European University Institute

Department of Economics

EUI Working Papers

ECO 2007/48

What Determines Entrepreneurial Clusters?

Luigi Guiso and Fabiano Schivardi

EUROPEAN UNIVERSITY INSTITUTE
DEPARTMENT OF ECONOMICS

What Determines Entrepreneurial Clusters?

LUIGI GUISO

and

FABIANO SCHIVARDI

EUI Working Paper **ECO** 2007/48

ISSN 1725-6704

Printed in Italy
European University Institute
Badia Fiesolana
I – 50014 San Domenico di Fiesole (FI)
Italy

http://www.eui.eu/
http://cadmus.eui.eu/

What determines entrepreneurial clusters?*

Luigi Guiso
University of Rome Tor Vergata & CEPR

Fabiano Schivardi
Bank of Italy & CEPR

July 11, 2006

Abstract

We contrast two potential explanations of the substantial differences in entrepreneurial activity observed across geographical areas: entry costs and external effects. We extend the Lucas model of entrepreneurship to allow for heterogeneous entry costs and for externalities that shift the distribution of entrepreneurial talents. We show that these assumptions have opposite predictions on the relation between entrepreneurial activity and firm level TFP: with different entry costs, in areas with more entrepreneurs firms' average productivity should be lower and vice versa. We test these implications on a sample of Italian firms and unambiguously reject the entry costs explanation in favor of the externalities one. We also investigate the sources of external effects, finding robust evidence that learning externalities are an important determinant of cross-sectional differences in entrepreneurial activity.

JEL Classification numbers: D24; D62; J23.
Keywords: Entrepreneurship, clustering, agglomeration economies

*This paper has greatly benefited from repeated discussions with Luigi Zingales. We thank Marianne Bertrand, Andrea Brandolini, Antonio Ciccone, Boyan Jovanovic, Riccardo Lagos, Claudio Michelacci, Gianluca Violante, Ekaterina Zhuravskaya, participants at the Workshop "Entry, Entrepreneurship and Financial Development" at the World Bank, the conference "Understanding Productivity Differences Across Sectors, Firms and Countries" in Alghero, and seminar participants at the Bank of Italy, CEMFI, New York University and Yale University for useful suggestions. Marco Chiurato and Elena Genito provided outstanding research assistance. Luigi Guiso thanks the EEC and MURST for financial support. The opinions expressed in the paper are our own and do not necessarily reflect those of the Bank of Italy. We are solely responsible for any mistakes. Email addresses: guiso@tin.it, fabiano.schivardi@bancaditalia.it

1 Introduction

There is a vast literature linking a country's endowment of "entrepreneurship" with economic prosperity. Environments where entrepreneurs can emerge easily are propitious to the creation of firms, their growth and their success. These ideas date back to Marshall (1890) and Schumpeter (1911): for example, the latter sees the entrepreneur as the carrier of innovation and hence the true engine of growth. But if entrepreneurship is so central to economic development, what drives it? Why are there so many entrepreneurs in some areas, such as Silicon Valley, and so few in others? Do these differences just reflect differences in opportunities driven by - say - the presence of Stanford University? Why do we find these clusters in some countries rather than others, and in particular areas within countries, such as in the Italian industrial districts or the Ruhr? These important questions have often been in the forefront of policy debate and government intervention.

To answer these questions, one must address the choice to become an entrepreneur. Perhaps the best known model of entrepreneurship is that of Lucas (1978), explaining who in a given population will become an entrepreneur using differences in exogenously given individual talents. In Lucas, "talent" is identified with the ability to extract output from a given combination of inputs. Thus, more talented individuals are those who can obtain a higher total factor productivity (TFP) if they start a business. Since individuals with more talent make more profits, they will choose to be entrepreneurs. But what can explain clusters if we do not believe in genetic differences, i.e. if the distribution at birth of individual ability is the same across populations?

In this paper we extend the Lucas model to investigate two potential explanations of differences in entrepreneurial density across locations: entry costs and external effects. One possibility is that there are heterogeneous costs of entry, and the locations with lower costs of setting up a firm end up with more entrepreneurs and more firms because even relatively less talented individuals will find it profitable to start a business. This is the approach implicitly followed by the large literature that focusses on factors - particularly financial - that keep the would-be entrepreneur from actually creating a new firm. Banerjee and Newman (1994), for instance, show that credit constraints can lead to poverty traps, since potential entrepreneurs cannot invest in profitable occupations involving set-up costs.

It is well documented empirically that limited access to financial markets can hinder the emergence of entrepreneurs.[1] More recently it has also been shown that regulation-induced high costs of entry hamper firm and employment creation (Klapper, Laeven and Rajan 2004, Bertrand and Kramarz 2002, Fonseca, Lopez-Garcia and Pissarides 2001).

The second possibility is that the distribution of individual productivity is shifted by local externalities, for instance because of learning opportunities, knowledge spillovers or intermediate input variety. A vast body of literature has pointed out that local externalities are an important determinant of firms' performance.[2] Externalities can be introduced in the Lucas model as shifter of the distribution of talents that make on average individuals more productive and therefore increase the share of agents that choose to start a business.

We show that the two assumptions have very contrasting implications regarding the relation between the propensity of individuals to become entrepreneurs and their average productivity: under entry costs differences, in areas with lower entry costs *i*) there should be a larger share of entrepreneurs in the population, and *ii*) their firms' average TFP should be lower. Thus, in equilibrium there should be a negative correlation between the number of firms in a given location (for a given population) and their TFP. On the contrary, with externality differences, in places with stronger externalities average entrepreneurial ability is higher and there are more entrepreneurs in relation to the population. In contrast with heterogeneous entry costs, the model with heterogeneous externalities implies therefore that in equilibrium there should be a positive correlation between the share of entrepreneurs in a given location and their firms' TFP.

We test these implications on a sample of Italian firms belonging to different clusters and find that the data unambiguously reject the entry costs story and support the externality one: areas with a higher share of firms in relation to the population are characterized by an higher average productivity.

[1] Evans and Jovanovic (1989) show that wealthier individuals who are currently employees are more likely than the less wealthy to become self-employed, a finding consistent with liquidity constraints. Blanchflower and Oswald (1998) show that liquidity constraints affect the decision to become an entrepreneur even after controlling for individual ability. More recently, Guiso, Sapienza and Zingales (2004) using individual level data for Italy find that in areas with a higher degree of financial development it is more likely that individuals become entrepreneurs, the rate of firm creation is higher and there are more firms per inhabitant.

[2] See for example the contributions in the *Handbook of Regional and Urban Economics, vol. 4* (Henderson and Thisse 2004).

Based on this result, in the rest of the paper we investigate further the sources of such externalities. Following Marshall (1890) and the literature on agglomeration economies, and in particular Duranton and Puga (2004), we identify three different channels through which agglomeration may affect firms' productivity: first, the opportunities to learn from other firms, either because of knowledge spillovers or through learning entrepreneurial abilities from the observation of other firms; second, the size of the local work force, which can increase the division of labor and the quality of job-worker matches; third, a greater variety of intermediate inputs. To discriminate between these three sources of externalities we run a horse race by constructing proxies for each one. As a proxy for learning opportunities and knowledge spillovers we use the number of firms in a location, the idea being that more firms offer more learning points; we proxy job-matching opportunities with the size of the local workforce; intermediate inputs variety is measured by the ratio of intermediate inputs to value added at the local level. We find strong evidence for the first channel, supporting evidence for intermediate inputs variety and very week or no evidence for externalities generated by labor pooling. In fact, the correlation between TFP and number of firms proves to be extremely robust to a series of controls.

We further corroborate our interpretation of the number of firms in terms of learning and knowledge spillovers. First, the findings are consistent with causality running from the number of firms to productivity rather than the reflection of unobserved local factors driving both productivity and the number of firms. In fact, when the number of firms is instrumented with the local population in 1861, following an idea of Ciccone and Hall (1996), the estimates are very similar to those obtained with OLS. Second, we directly test some implications that are unique to the learning model. As predicted by the knowledge diffusion models of Jovanovic and Rob (1989) and Eeckhout and Jovanovic (2002), learning should be more relevant for small firms, and knowledge dispersion across firms should positively affect the amount of spillovers, thus increasing productivity. Both predictions are supported by the data.

This paper relates to three strands in the literature. First, it is connected with the entrepreneurship literature: we sort out two alternative explanations of clusters. Second, it contributes to the agglomeration literature: we investigate the sources of local externalities.

Finally, it is related to the productivity literature. We provide evidence that differences in firm-level TFP may be due to the differing ability of entrepreneurs, that in turn could depend from the degree of knowledge spillovers.

Our results bear important policy implications. First, consistently with McKenzie and Woodruff (2003) for Mexico, we de-emphasize entry costs in explaining regional differences in entrepreneurial activity, adding an important element to the debate on the barriers to entrepreneurship. Second, our findings indicate that the density of firms might be a fundamental driving force of local externalities. This result is not confined to Italy: Henderson (2003) finds a positive effect of the number of plants at the local level on productivity in the US.

The rest of the paper is organized as follows. Section 2 sets out a simplified version of the Lucas model with exogenous factor prices, extended to incorporate the cost of setting up a firm. Exogenous and geographically heterogeneous costs of setting up a firm are a simple way to generate clusters. We then introduce the possibility of that the original distribution of talents can be shifted by a local externality and compare the predictions of this model with those of the set-up-costs model. We then empirically contrast a number of testable predictions from these two models, using Italian firm-level data matched with firm cluster information and described in Section 3. Section 4 presents our basic results, showing that, contrary to the pure set-up-costs model but in agreement with the externality model, entrepreneurial ability is greater where there are more entrepreneurs over the local population, and the probability mass on the left-hand side of the empirical ability distribution is lower where there are more entrepreneurs. In Section 5 we explore the externality explanation of the correlation between the mass of entrepreneurs and their quality, finding evidence for a leaning externality. Section 6 extends the analysis to test some unique implications of the learning model and runs a number of robustness checks. Section 7 summarizes and concludes.

2 The model

We use a modified version of Lucas (1978) model of entrepreneurial ability. The economy consists of N regions, within each of which firms produce output using labor (n) and capital

(k); the prices of the two inputs, denoted respectively w and u, are exogenously given and equal across regions. This assumption, which greatly simplifies the analysis, can be defended on the grounds that in our empirical analysis the territorial units are small enough to assume that they have no impact on aggregate factor prices and that factors are fully mobile.[3] Our conjecture is that with endogenous wages and user cost the results would still hold.

Given our partial equilibrium approach, we can confine the analysis to one representative region. We modify the basic model of Lucas by introducing a start-up cost c which has to be paid when becoming an entrepreneur. Individuals, who can be either entrepreneurs or employees, differ in entrepreneurial talent x; the more talented can get more output from a given combination of labor and capital if they choose to run a firm. Entrepreneurial talents are drawn from a random variable $\widetilde{x}$ distributed according to a distribution function $\gamma(x)$ over the support $(\underline{x}, \overline{x})$, $0 \leq \underline{x} < \overline{x} \leq \infty$, with corresponding cumulative distribution function $\Gamma(x)$. Output is produced according to the production function $xg[f(n, k)]$, where f is a constant returns to scale function and g is a concave transformation. Following Lucas, we interpret this as the span of control. Define $\phi(r) = f(n, k)/n$, where $r = k/n$. The first order conditions for an entrepreneur who maximizes profits can be written as:

$$\frac{\phi(r) - r\phi'(r)}{\phi'(r)} = \frac{w}{u} \tag{1}$$

$$xg'[n\phi(r)]\phi'(r) = u \tag{2}$$

from which it is immediately evident that the capital/labor ratio does not depend on x. The above FOCs give two equations in two unknowns, which can be solved implicitly to obtain two-factor demand functions in terms of entrepreneurial ability: $n(x)$, $k(x)$. It is immediately verifiable that $n'(x) > 0, k'(x) > 0$.

[3] Michelacci and Silva (2005) document that, in Italy, the share of entrepreneurs starting a business in the locality where they were born is substantially higher than the corresponding share of employees, arguing that becoming an entrepreneur requires some location-specific "relational" capital. This finding fully supports our modelling assumptions.

2.1 Start-up costs

We now depart from Lucas. When becoming an entrepreneur, the agent pays a cost c. The profits of an entrepreneur of ability x before the entry cost are

$$\pi(x) = xg[n(x)\phi(r)] - n(x)[w + ur]. \tag{3}$$

Using the optimal input choices condition (1) and (2), we get $\pi'(x) = g[n(x)\phi(r)] > 0$: the profits of an entrepreneur are increasing with ability. An individual becomes an entrepreneur if $\pi(x) - c \geq w$. Given that $\pi(x)$ is increasing, and that $\pi(0) = 0$, there exists one and only one value z at which the "marginal" individual is indifferent between being an entrepreneur and an employee:

$$\pi(z) - c = w \tag{4}$$

which implicitly defines the ability threshold value $z(c)$ above which it is optimal to become an entrepreneur. In this economy, the mass of workers will be $\Gamma(z)$ and that of entrepreneurs $(1 - \Gamma(z))$. By differentiating (4), we find that

$$\frac{dz}{dc} = \frac{1}{\pi'(z)} > 0 \tag{5}$$

The higher the start-up cost, the greater the ability of the marginal entrepreneur.

How can this model generate different levels of entrepreneurial activity across regions? A first possibility is that regions differ in terms of entry cost c, say because of differences in bureaucratic costs due to disparate efficiency of the public administration. Areas with lower costs will have a larger share of entrepreneurs:

$$\frac{d(1 - \Gamma(z))}{dc} = -\gamma(z)\frac{dz}{dc} < 0 \tag{6}$$

Define the average entrepreneurial quality as the expected value of x conditional on being an entrepreneur:

$$E(x|x \geq z) = \frac{\int_z^{\overline{x}} x\gamma(x)dx}{1 - \Gamma(z)}. \tag{7}$$

When c rises, the quality of the marginal entrepreneur increases, hence so does average entrepreneurial quality:

$$\frac{dE(x|z)}{dc} = \frac{[E(x|z) - z]\gamma(z)}{[1 - \Gamma(z)]}\frac{dz}{dc} > 0 \tag{8}$$

where, to facilitate notation, $E(x|z)$ stands for $E(x|x \geq z)$. The inequality follows from the fact that $\frac{dz}{dc} > 0$ and $E(x|x \geq z) > z$, where the last inequality formalizes the notion that the marginal entrepreneur z is of lower quality than the average. Equations (6) and (8) imply that if differences in the share of entrepreneurs across locations are explained by entry costs, we should expect a negative correlation between the share of entrepreneurs and their average quality.

We can obtain additional implications of heterogeneity in entry costs for the distribution of entrepreneurial ability. Define $\Omega(y|z)$ as the cumulative density function of the random variable obtained by truncating x at z:

$$\Omega(y|z) = \begin{cases} \frac{\Gamma(y)-\Gamma(z)}{1-\Gamma(z)} & \text{if } y \geq z \\ 0 & \text{otherwise} \end{cases} \tag{9}$$

$\Omega(y|z)$ is the share of entrepreneurs below any given level of ability y. As c increases, this share falls:

$$\frac{d\Omega(y|z)}{dc} = -\frac{\gamma(z)(1-\Gamma(y))}{(1-\Gamma(z))^2}\frac{dz}{dc} < 0 \tag{10}$$

This implies that heterogeneous entry costs induce a positive correlation of the share of entrepreneurs below any given level of ability with the overall mass of entrepreneurs, and a negative correlation above that level.

Summing up, heterogeneity in entry costs generates two sharp predictions: a larger mass of entrepreneurs should be associated with *i*) lower overall quality and *ii*) with a larger (smaller) share of entrepreneurs below (above) any quality level.

2.2 Local externalities

A second potential reason for different levels of entrepreneurial activity across locations is that the distribution of entrepreneurial skills is different, due in particular to local externalities. For example, in some locations the diffusion of knowledge and ideas is facilitated by environmental factors. As shown by Jovanovic and Rob (1989), the easier the circulation of knowledge the higher entrepreneurial quality. To model this possibility and derive its implications, we assume that the distribution of talent is parameterized by a shift factor λ, specific to each location: $x \sim \gamma(.,\lambda)$, with $\Gamma(x,\lambda)$ representing the cumulative distribution function. The parameter λ measures the intensity of local externalities. We assume that $\partial\Gamma/\partial\lambda < 0$: λ shifts the probability distribution to the right in the first order stochastic dominance sense. In this setting, clusters arise in areas with high λ:

$$\frac{d(1-\Gamma(z,\lambda))}{d\lambda} = -\frac{\partial\Gamma(z,\lambda)}{\partial\lambda} > 0. \tag{11}$$

Equation (11) implies that the higher λ, the larger the share of individuals with a given talent above the threshold z, and so the larger the mass of entrepreneurs.

As before, we define average entrepreneurial quality as the expected value of x conditional on being an entrepreneur. This value will now depend on λ:

$$E(x|z,\lambda) = \frac{\int_z^{\overline{x}} x\gamma(x,\lambda)dx}{1-\Gamma(z,\lambda)}. \tag{12}$$

The effect of a change in λ on average entrepreneurial quality is:

$$\frac{dE(x|z,\lambda)}{d\lambda} = \frac{[\int_z^{\overline{x}} x\frac{\partial\gamma}{\partial\lambda}dx - E(x|z,\lambda)\frac{\partial(1-\Gamma(z,\lambda))}{\partial\lambda}]}{(1-\Gamma(z,\lambda))} \tag{13}$$

Given that the first term is positive[4] and the second negative, this expression cannot be signed a priori. In fact, an increase in λ has two contrasting effects on average ability: on one side, for given z, it shifts ability to the right, i.e. increases average ability; on the other, some agents that would have been employees for a lower λ now become entrepreneurs. Given that they enter at the lower end of the talent distribution, more 'entry' implies that quality is diluted, thus reducing average quality - the second term in square brackets in (13). The sign of $\frac{dE(x|z,\lambda)}{d\lambda}$ depends on the shape of the distribution of talents and on how λ parametrizes it. However, $\frac{dE(x|z,\lambda)}{d\lambda} > 0$ holds for a general family of distributions, the log-concave

[4] Using the property of first order stochastic dominance, it can be shown that $\int_z^{\overline{x}} x\frac{\partial\gamma}{\partial\lambda}dx > 0$. In fact, for $d\lambda > 0$, stochastic dominance implies that $\int_z^{\overline{x}} x\gamma(x,\lambda+d\lambda)dx > \int_z^{\overline{x}} x\gamma(x,\lambda)dx$. Grouping terms and taking the limit for $d\lambda \to 0$ delivers the result.

distributions (Barlow and Proschan 1975).[5] This family of distributions includes, among others, the uniform, the normal and the exponential. For such distributions, a positive correlation between the share of entrepreneurs and their average quality will emerge.[6]

The same reasoning applies to the distribution of entrepreneurial talents conditional on $x \geq z$, $\Omega(x|z,\lambda)$: as before, while not determined a priori, the mass of entrepreneurs with talent below (above) an arbitrary threshold y, $\Omega(y|z,\lambda)$ can decrease (increase) with the density of firms. Therefore, with externalities, under mild conditions on the distribution function Γ there is a positive relation between the overall share of entrepreneurs and their quality.

To sum up, a model with different entry costs predicts a negative correlation between the number of entrepreneurs and their quality, while a model with externalities is compatible with a positive correlation. We confront the implications of the models with the data in the subsequent sections.

3 Data description

We test our propositions drawing on a dataset of Italian firms, the Company Accounts Data Service (in Italian, "Centrale dei Bilanci", CB), which provides standardized data on the balance sheets and income statements of about 30,000 Italian non-financial firms plus information on employment and firm characteristics. Data, available since 1982, are collected by a consortium of banks interested in pooling information about their clients. A firm is included if it borrows from a bank in the consortium. The focus on level of borrowing skews the sample towards larger firms. Furthermore, because most of the large banks are in the northern part of the country, the sample has more firms headquartered in the North than in the South. Finally, since banks are interested in creditworthy firms, those in default are not included, so the sample is biased towards better-quality borrowers. Despite these biases, previous comparisons with population moments indicate that the sample is not too far from being representative; moreover, it covers more than half of private sector sales

[5] A function h is said to be log-concave if its logarithm $\ln h$ is concave, that is if $h''(x)h(x) - h'(x)^2 \leq 0$.

[6] The log-normal, traditionally used to model firm size (Steindl 1990) and income distribution (Harrison 1981), does not satisfy this property. Some simulations indicate that even for this distribution the above condition will generically be satisfied, implying that average ability increases with the number of firms.

(Guiso and Schivardi 2005). Table 1, Panel A, gives summary statistics on employment, value added and the stock of capital at constant prices for the 1991 CB sample comprising 16,885 observations; we use 1991 as the reference year for our regressions but check for robustness when all the available years are used. Data are reported by industrial sector using a 10-industry classification; to avoid the usual problems of estimating productivity in services we have restricted the analysis to manufacturing.[7] The capital stock is constructed using the permanent inventory method with sectoral deflator and depreciation rates (see Cingano and Schivardi (2004) for details).

We complement the CB data with another dataset on Italian industrial clusters, the Local Labor Systems dataset. The territory of Italy has been divided by the National Statistical Institute (ISTAT) into 784 local labor systems (LLS) on the basis of working-day commuting areas.[8] The idea behind the algorithm is to define self-contained labor markets in terms of worker mobility. As such, this is the correct geographical boundary for an environment within which learning takes place and learning relations are formed. Since the Data Services gives the firm's LLS code, we can match firms with the corresponding LLS. The resident population of each LLS comes from the census while the number of manufacturing firms in the LLS is obtained from the files of the Italian Social Security Administration (INPS) on the population of firms with at least one employee for the years 1986-98. With respect to CB, the information on firms is much less detailed: for example, output is not reported so that TFP cannot be computed. For our purposes, the database contains the number of employees, the sector and the location of each firm, from which very precise measures of industrial density at the local and sectoral level can be constructed. Panel B shows summary statistics on the average number of firms per LLS for each of our 10 industries. It is clear that there is considerable geographical variation in the clustering of industry: for all industries, there are 92 firms per LLS with a standard deviation almost

[7] As will become clear in the next subsection, the sectoral classification balances the need for homogeneity of the production technology and that of a sufficient number of sectoral observations to properly estimate TFP.

[8] Even if defined using the same criteria (commuting ties), the concept of LLS differs from US Core Based Statistical Areas since there is no minimum population requirement. Hence, like the French "zones d'emploi", the Italian LLS entirely and continuously cover the national territory. The average land-area is 384 square kilometers, with a population density of 188 inhabitants per sq. km. Population ranges from 3,000 in the smallest LLS to 3.3 million in the largest.

three times the mean (269) and range of 1-8,392. Figure 1 gives a visual picture of the geographical clustering of entrepreneurship and its dispersion across local labor systems. Although the largest clusters are mostly in the North, there are several in the South. However, to make sure that our results are not driven by North-South differences, in our empirical analysis we always include area dummies.

3.1 Estimating entrepreneurial ability

In order to test the two alternative models described in Section 2 first we need a measure of entrepreneurial ability at the firm level. In Lucas, entrepreneurial ability is modeled as a shift in the production function: better entrepreneurs will get more output from any combination of inputs. Put in this way, entrepreneurial ability is simply equivalent to a firm's TFP. If this were the only feature affecting a firm's TFP, Lucas's model might serve as the basis for a theory of TFP, or at least of its dispersion across firms. The main limit of this theory is that the dispersion is simply assumed and inherited from the differences in entrepreneurial ability, which is taken as given. Our model of externalities is a way of providing a basis for an endogenous explanation of firm-specific TFP and the determinants of differences in average TFP across locations. To compute the contribution of entrepreneurial ability to TFP we assume that a firm's TFP has two components: one is common to all firms in the same industry and depends on their specific technology. The second is firm-specific, and in the spirit of our model we assume it reflects the ability of the firm's entrepreneur.

To obtain an estimate of the TFP of firm i we assume that output is produced with a Cobb-Douglas production function of the form $Y_{si} = x_{si} A_s * K_i^{\alpha} L_i^{\beta}$, where s indexes the industry, Y is output, and K and L denote the stock of capital and labor services. TFP is given by $TFP_{si} = x_{si} A_s$, and is the product of the sectoral component, A, and the firm-specific component, x. The latter is our measure of entrepreneurial ability. To obtain an estimate of TFP we need to compute values for α and β. To obtain estimates of the production function parameters that are robust to the endogeneity of some of the inputs (capital accumulation and labor demand may respond to unobserved productivity shocks) and the selection induced by exit (with some irreversibility, leaving the industry is more likely for firms with a lower capital stock when a bad productivity shock occurs) we use the multi-step estimation algorithm proposed by Olley and Pakes (1996), which accounts for both problems, allowing for consistent and unconstrained estimation of α_s and β_s.[9] To assess the reliability of the estimates, we also calculate the coefficients using Solow's assumptions.

Table 2 reports the estimated values of α_s and β_s with the two procedures. Production function estimates of $(\alpha_s + \beta_s)$ lie in the range 0.93-1.05. The model assumes decreasing returns to scale to avoid a degenerate equilibrium in which there exists only one firm supplying the whole market. Given that the capital coefficient is estimated using a semi-parametric procedure, we obtained its standard errors through a bootstrapping exercise based on 150 replications. As in Olley and Pakes (1996), standard errors are relatively large[10] and, given that the estimates of $(\alpha_s + \beta_s)$ are always somewhere around 1, the empirical model has no power to discriminate between different degrees of returns to scale. Formally, the null hypothesis $(\alpha_s + \beta_s) <1$ is never rejected in a one-sided test even at the 10 per cent confidence level.[11] In terms of single coefficients, compared with Solow's method the Olley and Pakes (1996) procedure tends to result in a higher coefficient for labor and a lower one for capital, arguably because of deviations of the factor markets from the competitive paradigm; apart from these differences, the two methods give broadly consistent results. In what follows, we use the TFP estimates obtained with the Olley and Pakes procedure to recover our measure of entrepreneurial ability.

Table 3 describes the distribution of our estimate of x, obtained by removing the industry-level component with a first-stage regression of estimated TFP on a set of industry dummies. To account for possible outliers we drop observations in the first and last percentile of the ability distribution by year-sector. The sample mean of entrepreneurial ability is 2.35 but there is considerable dispersion, as the high value of the standard de-

[9] To summarize, the procedure controls for endogeneity by approximating the unobserved productivity shocks with a non-parametric function of observable variables and for selection by introducing a Heckman-type correction term.

[10] Pakes and Olley (1995) discuss the asymptotic properties of the estimator, suggesting that the bootstrapping procedure might overestimate the true standard deviation of the capital coefficient, partially explaining why its values are higher than those for labor.

[11] Indeed, returns to scale might be initially increasing, due for example to fixed production costs, so that the "span of control" only kicks in for larger levels of operation. In fact, some small, growing firms might still be on the increasing part of the production function but, due to convex costs of adjusting the scale of operation, might not immediately exploit the full advantages of scale.

viation (0.5) implies. When the sample is split according to the density of firms in each geographical unit, entrepreneurial ability is higher where firms' density is above median than where it is below median (2.39 compared to 2.20), which is inconsistent with the start-up cost hypothesis but not with externalities. The table also shows the share of firms with ability below the 25^{th} and above the 75^{th} percentile both for the total sample and the two sub-samples of high-density and low-density areas. Contrary to the start-up cost model, there is a larger frequency mass to the left of the lower threshold in places with fewer firms (36 per cent in the low-density group compared with 22 per cent among the high-density locations) while in accordance with the externalities hypothesis the probability mass to the right of the upper threshold (the 75^{th} percentile) is greater where there are more firms for given population. Thus, the descriptive evidence clearly rejects the start-up cost theory in favor of externalities. The same conclusions can be drawn from Figure 2, which shows the distribution of entrepreneurial ability for firms in high-density and low-density local labor systems, computed using Gaussian kernel non-parametric smoothers evaluated at 25 points over the range of x. Notably, the distribution of entrepreneurial ability is shifted to the right in areas with a high density of firms, implying that entrepreneurs in high-density areas have greater ability. In the following sections we refine this descriptive evidence using formal regressions testing for statistical significance and controlling for any additional effects. Table 2 also reports summary statistics for some of these controls, such as the number of firms in the LLS and that in the LLS and industry, the number of workers in the LLS and the cumulated stock of capital in the LLS. Not surprisingly, in high-density areas there are also more workers and capital stock is larger.

4 Start-up costs or externalities? Testing the two models

According to equations (6) and (8), the start-up cost model implies that as we vary entry costs, entrepreneurial ability and the mass of entrepreneurs should move in opposite directions. Under reasonable assumptions, the externalities model implies a positive correlation between the two variables. Table 4 reports our basic test; the left-hand side is the log of our estimate of entrepreneurial ability (x in the model); on the right-hand side we include the share of entrepreneurs, i.e. the ratio between the number of firms, obtained from the yearly INPS archives, and the resident population in the local labor system, obtained from the population census ($1-\Gamma(z)$ in the model), a variable we call Entrepreneurial Incidence (EI). We also include as controls three geographical dummies for the Centre, Northeast and Northwest of the country (the South is the omitted region). Given that the independent variable only varies across LLS-year, we use standard errors adjusted for clustering. To check the robustness of our results, we use three different samples: a single cross section in 1991, which is the Census year when the population is counted; the de-trended firm average over the entire period[12] and the full panel with year dummies. The first column shows the estimates using the 1991 cross-section; the correlation between EI and TFP is positive and statistically significant at 10%; to give a sense of its magnitude, an increase in EI of one standard deviation would bring about an increase in TFP of a little more than 2%. Using firm averages and pooled data (columns 2 and 3), the estimates are slightly higher and much more precise. This clearly contradicts the start–up cost model of cluster formation already questioned by the previous descriptive evidence. This result is very robust across specifications.

To control for unobserved factors at the local level, we run the regressions including province dummies, a very fine geographical control.[13] The coefficient is slightly lower, an indication of possible spatially correlated effects, but still positive and statistically significant in two cases out of three.

The second panel of Table 4 sharpens the evidence on the validity of the start-up cost theory by looking at the relations between the number of firms in a cluster and the share with ability below a lower bound or above an upper one. According to this model, there should be a positive (negative) correlation between the number of firms and the frequency of firms with ability below (above) a certain bound. The intuition is that as the start-up cost declines and the number of firms increases, the new entrants are of lower quality, so there is a larger (smaller) mass of entrepreneurs with ability below (above) any given threshold. To test this implication we set the lower bound at the 25^{th} (and the upper at

[12] According to Bertrand, Duflo and Mullainathan (2004), the serial correlation in the independent variable can make inference problematic. As a simple solution, they propose running estimates on the collapsed data and ignoring the time series variation.

[13] Italy has 103 provinces, so that there are 8 LLS per province.

the 75^{th}) percentile of the empirical distribution of ability and construct an indicator that is equal to 1 if the firm's specific ability is below (above) the threshold. We then run a probit estimate on the entrepreneurial share and the geographical controls. The first three columns of Table 4 show the results for the share above the 25^{th} percentile for the three samples, using macro areas as geographical controls. They reveal a negative correlation, with highly significant coefficients. This pattern is broadly confirmed when provinces are used as geographical controls (last three columns). The last panel shows the share of firms above the 75^{th} percentile, finding a positive coefficient of the number of firms scaled by population, albeit significant only for the pooled sample. Taken together, these findings suggest that a larger share of firms is associated with a shift to the right in the distribution of entrepreneurial talent. Thus, the two main implications of the start-up cost model are strongly rejected by the data. On the other hand, they are consistent with externalities, which (under mild conditions) not only predict a positive correlation between ability and the share of entrepreneurs in the population, but also a negative correlation between the share of firms with ability below a lower bound and the entrepreneurial share in a cluster (and vice versa for the right tail).

The evidence in Table 4 is unequivocal: it strongly rejects theories of cluster formation based only on differences in entry and start-up costs, such as differences in the fixed costs or bureaucratic steps required to organize a firm. It lends support to models that emphasize differences in the distribution of entrepreneurial abilities due to local externalities.

To further strengthen our interpretation, we consider the correlation between EI and TFP at the local sectoral level. If differences in entrepreneurial incidence are due to entry costs, then they should apply independently of the sector of activity, so that the correlation should arise at the aggregate level.[14] Instead, we should expect externalities to have a strong sectoral component.[15] In fact, entrepreneurship entails some degree of specificity. Thus, if externalities are more important within an industry than across industries (due to some

[14] This is not to say that entry costs are equal across industries. Rather, it means that if two locations differ in the level of such costs, the differential effect on EI should be uniform across industries.

[15] There is a large literature on the relative importance of intra-industry and inter-industry spillovers, which has so far failed to reach a clear consensus, even if the intra-industry component tends to be important in most empirical studies (Rosenthal and Strange 2004). Using the same dataset as in this paper, Cingano and Schivardi (2004) find evidence in favor of intra-industry spillovers, and no evidence of inter-industry.

kind of specificity), then the correlation between the share of firms and entrepreneurial ability should be stronger in the same LLS and industry than overall. This is tested in Table 5, where we insert both the overall entrepreneurial share and that at the sectoral level, i.e. calculated using the number of firms in an LLS-industry. The first panel shows the results for the correlation between ability and the two indexes of entrepreneurial share. We always find that the effect is stronger and the significance higher for the industry-level index, independently of sample and type of geographical controls.

The second and third panels report the regressions for the probability that the firm has a specific component of TFP below the 25^{th} percentile of the distribution (Panel B) and above the 75^{th} percentile (Panel C). The pattern is very similar to that found in Panel A, with two exceptions for the first indicator, for which the overall share is sometimes significant. All in all, the evidence suggests that location-industry factors underlie the correlation, consistently with the externalities hypothesis and at odds with the idea that some locations have more firms because of lower start-up costs.

5 Which externalities?

Up to now, we have used the model to obtain equilibrium correlations between EI and ability, without any causal interpretation. We now take a further step and investigate the causes. In practice, by log linearizing (12), we can immediately verify that this amounts to identifying some measurable factors that shift the ability distribution (the variable λ in the model) and to running a regression of (log) ability on the (log) indicator of λ. The logical candidate to explain productivity differences according to density is local externalities. In this section, therefore, we contrast different sources of externalities to look for more direct evidence that can sort out their nature.

There is a large theoretical literature on agglomeration economies (see Duranton and Puga (2004) for a recent survey). This literature has maintained the original Marshallian idea (Marshall 1890) that the spatial concentration of production can be beneficial for three reasons. First, concentration fosters the circulation of ideas and the possibility of learning from other agents.[16] Second, a large concentration of workers in the same industry can have beneficial effects both in terms of the specialization that each worker can achieve and the quality of worker/job matches. Third, industrial clusters offer a wide variety of intermediate inputs, with potentially beneficial effects on productivity.[17] The empirical literature on the extent and scope of agglomeration economies suggests that localization economies are important. However, a consensus has not yet emerged on the relative merits of the different sources and investigation is continuing (see Rosenthal and Strange (2004) for an exhaustive assessment of the state of the empirical debate).

We distinguish among these different effects by proposing a proxy of each potential externality. To proxy for learning externalities we use the number of firms operating in a given industry and in a given area. If learning entrepreneurial abilities is not, as we think, a routine activity, then an obvious feature facilitating learning is the number of firms in a given location. If learning takes place mainly on the job and on the site, a larger number of firms offers more (and better) opportunities to acquire entrepreneurial abilities, since a potential entrepreneur can compare different working practices and business idea, possibly by working in different firms.[18] Moreover, the process of knowledge acquisition continues even after the business is started, because knowledge spillovers on alternative technologies or new markets keep accruing in regions with a large population of firms. The availability of intermediate inputs - the second reason why spatial concentration can raise firms' productivity - is easily measured by the ratio of intermediate inputs to value added at the local sectoral level. In fact, if greater concentration leads to higher productivity

[16] Guiso and Schivardi (2005) show that information flows are an important determinant for firms' employment choices in Italian industrial districts.

[17] Marshall (1890) wrote: "When an industry has thus chosen a locality for itself, it is likely to stay there long: so great are the advantages which people following the same skilled trade get from neighborhood to one another. The mysteries of trade become no mysteries; but are as in the air, and children learn many of them unconsciously.... Employers are apt to resort to any place where they are likely to find a good choice of workers with the special skill which they require. .. The advantages of variety of employment are combined with those of localized industries in some of our manufacturing towns, and this is a chief cause of their continued economic growth".

[18] According to Saxenian (1994), the mobility of workers across firms and their acquired capacity to start up new firms was one of the main reasons behind the success of Silicon Valley during the technology boom. This would also be consistent with the model and the empirical evidence of Lazear (2005), according to which the probability of becoming an entrepreneur is positively related to the number of tasks a worker is previously exposed to, because the entrepreneur needs to be able to understand and coordinate different activities: again, more firms could offer better opportunities of learning the complex set of skills required to manage a firm.

through more reliance on intermediate inputs,[19] we should find that TFP is positively related to this indicator. The third reason, labor market pooling effect, is measured by the number of workers operating in a given LLS-sector. Summary statistics for these variables are reported in Table 3.

In Table 6, Panel A, we regress firm-level TFP on the number of firms, the share of intermediate inputs over value added and the number of workers in the LLS-industry (all variables are in logs).[20] With four spatial controls, we find that the number of firms has a positive and significant coefficient in all specifications, with a value of around 0.05. The share of intermediate inputs is not significantly different from zero in the cross-sections, but is significant when using averaged data and in the pooled data, and thus there are indications that the availability of intermediate inputs might also foster local productivity, though the evidence is less clear-cut than for the number of firms. The number of workers is never significant, save in one case. The exception is the pooled data with 4 spatial controls, and its negative coefficient is at odds with the idea that local externalities are attributable to labor market pooling effects. To give a sense of the magnitude of these effects, using the pooled estimate of column 3 we calculate that increasing the number of firms by one standard deviation would bring about an increase in firms' productivity of about 9 per cent, which is quite large; doing the same with intermediate input intensity would increase TFP by a more modest 1.2 per cent. The last three columns of Table 6 repeat the exercise with 103 spatial controls (the provinces). The estimates for the number of firms and the intermediate inputs become somewhat smaller, but remain highly significant. The number of workers has no effect in any specification.

These patterns are confirmed by the analysis of the TFP percentiles. The second panel shows the regressions for the probability that the firm-specific component of TFP falls below the 25^{th} percentile of the distribution. The number of firms has a negative effect, as predicted by the learning model, and its coefficient is statistically significant in all specifications. The intermediate input indicator is also negative, although again less precisely estimated,

[19] Using US data, Holmes (1999) finds that sectoral concentration at the local level is positively related to intermediate input intensity, although the effect is rather modest.

[20] While the number of firms and that of workers can be computed from the INPS dataset, and therefore cover the respective populations, we have information on intermediate inputs and value added only for the CB sample, which is therefore used to compute the measure of intermediate input intensity.

and the labor market pooling indicator is marginally significant in only one case. The last panel shows the estimates for the probability that firm-specific TFP exceeds the 75^{th} percentile of the distribution, with results very much in line with the previous ones. Here, the intermediate inputs indicator is always significant while the number of workers is either insignificant or negative.

All in all, we conclude that the evidence supports both learning externalities and intermediate input variety, with the former playing a more prominent role. Controlling for these sources, no evidence of labor market pooling emerges.

6 Robustness and further implications

Having established that the number of firms is strongly correlated with firm-level TFP, we further investigate if we can correctly interpret this correlation as evidence in favor of learning externalities, as suggested above.

6.1 Unobserved heterogeneity

The OLS correlations face the endogeneity problem that plagues the empirical analysis of density and productivity. There could in fact be unobserved local factors, not accounted for by our geographical controls, as even the finer ones (the province dummies) refer to wider areas than the Local Labor System. For example, politicians might care about places with a high production density and provide business-oriented public goods, such as infrastructure, which raise productivity. While the province dummies absorb some of these effects, the transfers could take place at an even finer geographical level, leaving the regression residual correlated with a firm TFP. We address this issue in two ways. First, since in the estimates in Table 6 our regressors vary with the LLS and the industry, we can exploit the cross-industry variation while inserting geographical controls at the LLS level. In Table 7 we report the same regressions as in the previous one, adding dummies for the 784 LLS dummies. We find results that are similar to those of Table 6 with province dummies, losing significance only in the case of the cross-section sample. Similar conclusions are reached with the probit estimates in Panel B and C. Indeed, in a similar vain, Henderson (2003) finds a positive and robust correlation between number of plants and productivity in the US.[21] His findings therefore suggest that the correlation we find is not confined to Italy.

Second, we provide instrumental variables estimates. As an instrument for the number of firms, we need a variable that is correlated with the number of firms in a location but not with firm-level productivity in our sample. Following Ciccone and Hall (1996), we use as instrument the population at the LLS level in 1861 (the first Italian census). Clearly, the larger the population in a given location, the larger the number of firms, even if the primitive distribution of abilities is the same across areas. Moreover, given that location choices are persistent, due to moving costs and preference for "home", it can be maintained that population in 1861 is correlated with population today and thus with today's mass of firms. Indeed, a regression of the log of number of firms at the LLS-industry level on the population in 1861 at the LLS level produces an R^2 of 0.4. On the other hand, it is reasonable that local population size in the mid- 19^{th} century is not correlated with potential determinants of productivity in manufacturing over our sample period. This is our identifying assumption; it can be defended on the grounds that the industrial revolution in Italy did not even begin until the 1890s and that the biggest wave of industrialization occurred in the 1950s. If we presume that location choices before the industrial revolution were dictated mainly by agricultural fertility, and that this has no obvious relation to productivity in manufacturing in the later 20^{th} century, then the instrument satisfies the exogeneity condition.

Table 8 reports the results of a regression of firm efficiency on the number of firms in the LLS-sector, estimated by OLS (first three columns) and by IV (last three columns). For brevity, we only report the results with using provincial dummies as geographical controls. In Panel B and C we estimate a linear probability model to avoid the problems of IV estimates with probit models. All the regressions indicate that OLS and IV estimates are highly similar, with no evidence of a systematic bias in the OLS regressions. This suggests that our geographical controls are fine enough to capture any local factor that affects firms'

[21] Henderson's paper belongs to the literature assessing the industrial scope of spillovers, i.e. whether they are within or between industries, and he uses the number of own-industry plants as a measure of industrial concentration. Given that he does not aim at separating different sources of externalities, unlike us he does not control for the alternative channels. In line with what we find, he claims that the number of plants is the most robust indicator of intra-industry spillovers, and interprets it as evidence of knowledge externalities.

TFP and could be correlated with the number of firms in the area, lending support to our causal interpretation.

We have also performed robustness checks along the industry dimension. First, the two-digit classification we use might be too coarse and mix industries with different characteristics. While a more refined analysis is difficult because of limited sample size, particularly in the estimate of the production function coefficients, we can increase the number of industry controls in the baseline regression. We have run the basic regression on the pooled data including 296 dummies at the 4-digit level, finding no substantial difference in the estimates. A second problem is that we impose the same coefficient for the number of firms across different industries. While assessing learning opportunities at the industry level is beyond the scope of this paper, we have run a separate regression for each industry. In all industries we find that the number of firms has a positive and significant effect on TFP, with coefficients ranging from a low of 0.017 for basic metal to a high of 0.067 for leather and footwear.

6.2 Is the number of firms proxing for learning opportunities?

In this subsection we further corroborate the interpretation in terms of learning opportunities and knowledge spillovers. For example, the number of firms could be a proxy for competition, which would increase the productivity of the survivors through selection.[22] For learning externalities, we rely on specific theoretical predictions. Unfortunately, as noted by Duranton and Puga (2004), rigorous theoretical work on knowledge spillovers is rather scant. To our knowledge, Jovanovic and Rob (1989) and Eeckhout and Jovanovic (2002) are the only models that offer testable theoretical predictions.[23] The key to these models is that learning requires differences in knowledge among agents: if everybody knows the same thing, then there is no benefit from knowledge diffusion. This idea has two important

[22] This effect is often referred to in the literature as the "Porter effect", following Porter (1990) who studies the highly competitive and successful tile industry centered around Bologna in Italy. Syverson (2005) builds a model of the selection effect based on demand density and transportation costs, and tests it with data for the ready-mix concrete industry in the US, finding supporting evidence. A similar effect also emerges in the matching model of Lagos (2004).

[23] A related literature studies the diffusion of skills at the worker level - see Moretti (2004) for a survey. By studying the evolution of migrants' wages, Glaeser and Maré (2001) find evidence that cities favour the accumulation of human capital rather than simply increase individual productivity due to some thick-market externality.

implications that we can easily test with our data:

1. The least informed agents should gain more from spillovers, i.e. their productivity should be more sensitive to the amount of knowledge that can be accessed locally;

2. The dispersion of knowledge among firms should have a positive impact on productivity, because there is more to gain from knowledge diffusion.

We can test the first prediction by allowing the coefficient of the number of firms (our measure of potentially accessible knowledge) to differ across firms with different levels of knowledge. We therefore interact the number of firms with a dummy variable ("small") that is equal to 1 if the firm's capital stock is below the sample median, calculated at the LLS-industry level in each year, and add this interaction term to the regressions. The results, reported in Table 9, are strongly supportive of this prediction. In the OLS regressions, the elasticity of TFP to the number of firms is between 50 and 100 per cent higher for small firms across the different specifications. The increase in the estimates is very similar for the probability of being above the 75^{th} percentile and only slightly smaller (in absolute terms) for that of being below the 25^{th} percentile. Remarkably, the difference is significant at 1 percent in all 18 specifications, an indication of its robustness.[24]

To test the second implication, we need a measure of knowledge at the firm level. Following Eeckhout and Jovanovic (2002), we use the capital stock as a proxy of firm-level knowledge. First, the stock of capital is a measure of size, and it is reasonable to assume that larger firms have a higher stock of knowledge; second, insofar as knowledge is embodied in capital, it will be captured by this measure. We compute the standard deviation of the capital stock at the level of LLS-industry in each year, and use its log as an additional regressor. Table 10 reports the results of this set of exercises. For the OLS regressions, we find an elasticity of TFP to knowledge dispersion of about 0.04 (with higher values for the pooled regressions) and high levels of statistical significance. This finding is confirmed by the probit estimate, again more markedly for the probability of being above the 75^{th}

[24] As a further unreported check, we have also used an interaction between the capital stock and the number of firms, to avoid the arbitrariness of the split point of the regressions in Table 9. We find that the interaction is always negative (positive for the probability of being below the 25^{th} percentile), indicating that the larger the firm's stock of knowledge is the smaller is the effect of the number of firms.

percentile of the TFP distribution.

We have experimented with different measures of dispersion, using the ratio of the 90^{th} to the 10^{th} percentile of the capital stock, controlling for outliers and using a different functional form (i.e. in levels rather than in log), and alternative measures of firms' size, such as employment level. Again, results proved to be remarkably robust to all these changes. We conclude that, in line with theoretical predictions, knowledge dispersion has a positive effect on TFP. This is consistent with higher productivity levels being related to knowledge spillovers and learning externalities, and at odds with the competition effect, which should imply a negative impact of dispersion on productivity.

Taken together, the empirical exercises provide a compelling argument that differences in productivity across locations are at least partly due to the different learning opportunities and knowledge spillovers.

7 Conclusions

This paper has compared two alternative theoretical models of cluster formation, one based on the cost of setting up a business and another on local externalities. These models carry opposite implications on the sign of the correlation between entrepreneurial ability and entrepreneurial incidence, defined as number of firms over total population. This relation is negative if geographical agglomeration of firms is due to start-up costs and positive if it is due to differences in externalities. The models have also clear-cut implications for the relation between entrepreneurial incidence and the frequency mass at the two tails of the ability distribution. We have confronted these theoretical predictions with data on a large sample of Italian manufacturing firms coupled with information on the geographical clusters the firms belong to. We have found overwhelmingly that the start-up cost model is rejected and the externalities hypothesis supported.

When exploring the sources of externalities, we have found supporting evidence for intermediate input variety and especially for knowledge spillovers. We have indeed shown that the data agree with specific predictions of knowledge spillovers models. In future work we plan to investigate the modes through which they take place, focussing in particular on the possibility that in some locations in might be easier to accumulate entrepreneurial skills.

References

Banerjee, A. and Newman, A. (1994), 'Poverty, incentives and development', *American Economic Review Papers and Proceedings* **84**, 211–215.

Barlow, R. and Proschan, F. (1975), . *Statistical Theory of Reliability and Life Testing*, Holt, Rinehart and Winston, New York.

Bertrand, M. and Kramarz, F. (2002), 'Does entry regulation hinder job creation? evidence from the french retail industry', *The Quarterly Journal of Economics* **117**, 1369–1413.

Bertrand, M., Duflo, E. and Mullainathan, D. (2004), 'How much should we trust differences-in-differences estimates?', *Quarterly Journal of Economics* **119**, 249–275.

Blanchflower, D. G. and Oswald, A. (1998), 'What makes an entrepreneur?', *Journal of Labor Economics* **16**, 26–60.

Ciccone, A. and Hall, R. E. (1996), 'Productivity and the density of economic activity', *American Economic Review* **86**, 54–70.

Cingano, F. and Schivardi, F. (2004), 'Identifying the sources of local productivity growth', *Journal of the European Economic Association* **2**, 720–742.

Duranton, G. and Puga, D. (2004), Micro-foundations of urban agglomeration economies, *in* V. Henderson and J. Thisse, eds, 'Handbook of Regional and Urban Economics, vol. 4', Amsterdam: North Holland.

Eeckhout, J. and Jovanovic, B. (2002), 'Knowledge spillovers and inequality', *American Economic Review* **92**, 1290–1307.

Evans, D. S. and Jovanovic, B. (1989), 'An estimated model of entrepreneurial choice under liquidity constraints', *Journal of Political Economy* **97**, 808–827.

Fonseca, R., Lopez-Garcia, P. and Pissarides, C. (2001), 'Entrepreneurship, start-up costs and employment', *European Economic Review* **45**, 692–705.

Glaeser, E. L. and Maré, D. C. (2001), 'Cities and skills', *Journal of Labor Economics* **19**, 316–342.

Guiso, L. and Schivardi, F. (2005), 'Spillovers in industrial districts', *Economic Journal*. forthcoming.

Guiso, L., Sapienza, P. and Zingales, L. (2004), 'Does local financial development matter?', *The Quarterly Journal of Economics* **119**, 929–969.

Harrison, A. J. (1981), 'Earnings by size-a tale of two distributions', *Review of Economic Studies* **154**, 621–631.

Henderson, V. (2003), 'Marshall's scale economies', *Journal of Urban Economics* **53**, 1–28.

Henderson, V. and Thisse, J. (2004), *Handbook of Regional and Urban Economics, vol. 4*, Amsterdam: North Holland.

Holmes, T. J. (1999), 'Localization of industry and vertical disintegration', *Review of Economics and Statistics* **81**, 314–325.

Jovanovic, B. and Rob, R. (1989), 'The growth and diffusion of knowledge', *Review of Economic Studies* **56**, 569–582.

Klapper, L., Laeven, L. and Rajan, R. (2004), 'Entry regulation as a barrier to entrepreneurship', *Journal of Financial Economics.* Forthcoming.

Lagos, R. (2004), A model of TFP, Federal Reserve Bank of Minneapolis Staff Report, 345.

Lazear, E. P. (2005), 'Entrepreneurship', *Journal of Labor Economics* **23**, 649–680.

Lucas, Jr, R. E. (1978), 'On the size distribution of business firms', *Bell Journal of Economics* **2**, 508–523.

Marshall, A. (1890), *Principles of Economics*, London:Macmillan.

McKenzie, D. J. and Woodruff, C. (2003), Do entry costs provide and empirical basis for poverty traps? Evidence from Mexican microenterprises, mimeo, Stanford University.

Michelacci, C. and Silva, O. (2005), Why so many local entrepreneurs?, mimeo. CEMFI and CEP.

Moretti, E. (2004), Human capital externalities in cities, *in* V. Henderson and J. Thisse, eds, 'Handbook of Regional and Urban Economics, vol. 4', Amsterdam: North Holland.

Olley, S. and Pakes, A. (1996), 'The dynamics of productivity in the telecommunications equipment industry', *Econometrica* **64**, 1263–1297.

Pakes, A. and Olley, S. (1995), 'A limit theorem for a smooth class of semiparametric estimators', *Journal of Econometrics* **65**, 295–332.

Porter, M. (1990), *The Competitive Advantage of Nations*, Free Press, New York.

Rosenthal, S. and Strange, W. (2004), Evidence on the nature and sources of agglomeration economies, *in* V. Henderson and J. Thisse, eds, 'Handbook of Regional and Urban Economics, vol. 4', Amsterdam: North Holland.

Saxenian, A. (1994), *Regional advantage: Culture and competition in Sylicon Valley and Route 124*, Harvard University Press.

Schumpeter, J. (1911), *The Theory of Economic Development*, Harvard University Press, Cambridge, Mass.

Steindl, J. (1990), *Small and Big Business. Economic Problems of the Size of Firms*, Blackwell.

Syverson, C. (2005), 'Market structure and productivity: A concrete example', *Journal of Political Economy* **112**, 1181–1222.

Figure 1: Number of firms in the Italian LLSs

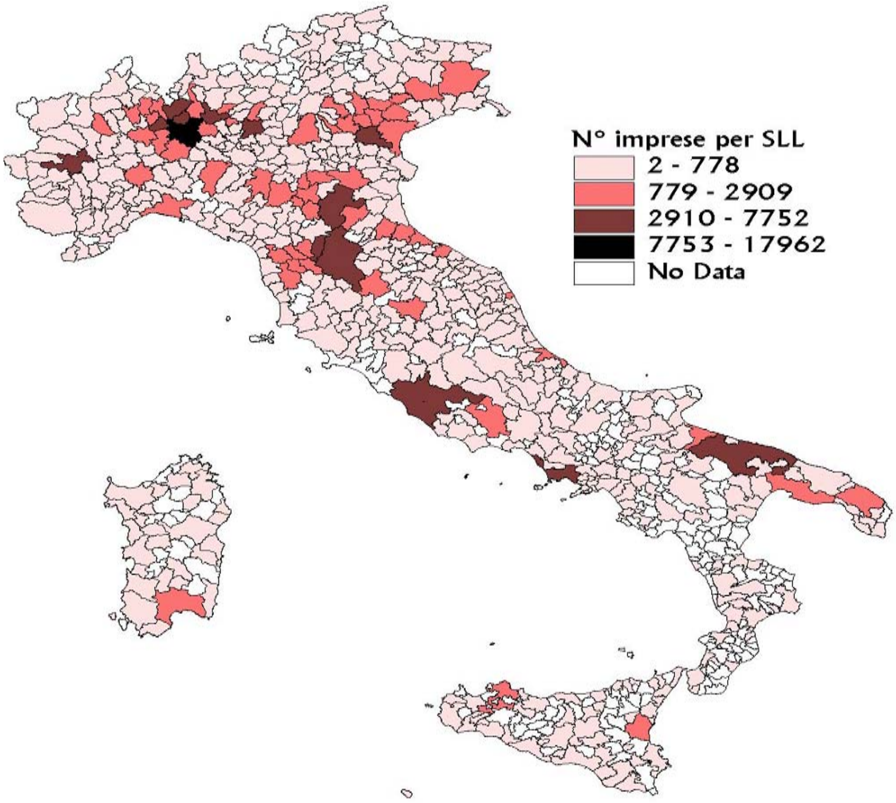

Figure 2: Distribution of log entrepreneurial ability for high (above the median) and low density (n. of firms over population) LLS

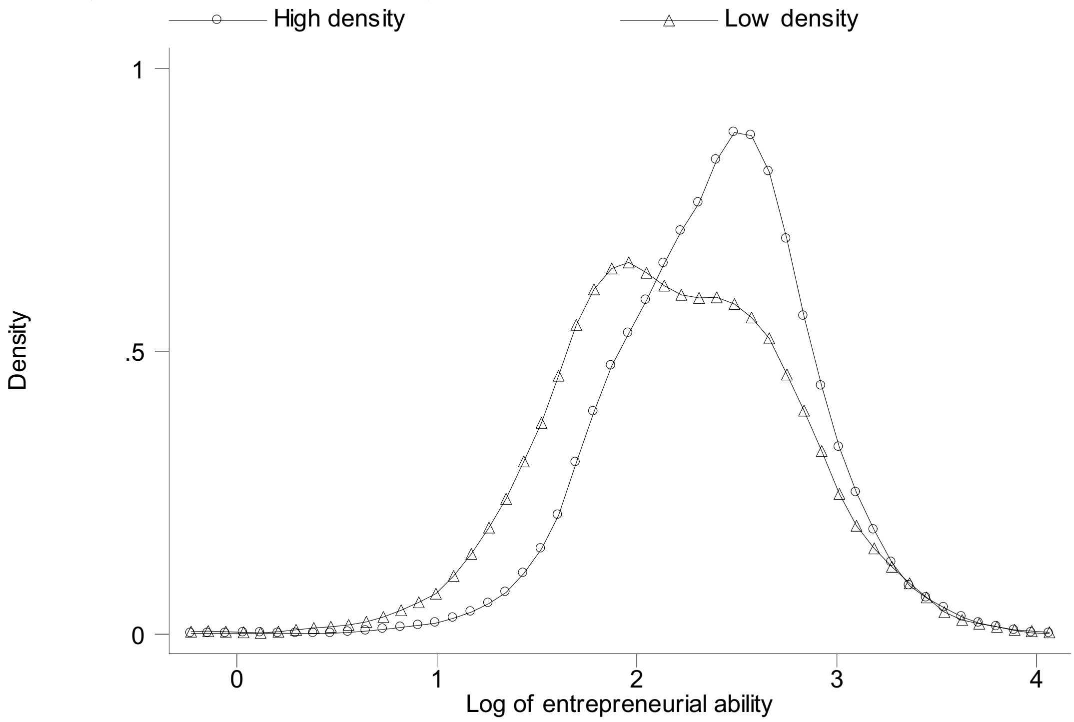

Table 1: Descriptive statistics for 1991

Panel A: Firms' characteristics (CB data)

Industry		Value added		No. employees		Capital Stock		No. obs.
		Mean	S.D.	Mean	S.D.	Mean	S.D.	
1	F	4856	20727	96	397	9891	35821	1367
2	T&C	2823	8183	83	230	4287	10282	2196
3	L&F	1659	2652	54	93	1611	3042	773
4	W&C	1786	2487	56	68	3130	5918	1110
5	T&Gl	4036	10981	87	201	9854	37443	1190
6	BM	6723	47871	162	1109	17499	118693	666
7	Mach	4487	27909	112	463	5739	42705	5262
8	Chem	7448	27959	129	461	14878	79981	1892
9	P&P	4454	15941	92	306	7404	29249	934
10	TEq	21021	174741	595	5172	38613	382168	447
	Total	4840	37277	114	969	8433	80014	15837

Panel B: Number of firms by LLS-industry (INPS data)

Industry		Average	S.D.	Max	Min .
1	F	62.5	89.7	722	1
2	T&C	150.5	270.4	2501	2
3	L&F	94.6	173.7	1159	1
4	W&C	89.3	148.3	1529	2
5	T	38.2	53.8	391	1
6	BM	21.1	48.6	374	1
7	Mach	234.2	568.0	8392	1
8	Chem	44.4	113.1	1636	1
9	P&P	71.5	221.4	2556	1
10	Teq	13.0	21.2	166	1
	Total	92.1	269.3	8392	1

Note: Value added and the stock of capital are in thousands of euros (at 1991 prices). Sectoral classification: F=Food, beverages and tobacco; T&C=Textiles and clothing; L&F= Leather and footwear W&C=Wood, products of wood and cork; T&Gl=Timber, construction materials and glass; BM=Basic metals; Mach=Metal products, machinery and equipment; Chem=Rubber, plastic and chemical products; P&P=Paper, printing and publishing; TEq=Transportation equipment

Table 2: Production function coefficients: factor share and direct estimates

Industry		Factor shares		Direct estimates		
		β	α	β	α	$\alpha+\beta$
		[1]	[2]	[3]	[4]	[5]
1	F	0.56	0.44	.63*** (.005)	0.39*** (.066)	1.02 (.064)
2	T&C	0.60	0.40	0.58*** (.003)	0.37*** (.035)	0.95 (.036)
3	L&F	0.61	0.39	0.62*** (.005)	0.43*** (.091)	1.05 (.091)
4	W&C	0.63	0.37	0.70*** (.005)	0.35*** (.077)	1.05 (.076)
5	T&Gl	0.58	0.42	0.67*** (.005)	0.37*** (.080)	1.04 (.078)
6	BM	0.65	0.35	0.60*** (.007)	0.33*** (.057)	0.93 (.054)
7	Mach	0.67	0.33	0.72*** (.002)	0.28*** (.013)	1.00 (.012)
8	Chem	0.60	0.40	0.70*** (.004)	0.29*** (.044)	0.99 (.043)
9	P&P	0.66	0.34	0.72*** (.005)	0.32*** (.039)	1.04 (.035)
10	TEq	0.74	0.26	0.70*** (.008)	0.26* (.144)	0.96 (.144)

Note: α is the capital coefficient and β the labor coefficient. The first estimates use the traditional Solow approach, the second the direct estimation of the production function coefficients using the Olley and Pakes (1996) procedure. Standard errors in parenthesis. Standard errors for the capital coefficient and for the sum of the coefficients computed by a bootstrapping procedure based on 150 replications. In column 4 and 5, *** indicates significance at 1%, ** at 5% and * at 10% (assuming normality for the bootstrapped standard errors). In column 6, the null $H_0 : \alpha + \beta = 1$ is never rejected at standard significance levels. See Table 1 for the industry labels.

Table 3: Ability and other characteristics by density of LLS

Variable	Total sample		High-density LLS		Low density LLS	
	Mean	S.D.	Mean	S.D.	Mean	S.D.
Ability	2.35	0.50	2.39	0.48	2.20	0.56
$I_{[Ability<25\%]}$	0.25	0.43	0.22	0.41	0.36	0.48
$I_{[Ability>75\%]}$	0.25	0.43	0.26	0.44	0.20	0.40
No. firms in LLS	7.36	1.42	7.62	1.28	6.37	1.47
No. firms in LLS-ind.	5.49	1.82	5.81	1.72	4.28	1.67
No. workers in LLS	10.02	1.68	10.33	1.54	6.47	2.35
No. workers in LLS-ind.	8.07	2.27	8.49	2.05	6.47	2.35
Int. inputs/VA in LLS	.88	.25	.87	.21	.89	.37
Int. inputs in LLS-ind.	.83	.43	.83	.37	.83	.60

All non dichotomous variables in log. High-density LLS are defined as those with the total number of firms over the population in the LLS above the median value by LLS. Ability is the estimate of TFP. $I_{[Ability<25\%]}$ is 1 if the ability is below the 25^{th} percentile of the ability distribution and zero otherwise. Ability and capital stock is from the CB sample; the number of firms and of workers are computed from the INPS dataset (the population).

Table 4: Firm efficiency and entrepreneurial share in the LLS

	Panel A. Dependent variable: log TFP, OLS estimates					
ES_{TOT}	6.28*	9.00***	8.38***	3.45	6.07**	5.52***
	(3.43)	(3.83)	(1.34)	(2.33)	(2.46)	(0.93)
R^2	.38	.38	.40	.40	.41	.42
No. obs.	15,837	27,173	137,907	15,837	27,173	137,907
	Panel B. Dependent variable: $I_{[Ability<25\%]}$, Probit estimates					
ES_{TOT}	−7.39***	−10.72***	−8.72***	−3.32	−6.52***	−4.44***
	(2.53)	(2.70)	(0.92)	(2.34)	(2.14)	(0.89)
Pseudo R^2	.03	.04	.03	.04	.06	.05
No. obs.	15,837	27,173	137,907	15,832	27,173	137,907
	Panel C. Dependent variable: $I_{[Ability>75\%]}$, Probit estimates					
ES_{TOT}	2.46	3.65	3.90***	1.51	3.15	3.52***
	(3.22)	(3.53)	(1.14)	(2.61)	(2.20)	(0.83)
Pseudo R^2	.01	.01	.01	.03	.04	.03
No. obs.	15,837	27,173	137,907	15,810	27,173	137,907
Sample	CS	F. Avg	Pool	CS	F. Avg	Pool
Geo ctrl	MA	MA	MA	Prov	Prov	Prov

Dependent variable: ability in Panel A (OLS estimates); $I_{[Ability<25\%]}$ in Panel B; $I_{[Ability>75\%]}$ in Panel C (Probit estimates). ES_{TOT} is the total number of firms over the population in the LLS. CS is the cross-section for 1991; Firm avg is the average over time at the firm level, after having de-trended all variables with a full set of year dummies; Pool is the whole sample with all firm-year observations. The geographical controls are Macro Area (MA, 4 dummies) and Provinces (Prov, 103 dummies). All regressions include industry and time dummies. Standard errors adjusted for clustering al the LLS-year level. ***indicates significance at 1%, ** at 5%, * at 10%.

Table 5: Firm efficiency and entrepreneurial share in the LLS-sector

	Panel A. Dependent variable: log TFP, OLS estimates					
ES_{TOT}	0.67 (2.23)	1.72 (2.19)	1.42 (.79)	−1.41 (2.30)	−.93 (2.11)	−0.47 (0.83)
ES_{SECT}	11.82* (6.58)	10.00** (5.93)	13.76*** (2.32)	8.52*** (3.13)	11.45*** (2.69)	10.22*** (1.33)
R^2	.39	.38	.40	.40	.41	.42
No. obs.	15,837	27,173	137,907	15,837	27,173	137,907
	Panel B. Dependent variable: $I_{[Ability<25\%]}$, Probit estimates					
ES_{TOT}	−4.22** (2.16)	−7.61*** (1.88)	−6.13*** (.75)	0.23 (2.53)	−2.25 (2.17)	−1.94** (0.98)
ES_{SECT}	−6.42 (4.64)	−6.13 (4.79)	−5.20*** (1.68)	−6.34** (3.09)	−6.66** (2.74)	−4.31*** (1.17)
Pseudo R^2	.03	.04	.03	.04	.06	.05
No. obs.	15,837	27,173	137,907	15,832	27,173	137,907
	Panel C. Dependent variable: $I_{[Ability>75\%]}$, Probit estimates					
ES_{TOT}	−4.10 (2.46)	−4.63** (2.25)	−3.77*** (0.80)	−3.87 (3.28)	−4.18* (2.27)	−2.89*** (1.00)
ES_{SECT}	13.45*** (5.30)	15.46*** (5.03)	14.74*** (1.69)	9.25** (3.76)	11.73*** (2.58)	10.77*** (1.17)
Pseudo R^2	.01	.01	.01	.03	.04	.03
No. obs.	15,837	27,173	137,907	15,810	27,173	137,907
Sample	CS	Firm Avg	Pool	CS	Firm Avg	Pool
Geo controls	MA	MA	MA	Prov	Prov	Prov

Dependent variable: ability in Panel A (OLS estimates); $I_{[Ability<25\%]}$ in Panel B; $I_{[Ability>75\%]}$ in Panel C (Probit estimates). ES_{TOT} is the total number of firms over the population on the LLS, ES_{SECT} is the total number of firms in the LLS-sector over the population in the LLS. Standard errors adjusted for clustering at the LLS-sector-year level. See Table 4 for the explanation of the specifications.

Table 6: Firm efficiency and externalities

	Panel A. Dependent variable: log TFP, OLS estimates					
No. firms	.051*** (.009)	.046*** (.009)	.048*** (.003)	.031*** (.007)	.024*** (.006)	.028*** (.003)
Int.Inputs/VA	.020 (.013)	.063*** (.011)	.028*** (.004)	.009 (.011)	.052*** (.010)	.015*** (.004)
Labor	−.007 (.005)	−.001 (.005)	−.005** (.002)	−.001 (.004)	.004 (.004)	.001 (.002)
R^2	0.39	0.40	0.41	0.41	0.41	0.42
No obs.	15,837	27,173	137,907	15,837	27,173	137,907
	Panel B. Dependent variable: $I_{[Ability<25\%]}$, Probit estimates					
No. firms	−.033*** (.008)	−.036*** (.007)	−.030*** (.003)	−.018*** (.008)	−.019*** (.006)	−.015*** (.002)
Int.Inputs/VA	−.012 (.011)	−.048*** (.010)	−.019*** (.004)	−.002 (.011)	−.039*** (.008)	−.008*** (.004)
Labor	.002 (.005)	.003 (.004)	.001 (.002)	−.004 (.004)	−.001 (.004)	−.004*** (.002)
Pseudo R^2	0.032	0.052	0.038	.047	0.066	0.049
No. obs.	15,837	27,173	137,907	15,832	27,173	137,907
	Panel C. Dependent variable: $I_{[Ability>75\%]}$, Probit estimates					
No. firms	.052*** (.008)	.052*** (.007)	.050*** (.003)	.036*** (.007)	.035*** (.006)	.033*** (.003)
Int.Inputs/VA	.038*** (.012)	.073*** (.010)	.048*** (.004)	.030*** (.012)	.066*** (.010)	.039*** (.004)
Labor	−.007 (.005)	−.006 (.005)	−.007*** (.002)	−.004 (.005)	−.004 (.004)	−.004*** (.002)
Pseudo R^2	0.027	0.031	0.026	0.038	0.042	0.037
No. obs.	15,837	27,173	137,907	15,810	27,173	137,907
Sample	CS	Firm Avg	Pool	CS	Firm Avg	Pool
Geo controls	MA	MA	MA	Prov	Prov	Prov

Dependent variable: ability in Panel A (OLS estimates); $I_{[Ability<25\%]}$ in Panel B; $I_{[Ability>75\%]}$ in Panel C (Probit estimates). The number of firms, capital and labor are in logs and computed at the LLS-industry level. Standard errors adjusted for clustering at the LLS-sector-year level. See Table 4 for the explanation of the specifications.

Table 7: Firm efficiency and externalities: controlling for LLS fixed effects

Panel A. Dependent variable: log TFP, OLS estimates

No. firms	.0115 (.009)	.018*** (.008)	.019*** (.003)
Int.Inputs/VA	.007 (.011)	.046*** (.010)	.006 (.004)
Labor	.005 (.005)	.004 (.004)	.002** (.002)
R^2	0.43	0.43	0.45
N obs.	15,837	27,173	137,907

Panel B. Dependent variable: $I_{[Ability<25\%]}$, Probit estimates

No. firms	−.007 (.008)	−.016*** (.007)	−.014*** (.003)
Int.Inputs/VA	−.000 (.011)	−.032*** (.009)	.001 (.004)
Labor	−.002 (.005)	−.001 (.004)	−.003* (.002)
Pseudo R^2	0.063	0.076	0.060
No. obs.	15,525	26,876	137,529

Panel C. Dependent variable: $I_{[Ability>75\%]}$, Probit estimates

N. firms	.015 (.009)	.027*** (.008)	.022*** (.003)
Int.Inputs/VA	.034*** (.013)	.069*** (.010)	.035*** (.004)
Labor	−.001 (.005)	−.004 (.005)	−.002*** (.002)
Pseudo R^2	0.044	0.049	0.044
N. obs.	15,837	26,476	136,694
Sample	CS	Firm Avg	Pool
Geo controls	LLS	LLS	LLS

Dependent variable: ability in Panel A (OLS estimates); $I_{[Ability<25\%]}$ in Panel B; $I_{[Ability>75\%]}$ in Panel C (Probit estimates). The number of firms, capital and labor are in logs and computed at the LLS-industry level. Standard errors adjusted for clustering at the LLS-sector-year-year level. See Table 4 for the explanation of the specifications.

Table 8: Firm efficiency and firm number: OLS and IV regressions

	Panel A. Dependent variable: log TFP					
No. firms	.030***	.030***	.029***	.035***	.019***	.025***
	(.004)	(.003)	(.001)	(.007)	(.006)	(.002)
R^2	.41	.40	.42	.40	.40	.42
No obs.	15,837	26,689	137,907	12,861	21,645	111,882
	Panel B. Dependent variable: $I_{[Ability<25\%]}$					
No.firms	−.023***	−.021***	−021***	−.024***	−.010*	−.017***
	(.004)	(.003)	(.001)	(.007)	(.006)	(.002)
R^2	.058	.079	.060	.066	.091	.069
No. obs.	15,837	26,689	137,907	12,861	21,645	111,882
	Panel C. Dependent variable: $I_{[Ability>75\%]}$					
No. firms	.030***	.031***	.029***	.034***	.028***	.027***
	(.004)	(.003)	(.001)	(.006)	(.004)	(.002)
R^2	.041	.043	.039	.046	.049	.044
No. obs.	15,837	26,689	137,907	12,861	21,645	111,882
Sample	CS	Firm Avg	Pool	CS	Firm Avg	Pool
Est. Meth.	OLS	OLS	OLS	IV	IV	IV

Dependent variable: ability in Panel A; $I_{[Ability<25\%]}$ in Panel B; $I_{[Ability>75\%]}$ in Panel C (linear probability estimates). The first three columns report OLS estimates; the fourth to sixth columns are IV estimates. The instrument is the population in the LLS in 1861. The Number of firms is in log and computed at the LLS-industry level. Standard errors adjusted for clustering at the LLS-sector-year level. See Table 4 for the explanation of the specifications.

Table 9: Knowledge spillovers and firm capital stock

	Panel A. Dependent variable: log TFP, OLS estimates					
No. firms	.037***	.028***	.036***	.024***	.012**	.022
	(.004)	(.004)	(.001)	(.003)	(.005)	(.001)
No. firms*Small	.014***	.029***	.016***	.014***	.030***	.016***
	(.002)	(.008)	(.001)	(.002)	(.008)	(.001)
R^2	.40	.40	.42	.41	.41	.43
No. obs.	15,837	27,173	137,907	15,837	27,173	137,907
	Panel B. Dependent variable: $I_{[Ability<25\%]}$, Probit estimates					
No. firms	−.028***	−.026***	−.027***	−.012***	−.012***	−.018***
	(.003)	(.004)	(.001)	(.004)	(.004)	(.001)
No. firms*Small	−.006***	−.012***	−.007***	−.006***	−.006***	−.007***
	(.001)	(.004)	(.001)	(.001)	(.001)	(.001)
Pseudo R^2	.03	.05	.04	.05	.05	.05
No. obs.	15,837	27,173	137,907	15,832	27,173	137,907
	Panel C. Dependent variable: $I_{[Ability>75\%]}$, Probit estimates					
No. firms	.037***	.028***	−.035***	.024***	.012**	.022***
	(.005)	(.003)	(.002)	(.004)	(.005)	(.001)
No. firms*Small	.017***	.030***	.018***	.017***	.030***	.018***
	(.002)	(.001)	(.001)	(.002)	(.008)	(.001)
Pseudo R^2	.04	.03	.04	.05	.04	.05
No. obs.	15,837	27,173	137,907	15,810	27,173	137,907
Sample	CS	Firm Avg	Pool	CS	Firm Avg	Pool
Geo controls	MA	MA	MA	Prov	Prov	Prov

Dependent variable: ability in Panel A (OLS estimates); $I_{[Ability<25\%]}$ in Panel B; $I_{[Ability>75\%]}$ in Panel C (Probit estimates). The Number of firms is in log and computed at the LLS-industry level. "Small" is a dummy =1 if the capital stock is below the median, calculated within each LLS-industry-year cell. Standard errors adjusted for clustering at the LLS-sector-year level. See Table 4 for the explanation of the specifications.

Table 10: Knowledge dispersion and productivity

	Panel A. Dependent variable: log TFP, OLS estimates					
No. firms	.039***	.038***	.040***	.026***	.024***	.026***
	(.005)	(.005)	(.002)	(.004)	(.004)	(.001)
Dispersion	.038***	.084***	.036***	.037***	.089***	.036***
	(.009)	(.008)	(.003)	(.008)	(.008)	(.003)
R^2	.38	.39	.40	.40	.41	.41
No. obs.	14,927	25,890	129,785	14,927	25,890	129,785
	Panel B. Dependent variable: $I_{[Ability<25\%]}$, Probit estimates					
No. firms	−.028***	−.029***	−.026***	−.018***	−.020***	−.017***
	(.004)	(.004)	(.001)	(.004)	(.003)	(.001)
Dispersion	−.009	−.030***	−.016***	−.011	−.035***	−.018***
	(.009)	(.007)	(.003)	(.008)	(.007)	(.001)
Pseudo R^2	.03	.05	.03	.04	.06	.04
No. obs.	14,297	25,890	129,785	14,916	25,888	129,785
	Panel C. Dependent variable: $I_{[Ability>75\%]}$, Probit estimates					
No. firms	.041***	.039***	−.041***	.030***	.027***	.027***
	(.006)	(.004)	(.002)	(.004)	(.005)	(.002)
Dispersion	.055***	.100***	.049***	.053***	.101***	.046***
	(.012)	(.012)	(.004)	(.012)	(.011)	(.004)
Pseudo R^2	.03	.03	.02	.04	.04	.04
No. obs.	14,927	25,890	129,785	14,884	25,869	129,773
Sample	CS	Firm Avg	Pool	CS	Firm Avg	Pool
Geo controls	MA	MA	MA	Prov	Prov	Prov

Dependent variable: ability in Panel A (OLS estimates); $I_{[Ability<25\%]}$ in Panel B; $I_{[Ability>75\%]}$ in Panel C (Probit estimates). The number of firms is in log and computed at the LLS-industry level; "Dispersion" is the log of the standard deviation of the capital stock within each LLS-industry-year cell. Standard errors adjusted for clustering at the LLS-sector-year level. See Table 4 for the explanation of the specifications.