluster wage, fewer firms join the cluster, and the cluster wage ends up being high. In the example, the low-wage, larger cluster dominates (in a Pareto sense) the high-wage, smaller cluster, suggesting that policy

[23]For simplicity, we do not consider the possibility that more than one cluster can simultaneously arise. This would simply complicate the analysis without providing additional insights. In fact, in the case of multiple coexisting clusters, it is easy to show that all of them would have the same average productivity (i.e., cluster wage) and would exhibit the same properties as the ones described below.

initiatives that promote larger clusters can be welfare-improving.

The possibility of (self-fulfilling) multiple equilibria suggests a potential role for public intervention.[24] Specifically, in the example, a contingent policy of wage subsidies (e.g., an offer to subsidize wages in the event that equilibrium wages exceed the level that occurs in the good equilibrium) will attract more of the low expected productivity firms to the cluster, which in turn generates lower wages, so that the subsidy will not in fact be required.[25]

5 Location analysis when workers also invest

Up to now, we have focused on the case where training is not affected by the workers' effort. This section extends the analysis to consider the location choice when both workers and firms can affect the creation of human capital. Formally, we assume that during the initial stage, each worker can exert costly effort l, which enhances the human capital provided by the firm. Specifically, we assume that effort multiplies the worker's acquired human capital by a factor $(1+l)$ and costs the worker $C(l) = k\frac{l^2}{2}$. Hence, each worker solves $\max_l (w\,(1+l) - k\frac{l^2}{2})$ and thus exerts effort,

$$l^* = \frac{w}{k}. \tag{24}$$

In addition, as in the basic case, we assume that the investment in human capital at $t=1$ (by both the firm and its workers, i.e., $(1+l)h_{1i}^l$) affects the firm's adjustment

[24] Public intervention may be required to the extent that the private sector fails to solve the coordination problem. See Rauch (1993) for an analysis of the role of history in the location of industrial clusters and how developers of industrial parks can partly overcome historical inertia.

[25] A number of individuals have argued for policy initiatives that promote clustering in knowledge based industries. For example, Lawrence Summers, the president of Harvard University expressed the following opinion on the importance of promoting clustering at the Massachusetts Life Sciences Summit held in Boston on 09/12/2003: "I am convinced that, as strong as (...) the life science cluster is today without combined efforts, it can be far stronger five years from now and still stronger a decade from now. And with all of our cooperation, Harvard is certainly prepared to do its part. I believe we can do a great deal for science, for humanity, and for the economy of this area."

costs at $t = 2$. Formally, this implies that firm i's production function at $t = 2$ is

$$Q_2 = a_i h_{2i} - \beta \frac{(h_{2i}^l - (1+l)h_{1i}^l)^2}{2}, \tag{25}$$

where we have used the super-index l to distinguish this case from the case in which workers' effort cannot affect their human capital.

5.1 Homogeneous firms

We consider in this sectionthe case where firms have identical distributions of their future productivities, i.e., $F_i(\tilde{a}_i) = F(\tilde{a}_i)$ for all i. Because in isolation firm's monopsony power allows them to capture all the benefits from the workers' human capital, workers have no incentives to contribute to their training (i.e., $l^* = 0$). Therefore, the value of the isolated firm, $V_{i,l}^I$, is still given by V_i^I from equation (6) i.e., the value in isolation when workers' effort does not affect the creation of human capital:

$$V_{i,l}^I = V_i^I = \frac{\alpha^2}{2\tau} + \frac{\alpha \bar{a}}{\tau} + \frac{\bar{a}^2}{2\tau}. \tag{26}$$

In clusters, however, since workers capture the benefits of the increase in their human capital, they do have an incentive to exert effort. In this case, firms solve at $t = 2$ the following problem:

$$\max_{h_{2i}^l} \; a_i h_{2i}^l - \beta \frac{(h_{2i}^l - (1+l)h_{1i}^l)^2}{2} - w h_{2i}, \tag{27}$$

whose f.o.c. implies that

$$h_{2i}^{l*} = (1+l)h_{1i}^l + \frac{a_i - w}{\beta}. \tag{28}$$

Because the supply of human capital per firm is $h_{2i}^l = (1+l)h_{1i}^l$, market clearing yields $w = \bar{a}$. Then, proceeding as in previous sections, we can solve firm i's problem at $t = 1$, (i.e., $\max_{h^l_{1i}} \alpha h^l_{1i} - \tau \frac{(h^l_{1i})^2}{2}$) to obtain $h^{l*}_{1i} = \frac{\alpha}{\tau}$.[26] Substituting h^{l*}_{1i}, h^{l*}_{2i} and l^* in the firm's objective function, we obtain the firm value in clusters, $V^C_{i,l}$:

$$V^C_{i,l} = \frac{\alpha^2}{2\tau} + \frac{\alpha \bar{a}}{\tau} + \frac{\sigma^2}{2\beta} + \frac{\alpha}{k} \frac{\bar{a}^2}{2\tau}, \tag{29}$$

and subtracting (26) from (29), we obtain the net benefit of clustering:

$$V^C_{i,l} - V^I_{i,l} = \frac{\sigma^2}{2\beta} - \frac{\bar{a}^2}{2\tau} \left(1 - \frac{\alpha}{k}\right). \tag{30}$$

From (30), we can make two observations. First, in addition to the factors previously identified (i.e., σ, β, τ, and $\bar{a}$), two other factors affecting firm location arise (i.e., α and k). In particular, clustering is more valuable when workers contribute substantially to the creation of human capital (i.e., low k and hence high l^*) and also when firms invest heavily in training at $t = 1$ (i.e., high α and hence high h^{l*}_{1i}). This last result holds because of the complementarity between firm training and worker effort, i.e., because workers' effort multiplies the human capital provided by firms.[27]

Second, we can identify a sufficient condition for clustering to dominate isolation (i.e., $\alpha > k$). This condition illustrates an additional rationale for firms to cluster even when there is (i) no uncertainty about firm productivity (i.e., $\sigma^2 = 0$) and (ii) no ex-ante differences in productivity (i.e., $F_i(\tilde{a}_i) = F(\tilde{a}_i)$). Intuitively, clusters induce workers to contribute to the creation of human capital by mitigating workers' hold-up concerns. This happens because, in contrast to isolation, where firms capture the gain associated with the additional human capital created by workers, in clusters workers with more human capital receive higher wages at $t = 2$.[28]

---

[26]Notice that workers' effort does not affect the firm's investment, i.e., $h^{l^*}_{1i} = h^*_{1i} = \frac{\alpha}{\tau}$ as given by (15). This is in contrast with the result obtained in the next subsection, when differences in productivity across firms are introduced.

[27]A recent paper by Rosenthal and Strange (2002) shows that, after controlling for worker specific characteristics, professional workers work longer hours in urban clusters which suggests that, consistent with the analysis here, cities encourage hard work.

[28]We are not the first to identify the role of the markets in mitigating hold-up problems. This

5.2 Heterogeneous firms

This section extends the previous analysis to the case of heterogeneous firms (i.e., to a setting like the one in section 4.2). The main purpose is to examine whether the ability of workers to contribute to their human capital can have different effects on the incentives of heterogeneous firms to cluster.

In isolation, the analysis remains unchanged because workers do not have incentives to invest in human capital. In clusters, however, the analysis changes. Substituting h_{2i}^{l*} (given by (28)) into the firm's objective function at $t = 1$ and simplifying the following problem for firm i at $t = 1$:

$$\max_{h_{1i}^{l}} \alpha h_{1i}^{l} - \tau \frac{(h_{1i}^{l})^2}{2} + (\bar{a}_i - w)(1 + l^*)h_{1i}^{l}. \tag{31}$$

Notice that now, in contrast to the case with identical firms, workers' incentives affects the firm's incentive to create human capital at $t = 1$ (i.e., note the factor $(1 + l^*)$). The reason is that workers' investment in human capital, because of the adjustment costs, increases the firm's demand for skilled labor at $t = 2$.[29] Solving we find,

$$h_{1i}^{l*} = \frac{\alpha + (\bar{a}_i - w)(1 + l^*)}{\tau}, \tag{32}$$

and the firm value in the cluster, $V_{i,l}^{C}$, which can be expressed as:

$$V_{i,l}^{C} = V_i^{C} + \left[l^* \cdot h_{1i}^{l*} \left(\bar{a}_i - \frac{w}{2} \right) \right] + \left[\left(Q(h_{1i}^{l*}) - Q(h_{1i}^{*}) \right) + (h_{1i}^{l*} - h_{1i}^{*})\bar{a}_i \right], \tag{33}$$

where h_{1i}^{*} and V_i^{C} are, respectively, the demand for labor and value of the clustered firm when workers cannot invest in human capital, i.e., expressions (20) and (21).

role has been recently considered in the Urban Economics literature, i.e., Rotemberg and Saloner (2000), and Matouschek and Robert-Nicoud (2003) and in the International Trade literature, i.e., McLaren (2000), and Grossman and Helpman (2002).

[29]Note that the adjustment costs, i.e., $-\beta \frac{(h_{2i}^{l} - (1+l)h_{1i}^{l})^2}{2}$ depends on workers and firm investments in human capital (i.e., l and h_{1i}^{l} respectively). Hence, as shown in (28), $h_{2i}^{l*} = (1 + l)h_{1i}^{l} + \frac{a_i - w}{\beta}$.

From expression (33), we can see the difference in value arising from the workers' investment in human capital. We denote this difference as $\Delta V_{i,l}^C \equiv (V_{i,l}^C - V_i^C)$ and refer to it as the "*effort-effects.*" Specifically, these effects are: (i) a *direct* effect due to the additional creation of human capital by workers (i.e., the term in the first bracket)[30] and (ii) an *indirect* effect stemming from the adjustment of the firm's demand for labor at $t = 1$ induced by workers' effort (i.e., term in the second bracket).[31] Expressing the effort-effects as

$$\Delta V_{i,l}^C = \frac{\alpha w}{\tau k}(\bar{a}_i - \frac{w}{2}) + \left[\frac{(\bar{a}_i - w)w}{2\tau k}\right]\left[\frac{\bar{a}_i w}{k} + (2\bar{a}_i + w)\right], \tag{34}$$

reveals that $\Delta V_{i,l}^C$ is positive for $\bar{a}_i > w$, negative for $\bar{a}_i < \frac{w}{2}$, and ambiguous for the remaining range. It can also be easily shown that $\Delta V_{i,l}^C$ increases with the firm's expected productivity $\bar{a}_i$. Hence, the following proposition can be stated:

Proposition 4 *Workers' investment in human capital (i.e., the effort-effects) increase the incentives of high productivity firms ($\bar{a}_i > w$) to cluster and of low productivity firms ($\bar{a}_i < \frac{w}{2}$) to isolate. Furthermore, the effort-effects increase with a firm's productivity (i.e., $\frac{d(\Delta V_{i,l}^C)}{d\bar{a}_i} > 0$).*

While it is easy to understand why the investment in human capital can promote clustering (in isolation workers do not invest in human capital), it is less straightforward to understand why it can lead firms to isolate. The potential incentive to isolate occurs because workers do not in general exert the effort most preferred by firms, and in some cases, exert too much effort. To be sure, while workers make their effort

[30]This term corresponds to the additional human capital produced by workers, $l \cdot h_{1i}^{l*}$, multiplied by the expected profits of per unit of human capital, (i.e., $\bar{a}_i - k\frac{l^*}{2} = \bar{a}_i - \frac{w}{2}$). The expected profits per unit consist of what the firm expects to obtain at $t = 2$, (i.e., $\bar{a}_i - w$) plus the reduction in its labor costs at $t = 1$ (i.e., $w - k\frac{l}{2}$).

[31]Notice that, due to the effort-effects, both h_{1i}^{l*} and (due to the adjustment costs) $E(h_{2i}^{l*})$ change by $\frac{(\bar{a}_i - w)l^*}{\tau}$. These adjustments yield an additional revenue to the firm of $Q(h_{1i}^{l*}) - Q(h_{1i}^*) = \frac{2l^* + {l^*}^2}{2\tau}(\bar{a}_i - w)^2$ at $t = 1$, and of $(h_{1i}^{l*} - h_{1i}^*)\bar{a}_i = \frac{(\bar{a}_i - w)l^*}{\tau}\bar{a}_i$ at $t = 2$.

decision at $t = 1$ based on the wage they expect to receive (i.e., $\max_l(wl - k\frac{l^2}{2})$), firms expect to receive from that effort only the firm's expected productivity, $\bar{a}_i$ (net of the cost of effort) at $t = 2$ (i.e., $\bar{a}_i l - k\frac{l^2}{2}$).[32] Specifically, if $\bar{a}_i < w$ ($\bar{a}_i > w$) workers' effort is too large (too small) from the point of view of the firm. Furthermore, when the firm's expected productivity is small enough (i.e., $\bar{a}_i < \frac{w}{2}$), workers' effort actually reduces the value of the firm (i.e., the firm would be better off if workers exert no effort rather than l^*).

The prior discussion focuses on the benefits of the effort-effects on clustering.[33] To examine how workers' effort affect the comparative statics on firm location discussed in section 4, we need to consider the other factors that influence clustering and how they interact with the effort-effects. For brevity, we focus next on the effect of a firm's expected productivity on location.

As proposition 4 states, workers' effort induces high productivity firms ($\bar{a}_i > w$) to cluster, and low productivity firms ($\bar{a}_i < w$) to isolate. In contrast, as result (iv) in proposition 2 shows, without workers' effort, the incentive to join the cluster increases for firms with extreme (i.e., very high and very low) expected productivity (i.e., $\frac{d(V_i^C - V_i^I)}{d\bar{a}_i} = \frac{\bar{a}_i - w}{\beta}$) so that, clusters exhibit a U-shape. The joint consideration of these two results implies that workers' effort reinforces the tendency of high productivity firms ($\bar{a}_i > w$) to cluster and weakens the tendency of low productivity firms ($\bar{a}_i < w$) to do so.[34] In fact, as the following proposition states, under certain conditions, "effort-effects" dominate and reverse the tendency of low productivity firms to join the cluster:

---

[32]Notice that, due to the adjustment costs, worker effort increases the firm's demand for skilled labor at $t = 2$, which has an expected productivity of $\bar{a}_i$.

[33]Effort-effects do not affect isolated firms, since $V_{i,l}^I = V_i^I$. Hence, $\Delta V_{i,l}^C$ is also a measure of how these effort-effects affect firms clustering vis-a-vis isolation, i.e., $(V_{i,l}^C - V_{i,l}^I) - (V_i^C - V_i^I) = \Delta V_{i,l}^C$.

[34]That is, if $\bar{a}_i > w$, $\frac{d(V_{i,l}^C - V_{i,l}^I)}{d\bar{a}_i} > \frac{d(V_i^C - V_i^I)}{d\bar{a}_i} > 0$ while if $\bar{a}_i < w$, $\frac{d(V_{i,l}^C - V_{i,l}^I)}{d\bar{a}_i} > \frac{d(V_i^C - V_i^I)}{d\bar{a}_i}$ although $\frac{d(V_{i,l}^C - V_{i,l}^I)}{d\bar{a}_i}$ may still be negative.

Proposition 5 *For a sufficiently large α, the incentive to cluster increases with the firm expected productivity,* $\frac{d(V_{i,l}^C - V_{i,l}^I)}{d\bar{a}_i} > 0$.

The intuition for the proposition is as follows: As discussed above, workers fail to consider the full effects of their effort on firm value, so firms can even experience a reduction in value due to their workers' effort. This externality is more pronounced when α is larger because, in this case, firms choose a higher h_{1i}^{l*}, and workers choose a higher effort, i.e., $l^* h_{1i}^{l*}$.

The discussion above, although made in the context of a given wage (i.e., partial equilibrium) has implications on the issue of the endogenous formation of clusters. In contrast to the case without effort-effects, where firms with low as well as high productivity have the highest incentives to cluster, effort-effects produce an additional impetus for firms with the highest productivity to join the cluster. This suggests that the empirical evidence on the benefits of clusters should be cautiously interpreted. For instance, an alternative interpretation to the findings by Henderson (1986) and Ciccone and Hall (1996), that productivity increases with the density of the economic activity and by Holmes and Stevens (2002), that plant sizes are higher within industry clusters, is that clusters attract the most efficient firms, rather than make existing firms more efficient.

We finish this section with a brief mention of one important assumption that we maintain throughout the analysis: the fact that we do not allow long-term contracts between workers and firms. While realistic legal reasons can justify this assumption, this section provides an additional justification for the exclusion of long-term contracts. Consistent with Grossman and Hart (1986), the impossibility to contract in the actual provision of human capital (both by firms and by workers) can severely reduce the ability of long-term contracts (e.g., a guaranteed wage set at $t = 1$ in exchange for the worker's labor services at $t = 2$) to induce the efficient human capital investment. In this setting, unless workers receive some benefits at the margin from their investments in human capital, they fail to provide effort and, hence, a suboptimal creation of human capital would prevail.[35]

6 Aggregate uncertainty, firm-specific risk and technology standards

So far, we have considered firm specific shocks and hence, our economy has been characterized by no aggregate uncertainty. In this section, we first introduce uncertainty about aggregate productivity (i.e., systematic shocks), and then we allow firms to design their production processes in ways that make them more or less sensitive to these systematic shocks. To simplify the exposition, we perform the analysis in the basic setting of sections 3 and 4 which abstracts from workers' effort in the creation of human capital.

6.1 Location and aggregate uncertainty

We introduce aggregate uncertainty (i.e., correlated productivity shocks) by modeling a firm's random productivity at $t = 2$ as:

$$\tilde{a}_i = \bar{a} + \gamma^{1/2}\tilde{v} + (1-\gamma)^{1/2}\tilde{b}_i \tag{35}$$

where $\bar{a} > 0$ is deterministic and $\tilde{v}$, a systematic shock, and $\tilde{b}_i$, a firm-idiosyncratic shock, are two independent random variables.[36] We assume that $0 \leq \gamma \leq 1$, $E(\tilde{b}_i) =$

[35] The analysis also suggests that there may be differences in how workers are compensated in clusters vis-a-vis isolated firms. In isolation, inducing worker effort is the more important incentive problem, suggesting that it may be optimal to consider incentive compensation contracts. In clusters, worker retention is the more important concern for firms, so we may expect to see longer-term compensation contracts to address this issue. Although we have abstracted from these issues in our model, these compensation issues would be interesting to explore in future work.

[36] Our previous analysis can be seen as a particular case of this model in which $\gamma = 0$.

$E(\tilde{v}) = 0$, $Var(\tilde{v}) \equiv \sigma_v^2$, and $Var(\tilde{b}_i) \equiv \sigma_b^2$ (and hence, $V(\tilde{a}_i) \equiv \sigma^2 = \gamma\sigma_v^2 + (1-\gamma)\sigma_b^2$). In addition, we further decompose the systematic shock $\tilde{v}$ into

$$\tilde{v} = \theta^{1/2}\tilde{v}_1 + (1-\theta)^{1/2}\tilde{v}_2 \tag{36}$$

where $0 \leq \theta \leq 1$, $E(\tilde{v}_1) = E(\tilde{v}_2) = 0$ and $Var(\tilde{v}_1) \equiv \sigma_{v1}^2$ and $Var(\tilde{v}_2) \equiv \sigma_{v2}^2$. We assume that the realization of $\tilde{v}_1$ (i.e., v_1) is known to firms after they locate at $t = 0$ but *before* they make their training decisions at $t = 1$ (i.e., h_{1i}). However, v_2, the realization of $\tilde{v}_2$, is known by firms only *after* they have chosen h_{1i}. The parameter θ measures how much of the aggregate shock is known by firms before they make their training choices.

As before, a firm's location choice boils down to a comparison of its value in isolation and in the cluster. In isolation, the human capital created at $t = 1$, h_{1i}, depends on the firm's expected productivity at $t = 2$, which, for a given realization of $\tilde{v}_1$ is: $E(\tilde{a}_i|v_1) = \bar{a} + \theta\gamma v_1$. Hence, firm value conditional on v_1 (i.e., $V_i^I|v_1$) can be obtained after substituting $E(\tilde{a}_i|v_1)$ for $\bar{a}$ in (6)

$$V_i^I|v_1 = \frac{[\alpha + E(\tilde{a}_i|v_1)]^2}{2\tau} = \frac{(\alpha + \bar{a} + \theta^{1/2}\gamma^{1/2}v_1)^2}{2\tau}. \tag{37}$$

The (unconditional) firm value can then be determined by integrating over all possible values of v_1:

$$V_i^I = \frac{(\alpha + \bar{a})^2}{2\tau} + \frac{\theta\gamma\sigma_{v1}^2}{2\tau}. \tag{38}$$

Equation (38) shows that although, an isolated firm finds the distinction between aggregate and firm-specific uncertainty irrelevant, the firm achieves a higher value (i.e., the term $\frac{\theta\gamma\sigma_{v1}^2}{2\tau}$ in (38)) when it can incorporate more information (i.e., the shock v_1) into their investment decision, h_{1i}.

In clusters, however, the distinction between aggregate and firm-specific risk is relevant because aggregate shocks, since they affect all firms, influence the wage at

$t = 2$. To find the value of a clustered firm, we first obtain firm and wage values conditional on given realizations of v_1 and v_2, and then, we successively integrate over v_2 and v_1 to find the ex-ante (unconditional) firm and wage values.

For a given v_1 and v_2, market clearing implies that $w|v_1, v_2 = \bar{a} + \gamma(\theta^{1/2}v_1 + (1-\theta)^{1/2}v_2)$. Furthermore, the value of a firm with realized productivity a_i (i.e., v_1, v_2, b_i) is

$$V_i^C|v_1, v_2, b_i = \frac{\alpha^2}{2\tau} + \frac{\alpha(\bar{a} + \theta^{1/2}\gamma^{1/2}v_1)}{\tau} + \frac{(1-\gamma){b_i}^2}{2\beta}. \tag{39}$$

Taking expectations over v_2 and b_i, we find that the expected wage is $E(w|v_1) = \bar{a} + \theta^{1/2}v_1$ and that the firm value conditional on v_1 is:

$$V_i^C|v_1 = \frac{\alpha^2}{2\tau} + \frac{\alpha(\bar{a} + \theta^{1/2}\gamma^{1/2}v_1)}{\tau} + \frac{(1-\gamma)\sigma_b^2}{2\beta}. \tag{40}$$

Finally, by integrating over v_1 we find the ex-ante wage, $E(w) = \bar{a}$, and firm value,

$$V_i^C = \frac{\alpha^2}{2\tau} + \frac{\alpha\bar{a}}{\tau} + \frac{(1-\gamma)\sigma_b^2}{2\beta}. \tag{41}$$

Examining (41), note that, in contrast to the case of isolation, firms in clusters do not take advantage of the early release of information about the aggregate shock, v_1, and, as a result, firm value is not affected by $\tilde{v}_1$. Notice that in the cluster, the creation of human capital at $t = 1$ is independent of the aggregate productivity shock (i.e., $h_{1i} = \frac{\alpha}{\tau}$). While a positive shock in a firm's expected productivity increases, *all else equal*, its creation of human capital, a positive aggregate shock also increases the wage at $t = 2$ (i.e., $w = \bar{a} + v_1$) which reduces the firm's incentives to create human capital.[37] In other words, as (20) shows, the creation of human capital by a clustered

[37]Formally, $h_{1i} = \frac{\alpha}{\tau} + E(\tilde{a}_i|v_1) - E(w|v_1) = \frac{\alpha}{\tau}$. The fact that $E(\tilde{a}_i|v_1) = E(w|v_1)$ is an artifact of the production function that we consider. However, the presence of two offsetting effects is quite general: an increase in expected productivity tends to increase wages and, hence, to discourage firms' creation of human capital.

firm is not determined by its expected productivity, but by the difference between expected productivity and the cluster wage, (i.e., $h_{1i}^* = \frac{\alpha+(\bar{a}_i-w)}{\tau}$).

Also, from (41), notice that the value of clustered firms depends on the firm-specific variance, σ_b^2, but not the variance of the systematic shock, σ_v^2. This is because clusters enhance firm value by reallocating human capital from firms with low productivity shocks to firms with high productivity shocks. In the limiting case where the shocks are perfectly correlated, i.e., $\gamma = 1$, there would be no reallocation of human capital in the cluster and, hence, no benefit to clustering.

To summarize, one can combine (38) with (41) and express the gains to cluster as

$$V_i^C - V_i^I = \frac{(1-\gamma)\sigma_b^2}{2\beta} - \frac{\theta\gamma\sigma_{v1}^2}{2\tau} - \frac{\bar{a}^2}{2\tau}, \tag{42}$$

which leads us to state the following proposition:

Proposition 6 *For a given level of total risk, the value of clustering vis-a-vis isolation increases with the relative importance of firm-specific risk (low γ) and the level of the aggregate risk that is not anticipated (low θ).*

Proposition 6 suggests that, empirically, clusters are more valuable in industries in which firms' productivity experience highly idiosyncratic shocks (i.e., high σ_b^2), and in which aggregate industry productivity is difficult to predict (i.e., low θ).

6.2 Firm-specific risk and technological standards

In this section, we endogenize the technology choice and consider choices involving the sensitivities of technologies to firm-specific and systematic risks. Specifically, we assume that firms can increase their exposure to firm-specific risk (and decrease systematic risk) by selecting a production process that deviates somewhat from what we will refer to as "the standard production process." As we have shown, in clusters, there is a benefit associated with increasing firm-specific risk that can lead clustered firms to deviate from the industry standard. Offsetting this benefit is the possibility that the adjustment costs associated with transferring workers from one firm to another are higher if a firm chooses a less standard production process.

To examine the choice of technological standards, we endogenize the parameter that captures the intensity of the adjustment costs at $t = 2$ (i.e., β_i). In particular, we assume that a higher β_i is associated with a less standard production process and therefore, with an increase in the firm's exposure to idiosyncratic risk, i.e., in (35) we set $(1-\gamma) = g(\beta_i)$ where $0 < g(\beta_i) < 1$, $g' > 0$ and $g'' < 0$, and hence:

$$\tilde{a}_i = \bar{a} + (1 - g(\beta_i))^{1/2}\tilde{v} + g(\beta_i)^{1/2}\tilde{b}_i, \tag{43}$$

where, as before, $\tilde{v}$ is a systematic shock and $\tilde{b}_i$ is a firm specific shock. Shocks $\tilde{v}$ and $\tilde{b}_i$ are independent, with zero mean, and variances σ_v^2 and σ_b^2 respectively, and therefore,

$$\sigma_i^2 = [1 - g(\beta_i)]\sigma_v^2 + g(\beta_i)\sigma_b^2. \tag{44}$$

Firms choose their technology (i.e., β_i) after locating, but before starting the initial stage of production. For simplicity, we assume that both the industry and the firm-specific shocks are unanticipated, i.e., $\theta = 0$.

Consistent with our analysis in section 6.1, in isolation, firm value is $V_i^I = \frac{(\alpha+\bar{a})^2}{2\tau}$, which is not affected by adjustment costs and the choice of risk.[38] However, in clusters, these choices do affect firm value:[39]

$$V_i^C = \frac{\alpha^2}{2\tau} + \frac{\alpha\bar{a}}{\tau} + \frac{g(\beta_i)\sigma_b^2}{2\beta_i}. \tag{45}$$

[38] This is because (i) in isolation, firms do not adjust their scale at $t = 2$ and (ii) the shocks (firm-specific and aggregate) are unanticipated.

[39] It is easy to check that, in this case, $w = \bar{a}$, $h_{1i}^* = \frac{\alpha}{\tau}$, and $h_{2i}^* = h_{1i}^* + \frac{a_i - w}{\beta_i}$.

Therefore, comparing firm values in clusters vis-à-vis in isolation, we get:

$$V_i^C - V_i^I = \frac{g(\beta_i)\sigma_b^2}{2\beta_i} - \frac{\bar{a}^2}{2\tau}. \tag{46}$$

As (46) shows, a more idiosyncratic technology, (i.e., a higher β_i) produces two opposing effects: (i) it makes the redeployability of human capital costlier and hence, reduces the value of locating in the cluster and (ii) it makes the possibility of redeploying human capital more valuable (i.e., higher idiosyncratic variance $g(\beta_i)\sigma_b^2$) and hence, increases the value of clustering. From the f.o.c. in (45) (i.e., $\max_{\beta_i} V_i^C$), we obtain:

$$\beta^* g'(\beta^*) - g(\beta^*) = 0. \tag{47}$$

As a final observation notice that previous results suggest that, empirically, the choice of technology by a firm is correlated with its location. This is because while idiosyncratic risk and low adjustment costs increase the real option value of being in the cluster, they do not have any impact on isolated firms. Therefore, if there is, say, a trade-off between the mean and variance of different technologies, an isolated firm will choose the technology that maximizes expected productivity, while clustered firms will face a trade-off between expected productivity and specificity (i.e., idiosyncratic risk and adjustment costs).[40] Specifically, firms within a cluster are willing to take on more risk, subject to being not too incompatible with their competitors.

7 Overconfidence and the location choice

Up to this point we have assumed that entrepreneurs make rational location choices. However, there is substantial evidence in the psychology literature that suggests that

[40] Isolated firms may actually choose technologies that are more or less risky than the technologies chosen by clustered firms. Firms within clusters benefit from (idiosyncratic) risk but also, from being compatible with other firms within the cluster.

individuals are overconfident about their abilities (e.g., Einhorn 1980), and there is an extensive literature that explores the implications of overconfidence on economic behavior.

To explore the effect of overconfidence on the location choice, we extend the basic model from sections 2 and 3 to allow for the possibility that entrepreneurs have biased perceptions of their firms' expected productivity. Specifically, at $t = 0$, entrepreneur i wrongly believes that firm i's expected productivity is above the average productivity of the economy, $\bar{a}_i^o = \bar{a} + \Delta$ with $\Delta > 0$, but correctly believes that the rest of the firms have the same expected productivity $\bar{a}$. Workers understand that entrepreneurs are overconfident and act accordingly.[41]

In isolation, the analysis with overconfidence corresponds to the analysis in the basic model with productivity $\bar{a}_i^o$, so $h_2^* = h_1^*$ and $h_1^* = \frac{\alpha+\bar{a}_i^o}{\tau}$. Hence, from (6), the "perceived" firm value at $t = 0$, $V_{i,o}^I$ can be expressed as:

$$V_{i,o}^I = V_i^I + (\alpha + \bar{a} + \frac{\Delta}{2})\frac{\Delta}{\tau}, \tag{48}$$

where $V_i^I = \frac{\alpha^2}{2\tau} + \frac{\alpha\bar{a}}{\tau} + \frac{\bar{a}^2}{2}$. We can also compute the firm value under the "true" distribution of firm productivity, $V_{i,T}^I$:

$$V_{i,T}^I = V_i^I - \frac{\Delta^2}{2\tau}. \tag{49}$$

Notice that while overconfidence increases the perceived value (which is the value on which the firm bases its location decision) $V_{i,o}^I > V_i^I$, it induces firms to overinvest in human capital and hence, reduces the true firm value (which we use to analyze the welfare implications of the location choice), $V_{i,T}^I < V_i^I$.

In clusters, overconfidence increases the creation of human capital by firms, i.e.

[41] In this setting, it is irrelevant what a given firm i thinks about other firms' overconfidence. As we show below, this is because w, the only variable that may alter firms' decisions, is not affected by firms' overconfidence.

$h_{1i}^* = \frac{\alpha+\Delta}{\tau}$ and the demand for skilled workers at $t = 2$, i.e., $h_{2i}^* = h_{1i}^* + \frac{a_i - w}{\beta}$.[42] Notice that, at $t = 2$, overconfidence increases the aggregate supply and demand of skilled labor at $t = 2$ by the same amount and, hence, leaves unaffected the cluster wage, $w = \bar{a}$.[43] Therefore, substituting previous expressions in (16), we obtain

$$V_{i,o}^C = V_i^C + \frac{\Delta}{\tau}(\alpha + \bar{a} + \Delta) + \frac{\Delta^2}{2}\left(\frac{1}{\beta} - \frac{1}{\tau}\right), \tag{50}$$

where $V_i^C = \frac{\alpha^2}{2\tau} + \frac{\alpha\bar{a}}{\tau} + \frac{\sigma^2}{2\beta}$, and computing the expectation of firm value under the true distribution of firm productivity, we get

$$V_{i,T}^C = V_i^C + (\bar{a} - \frac{\Delta}{2})\frac{\Delta}{\tau}. \tag{51}$$

Expression (50) shows that, as in isolation, overconfidence increases perceived value, $V_{i,o}^C > V_i^C$. Furthermore, in contrast to isolation, $V_{i,T}^C > V_i^C$ for moderate overconfidence, i.e., $\Delta < 2\bar{a}$, and $V_{i,T}^C < V_i^C$ for excessive overconfidence, i.e., $\Delta > 2\bar{a}$. Intuitively, overconfidence allows the firm to commit to training more workers and hence, ameliorates the problem of underinvestment of human capital. However, excessive overconfidence can lead to the creation of too much human capital and reduce firm value.

Given the above analysis, it is straightforward to show that overconfidence increases the tendency of firms to cluster. Specifically,

$$(V_{i,o}^C - V_i^C) - (V_{i,o}^I - V_i^I) = \frac{\Delta^2}{2\beta} > 0. \tag{52}$$

This is consistent with results from previous sections which indicate that firms with (expected) productivity above the cluster average have an incentive to join the cluster. In addition, overconfidence has welfare effects. We summarize all these results in the following proposition:

[42] $h_{2i}^* = h_{1i}^* + \frac{a_i - w}{\beta}$ follows immediately from (8) and $h_{1i}^* = \frac{\alpha+\Delta}{\tau}$ follows from (20).

[43] Although the wage remains unchanged, the linkages between periods that appear as a consequence of adjustments costs have important effects on the welfare results discussed below.

Proposition 7 *Assume that without overconfidence firms cluster, (i.e., $\frac{\sigma^2}{\beta} > \frac{\bar{a}^2}{\tau}$), then overconfidence increases welfare (i.e., true firm value) if and only if $\Delta < 2\bar{a}$. Assume alternatively that without overconfidence firms isolate (i.e., $\frac{\sigma^2}{\beta} < \frac{\bar{a}^2}{\tau}$) and define $\Delta_0 \equiv (\frac{\beta\bar{a}^2}{\tau} - \sigma^2)^{\frac{1}{2}}$ and $\Delta_1 \equiv 2\bar{a} - (\frac{\bar{a}^2}{\tau} - \frac{\sigma^2}{\beta})$ then overconfidence:*

(i) leaves firms in isolation and reduces welfare if $\Delta < \Delta_0$.

(ii) leads firms to cluster if $\Delta_0 < \Delta$, and increases welfare if and only if $\Delta_0 < \Delta < \Delta_1$.

We finish this section discussing the robustness of its main result, namely, the fact that overconfidence leads to more clustering.[44] While a comprehensive analysis of the effects of overconfidence on firm location is beyond the scope of this paper, we can examine some of the effects previously discussed. For instance, when one considers worker effort, overconfidence will enter the location choice through an additional channel. In particular, as we showed in section 5, when workers contribute to the creation of their human capital, firms with relatively high productivity benefit from being in a cluster, while firms with relatively low productivity benefit from being in isolation. In the terminology of section 5, effort-effects are positive for firms which *have or believe they have* future productivity that is above the average. This reinforces the conclusion that clusters are more likely to emerge when firms are overconfident about their future productivities.

We also could have considered an analysis in which workers, as well as entrepreneurs are subject to overconfidence. In this case, a relatively straightforward extension of our model would suggest an additional force that encourages clustering. Specifically, since high ability workers capture the gains associated with their superior human capital within a cluster, but not when they are isolated, workers who believe

[44] A parallel analysis with "underconfidence" (i.e., $\Delta < 0$) may have some similar effects on location but will not lead to value creation in clusters. The reason is that while overconfidence serves as a commitment to train workers, alleviating the time inconsistency problem, underconfidence will have the opposite effect.

that they have superior ability will prefer locating within the cluster, leading to lower expected wages that attract firms.[45]

8 Concluding remarks

In the early literature on location choice, transportation costs play a key role. Cities arise because of proximity to transportation hubs (e.g., ports) as well as to relatively immobile factors of production. While these theories are still quite important, they do not really apply to what we refer to as knowledge-based firms, which require skilled labor, very little transportation costs, and have no exogenous natural locations. There is substantial evidence that the share of aggregate output coming from knowledge-based firms continues to increase, so it is natural to ask whether this trend will have an influence on the development of our urban areas.[46]

In contrast to most of the discussion in the literature, the model developed in this paper suggests that firms in knowledge-based industries *will not necessarily choose to cluster.* Specifically, we show that the incentive for firms to locate in industry clusters is determined by how skills are developed, the nature of uncertainty, the expected growth rate of the industry and the ability of firms to expand and contract. When workers contribute to their own training, and when there is a substantial amount of firm specific uncertainty, there are likely to be substantial gains to clustering.

[45]However, not every conceivable extension of the model would lead toward a clustering effect of overconfidence. For instance, if a firm is overconfident about its ability to train workers (i.e., low τ_o) or about the industry prospects (i.e., high $\bar{a}$), then overconfidence exacerbates the *perceived* underinvestment problem in clusters (i.e., $h_i^I - h_i^C = \frac{\bar{a}_o}{\tau_o}$), and may foster isolation. We thank Ed Glaeser for suggesting these qualifications to our conclusions.

[46]Consistent with the increased importance of what we have characterized as knowledge-based firms, a recent paper by Glaeser and Kohlhase (2003) provide evidence that indicates that transportation costs for goods declined considerably in the 20th century, while transportation costs for people increased. They find that the U.S. population has moved away from transportation hubs to regions with better consumption opportunities (e.g., better weather). However, they do not examine whether cities have become more economically focused, which one would expect if knowledge-based firms tend to cluster to take advantage of knowledge spillovers.

However, in growing industries, in which firms invest substantial amounts in their workers' human capital, firms may be better off locating apart.

A number of policy issues are raised by the analysis. First, when there are economic advantages associated with clustering, there can be coordination issues that policymakers need to address. For example, our model suggests that there can be inefficient equilibria where too few firms cluster, which in turn, suggests that there may be potential gains from policies that indirectly promote clusters, perhaps by offering training subsidies. On the other hand, if entrepreneurs are overconfident there can be too much clustering, which suggests that policies that promote clustering can potentially be misguided.

Policymakers also like to think about issues relating to competing clusters. Indeed, representatives of the high tech community in Austin, Texas think about competing with the Silicon Valley cluster and the Hong Kong financial community often express concerns about their competitive advantage relative to Singapore. Since our model is restricted to equilibria with single clusters, we do not address this issue directly, however, the analysis has implications that can be extended to evaluate the competitive advantage of different clusters.[47] Specifically, the more successful clusters are likely to be those that attract firms that are different in dimensions that make them sensitive to different economic shocks, yet similar enough to share resources. In addition, clusters will be more successful in regions where workers have the ability to acquire human capital without a substantial investment by firms. Indeed, Saxenian (1994), in a study of the computer industries in Silicon Valley and Route 128, concluded that Silicon Valley was much more successful that Boston in the 1980s because cross-firm networking opportunities in Silicon Valley facilitated the reallocation of workers be-

[47]See Porter (1990) for an extensive discussion on this topic which includes both the implications on productive decisions by firms as well as on public policy.

tween firms (i.e., reduced adjustment costs) and helped the more industrious workers develop their human capital.

Finally, it should be noted that factors that determine the extent to which firms isolate and cluster in physical space can also be applied to the extent to which firms locate in other dimensions as well. For example, competitors may be better able to pool labor and other inputs if the firms have similar products, similar production processes, and similar organizational structures. We believe that the analysis in this paper can be applied to address issues relating to industrial clustering on these dimensions, as well as to a theory of corporate inertia, that would arise if corporations are reluctant to adopt innovations because of costs associated with deviating from the cluster.

9 Technical Appendix

9.1 Proofs

Proposition 1

It follows immediately from the text.

Proposition 2

Taking derivatives in (23): $\frac{d(V_i^C - V_i^I)}{d\sigma^2} = \frac{1}{2\beta} > 0$,$\frac{d(V_i^C - V_i^I)}{d\tau} = -\frac{(\bar{a}_i - w)^2 + \sigma^2}{2\beta^2} < 0$, $\frac{d(V_i^C - V_i^I)}{d\beta} = \frac{w^2}{2\tau^2} > 0$, and $\frac{d(V_i^C - V_i^I)}{d|\bar{a}_i - w|} = \frac{|\bar{a}_i - w|}{\beta} > 0$.

Proposition 3

(i) Because $\overline{a}_L \leq w \leq \overline{a}_H$, a sufficient condition for isolation is that, for the maximal differences with the wage, firms of extreme profitability (i.e., $\overline{a}_L$ and $\overline{a}_H$) find isolation desirable. That is, if $\bar{a}_i = \overline{a}_L$ and $w = \bar{a}_H$, then $\frac{(\bar{a}_L - \bar{a}_H)^2 + \sigma^2}{2\beta} - \frac{\bar{a}_H{}^2}{2\tau} < 0$ and if $\bar{a}_i = \overline{a}_H$ and $w = \bar{a}_L$, then $\frac{(\bar{a}_H - \bar{a}_L)^2 + \sigma^2}{2\beta} - \frac{\bar{a}_L{}^2}{2\tau} < 0$.

(ii) If all firms cluster, then $w = \bar{a}$. Substituting w for $\bar{a}$ in (23) yields $\frac{(\bar{a}_i - \bar{a})^2 + \sigma^2}{2\beta} - \frac{\bar{a}^2}{2\tau}$. Hence, a sufficient condition for the cluster to exist is $\frac{\sigma^2}{\beta} > \frac{\bar{a}^2}{\tau}$.

(iii) From $\frac{d(V_i^C - V_i^I)}{d\bar{a}_i} = \frac{\bar{a}_i - w}{\beta}$, it follows that if firm k (firm j) with expected productivity $\bar{a}_k < w$ ($\bar{a}_j > w$) belongs to the cluster, then firm k' (firm j') with $\bar{a}_{k'} < \bar{a}_k$ ($\bar{a}_{j'} > \bar{a}_j$) will also belong to the cluster. Therefore, a partial cluster will be composed of all the firms whose expected productivities belong to the set $[\overline{a}_L, \overline{a}_L^c] \cup [\overline{a}_H^c, \overline{a}_H]$, where $\bar{a}_L^c$ and $\bar{a}_H^c$ are the cluster limits. To characterize these cluster limits we need to solve for the $\bar{a}_i$'s that leave firms indifferent between the cluster and isolation. Setting $V_i^C - V_i^I = 0$ in (23), we get $\bar{a}_L^c(w) = w - (\frac{\beta}{\tau}w^2 - \sigma^2)^{1/2}$ and $\bar{a}_H^c(w) = w + (\frac{\beta}{\tau}w^2 - \sigma^2)^{1/2}$. Hence, an equilibrium where only firms with productivities $\overline{a}_i \notin (\bar{a}_L^c, \bar{a}_H^c)$ cluster is characterized by the simultaneous solution of the limit equations and the wage equation, i.e., $w = E\left[\bar{a}_i | \bar{a}_i \notin (\bar{a}_L^c, \bar{a}_H^c)\right]$.

Proposition 4

From (34), notice that if $(\bar{a}_i - \frac{w}{2}) < 0$ then $\Delta V_{i,l}^C < 0$, while if $(\bar{a}_i - w) > 0$ then

$\Delta V_{i,l}^C > 0$. From (33), $\frac{d\Delta V_{i,l}^C}{d\bar{a}_i} = l^* h_{1i}^{l*} + \frac{dh_{1i}^{l*}}{d\bar{a}_i} l^* \left(\bar{a}_i - \frac{w}{2}\right) - \frac{(2l^* + {l^*}^2)(\bar{a}_i - w)}{\tau} + \frac{(2\bar{a}_i - w)l^*}{\tau} = l^* h_{1i}^{l*} + \frac{(1+l^*)l^*(\bar{a}_i - \frac{w}{2})}{\tau} - \frac{(2l^* + {l^*}^2)(\bar{a}_i - w)}{\tau} + \frac{2l^*(\bar{a}_i - \frac{w}{2})}{\tau} = l^* h_{1i}^{l*} + \frac{l^*}{\tau}\left(\bar{a}_i + \frac{1+l^*}{2} w\right) > 0$.

Proposition 5

From the proof of proposition 4, $\frac{d\Delta V_{i,l}^C}{d\bar{a}_i} = l^* h_{1i}^{l*} + \frac{l^*(\bar{a}_i + \frac{1+l^*}{2} w)}{\tau}$ and taking the derivative in (23) with respect to $\bar{a}_i$, $\frac{d\left(V_i^C - V_i^I\right)}{d\bar{a}_i} = \frac{\bar{a}_i - w}{\beta}$. Adding up: $\frac{d(V_{i,l}^C - V_{i,l}^I)}{d\bar{a}_i} = \frac{d\Delta V_{i,l}^C}{d\bar{a}_i} + \frac{d\left(V_i^C - V_i^I\right)}{d\bar{a}_i}$. Furthermore, if α (and hence, h_{1i}^{l*}) is large enough then $\frac{d(V_{i,l}^C - V_{i,l}^I)}{d\bar{a}_i} > 0$.

Proposition 6

It follows immediately from the text.

Proposition 7

Because overconfidence induces clustering, firms that would cluster in the absence of overconfidence will also cluster in the presence of overconfidence. Therefore, overconfidence changes true firm value according to (51), which is positive if and only if $\bar{a} > \frac{\Delta}{2}$. However, firms that isolate in the absence of overconfidence (i.e., $V_i^C - V_i^I < 0$) may cluster due to overconfidence (i.e., $V_{i,o}^C - V_{i,o}^I > 0$). Specifically, from (52) and (17): $V_{i,o}^C - V_{i,o}^I = \frac{\sigma^2}{2\beta} - \frac{\bar{a}^2}{2\tau} + \frac{\Delta^2}{2\beta}$ which is positive if and only if: $\Delta > (\frac{\beta\bar{a}^2}{2\tau} - \sigma^2)^{1/2} \equiv \Delta_0$. Therefore, if $\Delta < \Delta_0$ firms remain isolated and, welfare is reduced (i.e., $\frac{d(V_{i,T}^I - V_i^I)}{d\Delta} < 0$). Otherwise, if $\Delta > \Delta_0$ firms cluster, and in that case: $V_{i,T}^C - V_i^I = \left(\bar{a} - \frac{\Delta}{2}\right) + \frac{\sigma^2}{2\beta} - \frac{\bar{a}^2}{2\tau} > 0 \Leftrightarrow \Delta < 2\bar{a} - (\frac{\bar{a}^2}{\tau} - \frac{\sigma^2}{\beta}) \equiv \Delta_1$.

9.2 Other technical derivations

Equation (16)

Substituting in (13) $w = \bar{a}$, $h_{2i}^* = h_{1i} + \frac{\tilde{a}_i - \bar{a}}{\beta}$, and $h_{1i}^e = h_{1i}^* = \frac{\alpha}{\tau}$, implies $V_i^C = \frac{\alpha^2}{\tau} - \frac{\alpha^2}{2\tau} + E[(\tilde{a}_i - \bar{a})(\frac{\alpha}{\tau} + \frac{\tilde{a}_i - \bar{a}}{\beta}) - \frac{(\tilde{a}_i - \bar{a})^2}{2\beta}] + \frac{\alpha\bar{a}}{\tau}$, which simplified yields (16).

Equation (19)

Substituting in (18) $w = \bar{a}$, $h_{2i}^* = h_{1i} + \frac{\tilde{a}_i - \bar{a}}{\beta}$, and $h_{1i}^e = h_{1i}^* = \frac{\alpha + \bar{a}}{\tau}$, implies $V_i^{FB} = \frac{\alpha(\alpha + \bar{a})}{\tau} - \frac{(\alpha + \bar{a})^2}{2\tau} + E[(\tilde{a}_i - \bar{a})(\frac{\alpha + \bar{a}}{\tau} + \frac{\tilde{a}_i - \bar{a}}{\beta}) - \frac{(\tilde{a}_i - \bar{a})^2}{2\beta}] + \frac{(\alpha + \bar{a})\bar{a}}{\tau}$, which simplified yields (19).

Equation (21)

Substituting in (13) $w = \bar{a}$, $h_{2i}^{*} = h_{1i} + \frac{\tilde{a}_i - w}{\beta}$, and $h_{1i}^{e} = h_{1i}^{*} = \frac{\alpha + (\bar{a}_i - w)}{\tau}$ implies, $V_i^C = \frac{\alpha(\alpha + (\bar{a}_i - w))}{\tau} - \frac{(\alpha + (\bar{a}_i - w))^2}{2\tau} + E[(\tilde{a}_i - w)(\frac{\alpha + (\bar{a}_i - w)}{\tau} + \frac{\tilde{a}_i - w}{\beta}) - \frac{(\tilde{a}_i - w)^2}{2\beta}] + \frac{(\alpha + (\bar{a}_i - w))w}{\tau}$, which simplified yields (21).

Equation (29)

Starting from $V_{i,l}^C = \alpha h_{1i}^l - \frac{\tau (h_{1i}^l)^2}{2} + E[(\tilde{a}_i - w)\, h_{2i}^{l*} - \beta \frac{(h_{2i}^{l*} - (1+l) h_{1i}^l)^2}{2}] + h_{1i}^{l^e}(w(1+l) - \frac{kl^2}{2})$ and substituting $w = \bar{a}$, $h_{2i}^{l*} = (1+l) h_{1i}^l + \frac{a_i - w}{\beta}$, $l = \frac{w}{k}$ and $h_{1i}^{l^e} = h_{1i}^{l*} = \frac{\alpha}{\tau}$ then, $V_{i,l}^C = \frac{\alpha^2}{\tau} - \frac{\alpha^2}{2\tau} + E[(\tilde{a}_i - \bar{a})(\frac{\alpha}{\tau}(1 + \frac{\bar{a}}{k}) + \frac{a_i - \bar{a}}{\beta}) - \frac{(a_i - \bar{a})^2}{2\beta}] + \frac{\alpha}{\tau}(\bar{a}(1 + \frac{\bar{a}}{k}) - \frac{k(\frac{\bar{a}}{k})^2}{2})$ which simplified yields (29).

Equation (33)

Starting as in the derivation of (29) and substituting $h_{2i}^{l*} = (1 + l^*) h_{1i}^l + \frac{a_i - w}{\beta}$ and $h_{1i}^{l^e} = h_{1i}^l$ then, $V_{i,l}^C = \alpha h_{1i}^{l*} - \frac{\tau {h_{1i}^{l*}}^2}{2} + \frac{E(\tilde{a}_i - w)^2}{2\beta} + h_{1i}^{l*}[(1 + l^*)\bar{a}_i - \frac{k \cdot {l^*}^2}{2}]$. Then, expressing $Q(h_i^{l*}) = \alpha h_{1i}^{l*} - \tau \frac{(h_{1i}^{l*})^2}{2}$ and $Q(h_i^*) = \alpha h_{1i}^* - \tau \frac{\left(h_{1i}^*\right)^2}{2}$ and simplifying we get (33).

Equation (34)

Substituting in (33), $Q(h_i^{l*}) = \alpha h_{1i}^{l*} - \frac{\tau (h_{1i}^{l*})^2}{2}$, $Q(h_i^*) = \alpha h_{1i}^* - \frac{\tau \left(h_{1i}^*\right)^2}{2}$, $h_{1i}^{l*} = \frac{\alpha + (\bar{a}_i - w)(1 + l^*)}{\tau}$, $h_{1i}^* = \frac{\alpha + (\bar{a}_i - w)}{\tau}$ and $l^* = \frac{w}{k}$, we obtain $\Delta V_{i,l}^C = \frac{w}{k} \frac{\alpha + (\bar{a}_i - w)(1 + \frac{w}{k})}{\tau} \left(\bar{a}_i - \frac{w}{2}\right) - \frac{[(\frac{w}{k})^2 + \frac{2w}{k}](\bar{a}_i - w)^2}{2\tau}] + \frac{\frac{w}{k}(\bar{a}_i - w)\bar{a}_i}{\tau}$, and simplifying we get (34).

Equation (39)

From (13), we get $V_i^C | \upsilon_1, \upsilon_2, b_i = \alpha h_{1i} - \frac{\tau h_{1i}^2}{2} + E[(\tilde{a}_i - w)\, h_{2i}^* - \beta \frac{(h_{2i}^* - h_{1i})^2}{2} | \upsilon_1, \upsilon_2, b_i] + h_{1i}^e E\left[w | \upsilon_1\right]$. Then, recognizing that $w | \upsilon_1, \upsilon_2 = \bar{a} + \gamma^{1/2}\left[\theta^{1/2}\upsilon_1 + (1 - \theta)^{1/2}\upsilon_2\right]$, $h_{2i}^* = h_{1i} + \frac{(1 - \gamma)^{1/2}\tilde{b}_i}{\beta}$, and $h_{1i}^e = h_{1i}^* = \frac{\alpha}{\tau}$, $V_i^C | \upsilon_1, \upsilon_2, b_i$ can be simplified to obtain (39).

Equation (49)

Plug $h_{2i}^* = h_{1i}^* = \frac{\alpha + \bar{a} + \Delta}{\tau}$ in (4) and simplify to get (49).

Equation (50)

Start from $V_{i,o}^C = \alpha h_{1i} - \frac{\tau h_{1i}^2}{2} + E_o\left[(\tilde{a}_i - w)\, h_{2i}^* - \beta \frac{(h_{2i}^* - h_{1i})^2}{2}\right] + h_{1i}^e w$, and insert $w = \bar{a}$, $h_{2i}^* = h_{1i} + \frac{(\tilde{a}_i - w)}{\beta}$, $h_{1i}^e = h_{1i}^* = \frac{\alpha + \Delta}{\tau}$, and $E_o\left(\tilde{a}_i\right) = \bar{a} + \Delta$ to get, $V_{i,o}^C = \frac{\alpha(\alpha + \Delta)}{\tau} - \frac{(\alpha + \Delta)^2}{2\tau} +$

$E_o[(\tilde{a}_i - \bar{a})(\frac{\alpha+\Delta}{\tau} + \frac{\tilde{a}_i-\bar{a}}{\beta}) - \frac{(\tilde{a}_i-\bar{a})^2}{2\beta}] + \frac{(\alpha+\Delta)\bar{a}}{\tau}$, which simplified is (50).

Equation (51)

Plug $w = \bar{a}$, $h^*_{2i} = h_{1i} + \frac{\tilde{a}_i-\bar{a}}{\beta}$, and $h^e_{1i} = h^*_{1i} = \frac{\alpha+\Delta}{\tau}$ in (13) and simplify to get (51).

Example of multiple (Pareto ranked) equilibria.

Individual firms technologies: $\beta = 1$; $\tau = \frac{11}{8}$, and $\sigma^2 = \frac{189}{11}$ and the empirical distribution of $\bar{a}_i$ is piecewise uniform (i.e., uniform in subintervals [2,3], [3,4], [4,6], [6,9] and [9,10]) such that G(2)=0; G(3)=0.2; G(4)=0.6; G(6)=0.72; G(9)=0.8; G(10)=1, where G(x)$\equiv \int_2^x g(\bar{a}_i)$. Under these conditions, and using the expressions derived in the proof of proposition 3 above, the following clusters can emerge:

(1) A "high-wage" cluster with $w = 6$ and limits $\{\overline{a}^c_L, \overline{a}^c_H\} = \{3, 9\}$. That is: $w = E[\bar{a}_i \mid \bar{a}_i \notin [3, 9]] = 6$; $\overline{a}^c_L(6) = 3$ and $\overline{a}^c_H(6) = 9$.

(2) A "low-wage" cluster with $w = 5$ and limits $\{\overline{a}^c_L, \overline{a}^c_H\} = \{4, 6\}$. That is: $w = E[\bar{a}_i \mid \bar{a}_i \notin [4, 6]] = 5$; $\overline{a}^c_L(5) = 4$ and $\overline{a}^c_H(5) = 6$.

Notice that the low-wage cluster Pareto-dominates the high-cluster one, because, even the lowest expected productivity firm (i.e., $\overline{a}_L = 2$) benefits from having a $w = 5$ rather than $w = 6$. In other words, $\frac{d\left(V_i^C - V_i^I\right)}{dw} = \frac{-(\bar{a}_i - w) - \frac{\beta}{\tau}w}{\beta} = \frac{-\bar{a}_i + \frac{3}{11}w}{\beta} < 0 \Longleftrightarrow \frac{3}{11}w < \overline{a}_i$, which holds for $\overline{a}_L = 2$, $w = 6$.

REFERENCES

Ciccone, A. and Hall, Robert E. "Productivity and the Density of Economic Activity." *American Economic Review,*1996, Vol. 86, pp. 54-70.

Combes, Pierre-Phillippe and Duranton, Gilles. "Labour Pooling, Labour Poaching, and Spatial Clustering." Center for Economic Policy Research (London) Discussion Paper No. 2975, 2001.

Dumais, Guy, Ellison, Glenn and Glaeser, Edward, L. "Dynamic Concentration as a Dynamic Process", 1997, NBER Working Paper no. 6270.

Duranton, Gilles and Puga, Diego. "Microfoundations of Urban Agglomeration Economies." London School of Economics and University of Toronto, 2003.

Einhorn, Hillel J. "Overconfidence in Judgement." *New Directions for Methodology of Social and Behavioral Science* 4, 1-16.

Ellison, Glenn and Glaeser, Edward L. "Geographic Concentration in U.S. Manufacturing Industries: A Dartboard Approach." *Journal of Political Economy,* 1997, 109(5), pp. 889-927.

Fujita, Masahisa, Krugman, Paul and Venables, Anthony J. *The Spatial Economy.* Cambridge, MA: MIT Press, 1999.

Fujita, Masahisa, Thisse, Jaques-Francois. *Economics of Agglomeration.* Cambridge, UK: Cambridge University Press, 2002.

Ellison, Glenn and Glaeser, Edward L. "The Geographic Concentration of Industry: Does Natural Advantage Explain Agglomeration?" *American Economic Review, 1999,* Vol. 89, No. 2, pp. 311-316.

Glaeser, Edward L. and Kohlhase, Janet E. "Cities, Regions and the Decline of Transportation Costs." HIER Working Paper No. 2014, July 2003.

Grossman, Gene M. and Helpman, Elhanan. "Integration in a Global Economy." National Bureau of Economic Research (Cambridge) Working Paper No. 8728, 2002.

Grossman, Sanford J. and Hart, Oliver D. "The Costs and Benefits of Ownership: A Theory of Vertical and Lateral Integration." *The Journal of Political Economy*, 1986, 96(4), pp. 691-719.

Hart, O. and Moore, John. "A Theory of Debt Based on the Inalienability of Human Capital." *Quarterly Journal of Economics*, 1994, 109, pp. 841-79.

Helsley, Robert W. and Strange, William C. "Matching and Agglomeration Economies in a System of Cities." *Regional Science and Urban Economics*, 1990, 20(2), pp. 189-212.

Helsley, Robert W. and Strange, William C. "Agglomeration Economies and Urban Capital Markets." *Journal of Urban Economics*, 1991, 29(1), pp. 96-112.

Henderson, J. Vernon. "Efficiency of Resource Usage and City Size." *Journal of Urban Economics*, 1986, 19, pp. 47-70.

Holmes, Thomas. J. and Stevens, John J. "Geographic Concentration and Establishment Scale" *Review of Economics and Statistics*, 2002, 84(4), pp. 682-691.

Krugman, Paul. *Geography and Trade.* Cambridge, MA: MIT Press, 1991.

Manes Stephen and Andrews,Paul. *Gates: How Microsoft's Mogul Reinvented an Industry-And Made Himself the Richest Man in America.* New York: Touchstone Books, 1994.

McLaren, John. "Globalization and Vertical Structure." *American Economic Review*, 2000, 90(5), pp. 1239-1254.

Marshall, Alfred. *Principles of Economics.* London: Macmillan, 1890.

Mortesen, Dale T. and Pissarides, Christopher A. "New Developments of Models of Search in the Labor Market." In Orley Ashenfelter and David Cars (eds.) *Handbook of Labor Economics,* Vol. *3.,* pp. 2567-2627. Amsterdam: Elsevier, 1999.

Matouschek, Niko and Robert-Nicoud, Frederic. "The Role of Human Capital Investments in the Location Decision of Firms." Northwestern University and University of Geneva, 2003.

Porter, Michael E. *The Competitive Advantage of Nations.* New York : Free Press, 1990.

Pissarides, Christopher. *Equilibrium Unemployment Theory.* Cambridge MA: MIT Press, 2000.

Rauch, James E. "Does History Matter Only When It Matters Little? The Case of City-Industry Location." *The Quarterly Journal of Economics*, 1993, 108 (3), pp. 843-867.

Rosenthal, Stuart S. and William C. Strange. "The Urban Rat Race," Syracuse University Working Paper, 2002.

Rotemberg, Julio J. and Garth Saloner. "Competition and Human Capital Accumulation: A Theory of Interregional Specialization and Trade. " *Regional Science and Urban Economics*, 2000, 30(4), pp. 373-404.

Saxenian, Annalee. *Regional Advantage.* Cambridge, MA: Harvard University Press, 1994.