American Economic Review: Papers & Proceedings 2012, 102(3): 447–452
http://dx.doi.org/10.1257/aer.102.3.447

Innovation in Space[†]

By Klaus Desmet and Esteban Rossi-Hansberg*

Why do private firms invest in innovation? Surely the answer to this question is related to the benefits that firms obtain from improving their technology or production processes. In particular, the rate of innovation has been traditionally linked to the capacity of firms to generate profits from these innovations. These profits, in turn, are used by firms to finance their investments in technology. In fact, this logic has been used in the economic growth literature to justify the presence of imperfect competition and monopoly power. How can firms finance their innovations if perfect competition eliminates all their profits?

In this short article we discuss a mechanism that can lead to private innovation by firms even in the presence of perfect competition (see Desmet and Rossi-Hansberg 2011a, where we introduced this logic to analyze development over space). The key elements are the presence of nonreplicable factors of production that are essential for the production process and innovations that are to a degree specific to these factors. In the presence of these nonreplicable factors, perfect competition in the product market, together with competition in the inputs markets, can lead to optimal innovation (absent intertemporal externalities). That is, the presence of these nonreplicable factors leads to optimal innovation even in the absence of any market power by firms. The main example of such a nonreplicable factor is land. Locations are given and cannot be reproduced. Even if other land is available in nearby locations, that land is not identical in that it is located somewhere else. A store on Fifth Avenue in New York or the Magnificent Mile in Chicago is quite a different business from a store in one of their side streets.[1]

The mechanism is simple. The standard logic that requires firms to have market power to finance their innovation assumes that, in the absence of such power, variable profits will be driven to zero, so that firms that innovate will obtain negative total profits. This happens because innovation is costly, and innovations diffuse freely. This logic does not apply in the presence of nonreplicable factors like land. Fix a location in space and think about the competition of firms for that location. The owner of that lot of land wants to maximize the gains from owning it, so he will sell or rent it to the highest bidder. As firms compete for that location, they bid as high as they can, namely, until they just break even. The key is that they can enhance their bid by innovating and making that land more valuable. If firms bid for land with a plan to innovate, then if they win, they obtain the benefits from the innovation since no one else can produce in that location but them. Firms will therefore invest in innovation as long as the gains from these investments outweigh the costs. The result is firms that invest optimally in order to compete for these nonreplicable and essential factors.

The argument above assumes that innovations are location specific. Hence, a bidder who wins the competition for a location guarantees that he will be the only one using the technology. This assumption can be relaxed to allow for imperfect appropriability of the innovation by others. This type of externality does not eliminate innovation as long as technologies invented for a particular location are not as productive in other locations, or as long as there are temporal lags

* Desmet: Department of Economics, Universidad Carlos III, Calle Madrid 126, 28903 Getafe (Madrid), Spain (e-mail: klaus.desmet@uc3m.es); Rossi-Hansberg: Department of Economics, Princeton University, 309 Fisher Hall, Princeton, NJ 08544-1021 (e-mail: erossi@princeton.edu). We thank Francisco Buera and Stephen Redding for very useful comments.

[†] To view additional materials, visit the article page at http://dx.doi.org/10.1257/aer.102.3.447.

[1] Besides land, other examples of nonreplicable factors are particular talents embedded in individuals. The entrepreneurial talent of Bill Gates or Warren Buffett, or the artistic talent of Pavarotti or Rothko, are hard to replicate, at least in the short run.

to appropriate them.[2] In these cases innovation will still be present, although it will, in general, not be optimal due to the externalities.

The argument has at least two important implications. First, when we think about growth through technological innovation in a spatial setup, land owners will appropriate a large fraction (all, if land is the only necessary and nonreplicable factor) of the gains from these innovations, even though they will not, in general, be the generators or promoters of the discoveries. This implies that all profits in excess of the cost of innovation go to land rents, an implication we explore in the empirical section of this article. Second, even though this mechanism does not, of course, prevent the presence of market power, it does imply that, absent externalities, market power is not necessary or desired in order to stimulate innovation. Hence, as in Boldrin and Levine (2004), it implies that the granting of monopoly power through patent policy is undesirable if there are no externalities. The main policy to promote innovation should be based on assigning clear property rights of land, as well as deregulating and eliminating frictions in the market for land and other factors with similar characteristics.

I. A Simple Model of Innovation in Space

Consider space in a continuum of locations $\ell \in [0, 1]$ with land density equal to one. Production requires land and labor. Production at location ℓ in some industry i by any firm is given by a constant-returns-to-scale (CRS) production function $A_i(\ell)\, F_i(L_i(\ell), N_i(\ell), \ell)$, where $A_i(\ell)$ denotes technology, $L_i(\ell)$ denotes the amount of land employed and $N_i(\ell)$ the amount of labor, in industry $i \in \{1, \ldots, I\}$ at location ℓ.[3] We assume that $F_i(\cdot)$ is strictly increasing and strictly concave in land and labor. Note that we also allow for exogenous differences in location quality by introducing the location name, ℓ, as an argument of the production function. Firms are ex ante identical and given CRS the boundary of the firm is indeterminate. For expositional purposes let $L_i(\ell) = 1$ all f, i, ℓ. Land then becomes a fixed factor of production, which determines the boundary of the firm.

$\overline{N}$ identical agents live where they work and are endowed with one unit of time which they supply inelastically in the labor market. The income of agents working at location ℓ is given by $w(\ell) + r$, where $w(\ell)$ denotes the location-specific wage and r the return to land ownership, which is assumed not to be location-specific as agents own a diversified portfolio of land. They derive utility according to a utility function $u(\{c_i\}_{i=1}^{I})$, which we assume homothetic, and increasing and strictly concave in each argument. This implies that given a set of prices for goods in all industries, $\{p_i(\ell)\}_{i=1}^{I}$, and the income of agents at location ℓ, there is a well defined set of consumption choices, $\{c_i(\ell)\}_{i=1}^{I}$ all $\ell \in [0, 1]$ of agents at all locations. Denote by $\overline{u}(\ell)$ the indirect utility implied by these choices. Agents can move freely in space, which implies that $\overline{u}(\ell) = \overline{U}$ all ℓ, where $\overline{U}$ is determined in equilibrium.

We assume free entry of firms that compete for land and labor and decide on their level of technology $A_i(\ell)$. Firms face a strictly convex cost of innovation. So a firm that wants to choose a level of technology $A_i(\ell)$ has to pay $\phi(A_i(\ell))$ units of output, where $\phi'(\cdot) > 0$ and $\phi''(\cdot) > 0$.

Competition for labor together with free mobility implies that $w(\ell)$ is determined by $\overline{u}(\ell) = \overline{U}$ and that the level of wages is such that the firms hire the total amount of labor in the economy, $\overline{N}$.

Competition for land at each location implies that land is assigned to the highest bidder. Namely, land rents at location ℓ are given by $R(\ell) = \max_{i \in \{1,\ldots,I\}} R_i(\ell)$ where $R_i(\ell)$ is the highest bid of potential entrant firms at location ℓ. $R_i(\ell)$ is thus given by

$$(1) \qquad R_i(\ell) = \max_{A_i, N_i} p_i(\ell)\, A_i\, F_i(1, N_i, \ell) - w(\ell)\, N_i - p_i(\ell)\, \phi(A_i),$$

since this is the land bid rent that implies that profits at location ℓ are equal to zero, $\pi_i(\ell) = 0$. Note that if $\pi_i(\ell) > 0$ at some location, then there would be an alternative entrant that would be willing to pay a higher rent and sacrifice some of these profits in order to obtain that location to produce.

[2] See Mansfield (1985) and, more recently, the work by Comin, Dmitriev, and Rossi-Hansberg (2011) that shows how particular technologies diffuse slowly over time and imperfectly over space.

[3] The finite number of industries, together with the continuum of locations, implies that in equilibrium there will be an interval of positive measure specializing in each industry.

The level of technology at location ℓ is therefore implicitly given by the first-order condition with respect to A_i, namely,

$$\frac{\partial \phi(A_i(\ell))}{\partial A_i} = F_i(1, N_i(\ell), \ell),$$

where $N_i(\ell)$ denotes the optimal amount of labor used at location ℓ, given implicitly by

$$\frac{\partial A_i(\ell)\, F_i(1, N_i(\ell), \ell)}{\partial N_i} = \frac{w(\ell)}{p_i(\ell)}.$$

Firms at a given location ℓ compete for land anticipating the price of their product and the wage. They decide given these prices their level of technology and employment. The firm that wins the bid at a given location gets to produce there. This is essential. Since technology is location specific, firms know that at a given location only one producer will win the competition for land, and so they will not be undercut by some other producer in the same location leading to losses as in the standard theory.

Of course, although land is nonreplicable within locations, it is replicable across locations. This implies that while individual firms face decreasing returns to scale, the industry as a whole is constant returns to scale.[4] That is, the industry expands along the extensive, rather than along the intensive, margin. As a result, in equilibrium price equals marginal cost, which in turn equals average cost, since it includes expanding both land and labor at land prices that include the gains and costs of innovation. The increasing marginal cost faced by an individual firm puts a bound on its size, ensuring a perfectly competitive market structure.

If land was not an input in production, price would be equal to the marginal cost, given $A_i(\ell)$, and so firms would make a loss equal to their investment in technology. Hence, their decision would be not to invest. This is the classic result in this literature that absent market power firms do not have incentives to invest. With land markets, the price of land that firms are willing to pay for a location takes into account the benefits and costs of localized innovation. Hence, even though in equilibrium prices will be equal to marginal cost, properly adjusted by transport costs, they will make zero profits because the rent they pay for land already was discounted by the cost of innovation and augmented by the benefits.

To find an equilibrium in this economy we just need to clear labor and goods markets. We assume that goods are costly to transport. Transporting an industry-i good from r to ℓ implies that only $e^{-\tau_i|r-\ell|}$ units of the good arrive in ℓ, where $\tau_i \geq 0$ all i. No arbitrage then implies that if location r exports the good to location ℓ,

$$p_i(r) = e^{-\tau_i|r-\ell|}\, p_i(\ell).$$

As in Rossi-Hansberg (2005), define $H_i(\ell)$ as the stock of excess supply in industry i at location ℓ. Then, by definition, $H_i(0) = 0$ and

$$(1) \quad \frac{\partial H_i(\ell)}{\partial \ell} = A_i(\ell)\, F_i(1, N_i(\ell), \ell) - \phi(A_i(\ell)) - c_i(\ell)\, N(\ell) - \tau_i |H_i(\ell)|,$$

where $N(\ell) = \sum_{i=1}^{I} N_i(\ell)$. Goods market clearing is then guaranteed by $H_i(1) = 0$ all i. Equilibrium in the labor market is guaranteed by $\int_0^1 N(\ell)\, d\ell = \overline{N}$, and the diversified ownership of land implies that $r = \int_0^1 (R_i(\ell)/\overline{N})\, d\ell$.

We can also study the problem of the planner that maximizes the utility of the representative agent in this economy. The planner chooses $\{c_i(\ell), N_i(\ell), A_i(\ell)\}_{i=1}^{I}$ so as to maximize

$$(3) \quad \max \int_0^1 u\Big(\{c_i(\ell)\}_{i=1}^{I}\Big)\, N(\ell)\, d\ell$$

$$\text{s.t}\; H_i(1) \geq 0 \text{ all } i,$$

$$\int_0^1 N(\ell)\, d\ell = \overline{N}.$$

PROPOSITION 1: *In equilibrium the level of technological innovation is efficient.*

PROOF:

Note that $A_i(\ell)$ only enters condition (2). Furthermore, the planner wants to relax the first constraint in (3) as much as possible, which is achieved by maximizing $\partial H_i(\ell)/\partial \ell$ for any $\{c_i(\ell)\}_{i=1}^{I}$. Hence the planner chooses $A_i(\ell)$ so as to maximize $\partial H_i(\ell)/\partial \ell$ which yields

[4] This is similar to the discussion in Hellwig and Irmen (2001).

$\partial \phi(A_i^*(\ell))/\partial A_i = F_i(1, N_i^*(\ell), \ell)$, where $N_i^*(\ell)$ is the optimal amount of labor in industry i at location ℓ. Since there are no frictions in the labor market, $N_i^*(\ell) = N_i(\ell)$. Hence, $A_i^*(\ell) = A_i(\ell)$.

The conclusion that innovation in an economy where land is an essential input in production is optimal is subject to some caveats. The main is that we have derived this implication in an economy without externalities. The presence of dynamic externalities *within* locations would not affect the conclusion as long as firms can write long term land use contracts with land owners or directly buy the land from them. This would imply that they will maximize the present discounted value of profits in that location and therefore will innovate optimally even if some alternative firm could access the same technology in that location in the future. The reason is that they have secured the land, thereby preventing the entry of other firms in that location. In the absence of these long term contracts or land purchases, this type of dynamic externalities will imply that investment will maximize only the land rents over the time horizon over which contracts can be effectively signed and enforced.

As we show in Desmet and Rossi-Hansberg (2011a), the presence of dynamic externalities *across* locations again implies that now firms will only invest so as to maximize the rents over the period that it takes other firms to adopt that technology. If, for example, inventions today are diffused over space next period, firms will invest so as to maximize only current rents. This implies that investment will in general be suboptimal. However, even in this case, as long as diffusion is not instantaneous, innovation will be positive, since firms still need to maximize their bids for land today. If, in contrast, externalities *across* locations occur immediately, innovation will still be positive only as long as the innovations are not as well suited when used in other locations.

II. Empirical Implications

The theory says that land rents are bid up until a firm's profits, net of the cost of innovation, are zero. The cost of innovation can then be computed as the difference between total output and total payments to factors of production including land. In this section, we use this theory to calculate a back-of-the-envelope measure of the cost of innovation at the level of US metropolitan areas (MSAs). To determine whether this gives us a reasonable estimate of the cost of innovation, we investigate whether it is increasing and convex in the level of productivity across MSAs.

In the theoretical model labor and land are the only factors of production; in the empirical exercise we also allow for capital. According to the theory, the cost of innovation in MSA m should be:

$$(4) \quad p(m)\phi(A(m)) = \frac{p(m)A(m)}{L(m)} F(L(m), N(m), K(m)) - w(m)\frac{N(m)}{L(m)} - q(m)\frac{K(m)}{L(m)} - R(m),$$

where $K(m)$ is the stock of capital, $q(m)$ denotes the rental price of capital, and the other variables have already been defined. Consistent with the theory that innovation is localized, the above expression assumes that the cost of innovation is paid per unit of land.

To see the logic of this exercise, it is key to understand the nature of innovation. If innovation were to take the form of buying intermediate inputs, or if innovation were to be done by a firm's employees, then the cost of innovation would already be included in the difference between total output and total payments to factors of production. In that case, the cost of innovation, as measured in (4) above, would be zero and, thus, have no relation with the level of technology. This would be a simple consequence of the accounting identity that requires the value of output to be equal to the value of factor payments. However, if it is the owner of the firm who innovates, the increased productivity would show up as higher profits, without affecting factor payments. If, then, part of the profits were to go to the owner as compensation for his efforts, with the remainder going to land rents, then the cost of innovation, as measured in (4), should be positive, and increasing and convex in the level of technology.

Clarifying further, it is not necessary that all (or even any) innovation is done by the owner, but any other type of innovation would yield no relation between the cost of innovation, as measured in (4), and the level of technology. This implies that we may very well find no relation, but if we do, it should be positive and convex.

Given the focus on local innovation, our ex ante hypothesis is that firm owners will be responsible for at least part of the productivity gains. What we have in mind is the local bank improving client relations, the local coffee shop making the decor more attractive, the music hall adjusting to the local community's taste, etc. In those examples, we are not talking about buying blueprints in the marketplace or inventing new products, but rather about local entrepreneurs and business owners improving local productivity.

We use data on 229 metropolitan statistical areas (MSAs) for the years 2005–2008. Output is measured as GDP. Payments to labor are taken to be the product of the number of jobs and the average wage per job. Payments to capital and land are computed as the user cost times the value of the capital stock and the land stock, which is measured as the sum of the value of nonresidential capital (including the land on which it is located) and the value of the housing stock.[5] The productivity level of an MSA is measured as TFP and is computed using data on output, the capital stock, and hours worked.[6] Once we have estimates for the different variables, we compute the cost of innovation as a residual, following equation (4).[7] As for the data sources, data on output, jobs, wages, and nonresidential capital come from the Bureau of Economic Analysis; data on the housing stock come from the American Community Survey; and data on hours worked come from the Current Population Survey.

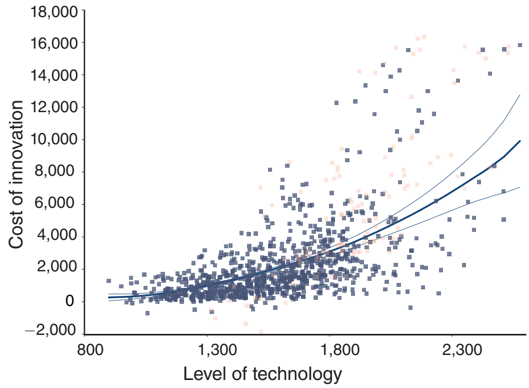

FIGURE 1. INNOVATION COST AND LEVEL OF TECHNOLOGY

Notes: The round markers indicate the levels of technology and the costs of innovation in 229 MSAs in years 2005–2009. The curves represent the predicted values (and the 95 percent confidence intervals) from the kernel regression. The crossed markers indicate the levels of technology and the costs of innovation in 29 MSAs over the period 2005–2009. They correspond to the robustness check where we use MSA-level information on the cost of land in the cost of housing.

The main curve in Figure 1 depicts the result of a kernel regression of the cost of innovation on the level of technology (using an Epanechnikov kernel with 0.5 bandwidth). The lighter curves represent the 95 percent confidence interval, and the round markers show the actual data. As predicted by the theory, the relation in the data is increasing and convex. This suggests that the difference between total output and total payments to factors of production is indeed related to the cost of innovation. We view this as evidence that the competition for land takes into account the local cost of innovation. That is, land prices are bid up until profits, net of innovation costs, are zero.

In this benchmark exercise we estimated the value of the land on which nonresidential structures are located by using information on the *average* share of land in the value of housing across MSAs, rather than the *individual* shares at the level of MSAs. We did so in order not to lose too many observations, since using data for individual MSAs would have reduced our sample from more than 225 MSAs to fewer than 50 (Davis and Palumbo 2008). Still, it seems key

[5] The value of nonresidential capital at the MSA level is proxied by allocating sectoral nonresidential capital stock data at the US level (from the National Economic Accounts from the BEA) to the different MSAs as a function of their sectoral weights (see Desmet and Rossi-Hansberg, 2011b, for more details). When doing so, we distinguish between the nonresidential capital stock excluding structures and nonresidential structures. In order to add the value of land to that of nonresidential structures, we use information on the average cost of land in the cost of housing across 50 MSAs from Davis and Palumbo (2008). The value of the housing stock is the sum of the value of owner-occupied houses and the value of rental properties from the American Community Survey. The user cost of capital, r, is set equal to 0.04 and reflects the sum of the interest rate and the depreciation, which are taken to be common across all MSAs.

[6] In particular, productivity is obtained by dividing output by the capital stock to the θ and hours worked to the $1 - \theta$, where $\theta = 0.33$.

[7] Since both the level of technology and the cost of innovation are measured as residuals, measurement errors might generate a mechanical relation between the two. However, nothing suggests it should be increasing and convex.

to see whether our result is robust to correcting for differences in land prices across MSAs. The data corresponding to that alternative estimation are given by the crossed markers in Figure 1. Consistent with the theory, the implied relation between the level of technology and the cost of innovation is clearly increasing, though somewhat steeper than in the benchmark exercise. Another concern is that we are not appropriately accounting for all potential congestion costs. Local business owners may want to be compensated for these costs. In a further robustness check, not shown here in the interest of space, we assumed that congestion costs are convex in city size, and on average equal to 5 percent of GDP. Under this alternative specification, we again found an increasing and convex relation between innovation costs and technology level, with a tighter fit of the data than in Figure 1.

III. Conclusion

This article has proposed an environment in which firms optimally choose to innovate, in spite of the market being perfectly competitive. The key elements driving this result are innovation being localized, land being nonreplicable, and land markets being competitive. Firms that bid for a certain location will want to innovate if that increases their bidding power. They will be willing to do so because if they win the bid and innovate, no one will be able to produce in that location except them. Perfect competition is ensured because the nonreplicability of land implies that firms face decreasing returns to scale. Given the spatial nature of innovation and the importance of land in production, the argument that innovation arises naturally in a competitive environment may be more widespread than previously thought. We present some evidence that innovation gains and costs are in fact embedded in land prices across cities. Although the model presented here is static in nature, it can easily be extended to a dynamic framework to study the evolution of economic activity over time and space. In Desmet and Rossi-Hansberg (2011a) we develop such a spatial endogenous growth model and use it to analyze the spatial evolution of the United States over the last decades.

REFERENCES

Boldrin, Michele, and David K. Levine. 2004. "2003 Lawrence R. Klein Lecture: The Case against Intellectual Monopoly." *International Economic Review* 45 (2): 327–50.

Comin, Diego, Mikhail Dmitriev, and Esteban Rossi-Hansberg. 2011. "The Spatial Diffusion of Technology." Unpublished.

Davis, Morris A., and Michael G. Palumbo. 2008. "The Price of Residential Land in Large US Cities." *Journal of Urban Economics* 63 (1): 352–84.

Davis, Morris A., and Michael G. Palumbo. 2008. "The Price of Residential Land in Large US Cities: Dataset." http://www.lincolninst.edu/resources/.

Desmet, Klaus, and Esteban Rossi-Hansberg. 2011a. "Spatial Development." Unpublished.

Desmet, Klaus, and Esteban Rossi-Hansberg. 2011b. "Urban Accounting and Welfare." Unpublished.

Hellwig, Martin, and Andreas Irmen. 2001. "Endogenous Technical Change in a Competitive Economy." *Journal of Economic Theory* 101 (1): 1–39.

Mansfield, Edwin. 1985. "How Rapidly Does New Industrial Technology Leak Out?" *Journal of Industrial Economics* 34 (2): 217–23.

Rossi-Hansberg, Esteban. 2005. "A Spatial Theory of Trade." *American Economic Review* 95 (5): 1464–91.