

PDF hosted at the Radboud Repository of the Radboud University Nijmegen

The following full text is an author's version which may differ from the publisher's version.

For additional information about this publication click this link.

<http://hdl.handle.net/2066/162779>

Please be advised that this information was generated on 2019-06-20 and may be subject to change.

Heterogeneous firms and cluster productivity: a neglected externality through survival of the weakest

Cornelis W. Haasnoot¹
(GIGA German Institute of Global and Area Studies, Germany)

&

Albert de Vaal²
(Institute for Management Research, Radboud University Nijmegen, Netherlands)

This version: August 2014

Preliminary version: please do not quote

Abstract

We argue that geographical clustering helps the weakest firms to survive. We model this neglected negative externality by adapting the heterogeneous firms model of Melitz (2003) to include firms that invest in R&D and firms that do not. Separating the chance of post-entry market exit into a system risk that is exogenous to all firms and a firm-specific risk that can be reduced by doing R&D, we find that only the most productive firms will invest in R&D. Incorporating knowledge spillovers to proxy for geographical proximity, the benefits from these R&D investments may spill over to other firms. This occurs either directly by reducing the firm-specific risk for non-R&D firm, or indirectly, by making R&D investments cheaper. The effects on innovation are different for these two cases: direct spillovers reduce innovation while indirect spillovers increase it. However, for both cases it holds that the effect on average productivity within the region is clearly negative.

Keyword: geographical clustering; R&D spillovers; heterogeneous firms; government policy.

JEL-codes: L11, O33, R11

¹Address: GIGA Institute of Asian Studies, Rothenbaumchaussee 32, 20148 Hamburg, Germany. E-mail: haasnoot@giga-hamburg.de.

²Corresponding author. Address: Radboud University Nijmegen, Department of Economics, P.O. Box 9108, 6500 HK Nijmegen, The Netherlands. Email: a.devaal@fm.ru.nl. URL: <http://www.ru.nl/fm/vaal/>.

1. Introduction

In this paper we want to draw attention to an unexpected negative effect of geographical clustering on regional output and growth through an inefficient composition of firms in the cluster. Firms differ by their intrinsic productivity, leading some firms to be larger and more profitable than others in the industry (e.g. Bailey et al., 1992, Jensen et al., 1997). However, the benefits of clustering will be relatively larger for firms at the low end of the productivity-profitability spectrum. This increases the market survival rate of low-productivity firms in particular, increasing the share of low-productivity firms in the cluster. The positive agglomeration effect of clustering for firms affects the composition of firms in the cluster, reducing the overall efficiency of the cluster (and the economy), even if the end result of agglomeration is still a net positive. We make this neglected externality of clustering explicit by setting up a model of heterogeneous firms à la Melitz (2003) where firms may invest in R&D to reduce the risk of leaving the market and where agglomeration benefits take the form of knowledge spillovers.

In innovative environments firms need to constantly rethink their strategy and adapt their products to changed circumstances in the market. In these environments, firms that have a larger existing knowledge base are better able to assess, access and address knowledge that is developed externally (Cohen and Levinthal, 1990). On the other hand, a larger existing knowledge base will also mean that more knowledge spills out of the firm into the rest of the cluster (Shaver and Flyer, 2000), undermining the competitive advantage that stems from this knowledge base in the first place (Alcácer and Chung (2007) , Pouder and StJohn, 1996). This suggests that there are differences between firms within a cluster as far as utilizing knowledge goes. In our model, we will look at both effects separately: first we will look at a spillover which disproportionately benefits firms who are not innovative, and then we will look at a spillover which disproportionately benefits firms who are innovative. In line with the (empirical) literature on firm survival we assume that firms make a continuous effort to stay in the market (Caves, 1998; Cefis and Marsili, 2005; Pérez et al., 2004).

To include these insights in a Melitz-type of framework, we split the risk of firm exit into two different and independent factors. On the one hand, firms face a systemic risk of market exit that is comparable to the fixed and exogenous probability of firm death in Melitz (2003). On

the other hand, firms face a firm-specific risk, which they can control by doing R&D. This is the effort that is required for a firm to remain a strong player in the market.³ In the terminology of the literature on product innovations our concept of innovation comes closest to what is known as ‘incremental product innovation’.⁴ As an example could serve the Apple or Samsung business models, with a new version of the same product coming out each year. The benefits of R&D in this respect depend on the firm’s productivity level⁵. Firms that draw a high productivity level have a stronger incentive to stay in the market than firms drawing a low productivity level. In our modeling specification this will imply R&D becomes a binary choice: firms above an endogenously determined productivity threshold will invest in R&D to stay in the market, while firms below the threshold will leave the market after one period.

One of the alleged advantages of the geographical proximity within clusters is that knowledge spillovers may occur. Knowledge spillovers are seen as one of the Marshallian *raisons d’être* of clusters and there is a vast literature that has argued in favor of a positive relation between knowledge spillovers and geographical proximity.^{6,7} In our model we analyze two different perspectives on how these knowledge spillovers could occur. The first perspective is that knowledge spillovers imply that R&D reduces the firm-specific exit risk of firms that do not invest as well. This perspective sees knowledge spillovers as the direct imitation of R&D

³ As such, our approach to R&D is also related to Grossman and Helpman’s (1991) modelling of quality ladders: continuous technological progress is required to remain in the market, and those firms that do not climb to the next rung will inevitably fall off.

⁴ In their critical review of the plethora of definitions regarding product innovation, Garcia & Calantone (2002: 123) define incremental innovations as “products that provide new features, benefits, or improvements to the existing technology in the existing market.”

⁵ The argument we develop has its antecedents in Melitz (2003), who provides a formal framework for studying the effects of productivity heterogeneity amongst firms in a model of international trade. In his model, more productive firms benefit more from trade than less productive firms. Baldwin and Okubo (2006) use a similar argument to ours when investigating the effects of regional subsidies on firm location. Our work also formalizes the possibility suggested by Swann (2006) that governmental policy encouraging clusters may attract firms from the ‘shallow end of the spectrum’ (p. 269).

⁶ Jaffe and Trajtenberg (2002), for instance, looked at patent citations in relation to geographical proximity finding that patents are more likely to build upon previous patents if they were filed near to each other. Other contributions showcasing the importance of geographical proximity for knowledge spillovers are Jaffe et al. (1993), Audretsch and Feldman (1996), Zucker and Darby (1998), Keller (2002), and Asheim and Gertler (2005). Audretsch & Feldman (2004) provide an overview of the literature.

⁷ We refrain from explicitly modelling economic clustering and see the existence of knowledge spillovers as a proxy for it. This helps us make our main point without entangling ourselves in the analytical complexity of Krugman-like models of geographical economics. Though we recognize that the negative effect of cluster composition will also have an impact on the incentives for firms to form a cluster, it suffices for our purposes to keep clustering exogenous. Our main interest is to unravel a neglected externality due to the heterogeneity of firms within clusters and not to analyze the determinants of clustering itself.

practices. Such spillovers could for instance occur as a result of R&D labor turnover effects⁸ or due to the inspection of patents⁹. In line with the literature on absorptive capacity (e.g. Findlay, 1978, Cohen & Levinthal, 1989, 1990), we assume that without engaging in R&D themselves knowledge spillovers by direct imitation will be imperfect and we use appropriate parameterization to account for that. The second perspective on knowledge spillovers we analyze is that the benefits of firms investing in R&D reach other firms in the cluster indirectly, by making R&D investments cheaper for them (and for investing firms). This perspective bears a relation with the idea that knowledge creation entails a positive externality as in models of endogenous growth and trade (e.g. Romer 1986, Grossman and Helpman, 1991). We use a standard parameterization to account for these effects, but also verify outcomes in case the spillovers of indirect imitation depend on the number of firms in the cluster. As we show, the effect on the extent of innovation in the cluster is different for these two different cases. Spillovers by direct imitation reduce innovation, whereas spillovers by indirect imitation increase it. In both cases the effect on average productivity within the cluster is negative when moving from no spillovers to full spillovers. However, there may be an intermediate level of spillovers where average productivity of the cluster is higher than without spillovers. Knowledge spillovers help less efficient firms to survive, reducing the efficiency of the cluster at large. The knowledge spillovers that come along with geographical clustering thus gives rise to a neglected externality. While having positive effects for individual firms, the effects on cluster performance are potentially negative. For governmental policy this implies that stimulating geographic clusters may have less positive effects than previously thought.

Our focus on how the composition of firms in a cluster may affect cluster productivity is different from most of the literature on geographical clustering of industrial activity. Clusters have received abundant academic attention, but the focus has been primarily on how clusters

⁸ Empirical investigations of these include Maliranta, Mohren and Rouvinen (2009), Møen (2005) and Almeida and Kogut (1999).

⁹ Patent applications must include citations that represent the base of knowledge that the patent was built on. The patent examiner determines which patent citations are necessary or pertinent, as they delineate the exact property rights related to the patent. Since most citations are not self-citations (they do not refer to patents owned by the same company), these represent knowledge spillovers from other firms or institutions. Jaffe, Trajtenberg and Fogarty (2000) surveyed patent owners to investigate whether or not patent citations were a valid proxy for knowledge spillovers and found that while a significant number of patent citations were not associated with knowledge spillovers, patent citations still function as a proxy.

affect firm performance. Research on clusters includes general advancements that try to establish the broader economic and geographical reasons for clustering (Fujita et al., 1999; Brakman et al., 2009), empirical literature to establish these reasons in reality (Baptista and Swann, 1998; Bell, 2005; DeCarolis and Deeds, 1999; Molina-Morales and Martinez-Fernandez, 2003), research that deals with the evolution of clusters over time (e.g. Iammarino & McCann, 2006; Feldman et al, 2005) and scholarly articles on the ways a cluster affects the operations of firms (e.g. Wennberg & Lindqvist, 2010; Molina-Morales and Martínez-Fernández, 2009 ; Pouder and St. John, 1996; Boschma, 2005) and how this could differ depending on the characteristics of firms (McCann and Folta, 2008, Braunerhjelm and Feldman, 2006). Clustering has also received much attention in the policy arena, where it has been deemed a desirable trend to foster regional development and economic growth (e.g. European Commission, 2003; Sölvell et al., 2003).

The effect of firm composition on cluster productivity has hardly received attention. One exception is Shaver and Flyer (2000), who look at the effect that agglomeration advantages have on firm behavior. Their argument is that weaker firms will enter the cluster to benefit from it, while stronger firms will stay away to survive on their own. Our paper also looks at the diverging effects clusters have on different types of firms, but we go one step further by linking it back to the effect that this has on the cluster as a whole. Policy analysis has to take these effects into account, because policies promoting clusters might have a negative effect on cluster composition. This negative effect would then counteract the intended effects of the policy. Looking at differential effects of clustering on firms would also make it possible to reconcile the positive and negative results in one theoretical framework. By decomposing the effects of clustering into positive and negative effects, economic policy can target these decomposed effects instead of targeting clusters as a whole. In that sense, we follow the recommendations of Nathan and Overman (2013), who argue that “an ‘agglomeration policy’ approach (...) seeks to develop interventions that increase the benefits of urban location while damping down the disadvantages.” .

A novelty of our paper regarding the literature of heterogeneous firms is our focus on R&D affecting the chance of exit from the market. When addressing R&D investments in a heterogeneous firms framework à la Melitz (2003), the focus of the literature has been on purposeful investments in productivity enhancement. For instance, in the general equilibrium

model of Atkeson & Burstein (2010) heterogeneous intermediate goods producing firms that survive an exogenously given probability of market exit invest in R&D to enhance the probability of experiencing a positive productivity shock next period. The authors refer to this type of R&D expenditure as process innovation. Product innovation is also included in their framework, but takes the normal form of firms making a fixed market entry cost before finding out their (initial) productivity. Contributions using a similar set-up are Atkeson and Burstein (2011), and Burstein and Melitz (2011). Other papers that have incorporated explicit decision making of firms regarding their productivity level in a Melitz-type of framework are Yeaple (2005) and Bustos (2011). In these papers, each firm can choose to pay an additional fixed cost that would lead to a reduction in marginal costs that is equivalent to an increase in productivity. Vannoorenberghe (2009) makes the productivity improvement that is reached dependent on how much a firm decides to invest. Van Long et al. (2001) discuss process innovation in a framework that is related to Atkeson and Burnstein's (2010) paper. In their set-up firms decide on R&D investments *prior* to knowing their stochastic productivity level. Moreover, they use Cournot competition as their modeling framework.

There are some papers that have looked at the combination of heterogeneous firms and clustering. For example, Combes et al. (2012) look at the effects that clustering has on the distribution of firm productivities. They argue that there are two effects: a selection effect and an agglomeration effect. The selection effect is a left truncation because the environment within the cluster is more competitive than the environment outside the cluster, eliminating the least productive firms; the agglomeration effect is the general positive effect of clusters, in their case higher labor productivity because of more interaction between employees within a cluster. This causes a right-shift of the productivity distribution, because all firms gain, plus a dilation effect, because the effect of increased productivity is stronger for firms that are already more productive. They find empirical evidence for an agglomeration effect, but not a selection effect. Our paper analyzes the effects of spillovers in a similar fashion, and we find that spillovers of themselves can dilate the distribution of firms on the left-hand side, which would cancel out the left truncation of the selection effect.

This paper also has some implications for the literature on firm survival. There are a number of papers which explicitly look at firm performance within clusters. For an overview of the effects of clustering on firms, see Rocha (2004). Many papers, like Wennberg and Lindqvist

(2010), have found a positive effect of clusters on firm survival. Wennberg and Lindqvist even find positive effects of clustering on firm performance. However, they use firm fixed effects, which mean that they are effectively ignoring the exact effect that we describe here, that firms within clusters tend to be less productive. Our point is that the presence of a cluster does not only affect firm performance, but also firm characteristics. Empirical analyses might benefit from taking these interdependencies into account.

The structure of this paper is as follows: Section 2 offers the benchmark model, deriving results on aggregate productivity in the absence of clustering. Section 3 applies the model to industrial clusters by including knowledge spillovers. Section 4 offers several extensions of the model and Section 5 concludes.

2. The benchmark model

To investigate the consequences of firm heterogeneity for the aggregate productivity effects of clustering we adapt the standard Melitz (2003) model of heterogeneous firms and include knowledge spillovers from firm level R&D investments as a proxy of geographical proximity. As argued, we see R&D as investments firms make to reduce their firm specific risk of market exit. In the original Melitz model, the probability of firm exit is denoted by δ and is considered to be fixed and exogenous. This could be seen to reflect the industry-specific risk of exit firms face. To set this apart from the risk firms could influence, we assume that the chance of firm exit consists of two different and independent components. On the one hand, we consider some form of strong systemic risk: that is, the risk that a firm goes out of business for reasons it cannot control. We parameterize this systemic risk by ζ . This type of risk is comparable to the exogenous chance of firm exit in the Melitz model. On the other hand, we assume a firm-specific risk ε , which can be controlled by doing R&D. The firm-specific risk can be seen as the constant innovation that is required for a firm to remain a strong player in the market.

The entry and exit of firms in industry is essentially the same as in Melitz (2003). Firms are uncertain about their inherent productivity and therefore base their entry decision on a comparison of the one-time market entry costs and their expected profits of post-entry production. Once firms enter, they find out about their actual productivity and decide whether

to stay in the market (in case of positive profits) or to exit (in case their productivity level is too low to sustain positive profits)¹⁰. However, firms also face a risk to exit the market – the aforementioned exogenous systemic risk ζ and the endogenous firm-specific risk ε ($0 \leq \zeta, \varepsilon \leq 1$). The consolidated chance of survival in the market thus becomes $(1 - \zeta)(1 - \varepsilon)$, implying a chance of post-entry exit δ of:

$$\delta \equiv \zeta + \varepsilon - \zeta \cdot \varepsilon \quad (1)$$

We assume that firms can invest in R&D to lower their firm-specific risk. The decision to invest depends on a comparison of the additional profit reached over time and the additional investment cost. Suppose that ε is a linearly declining function in R&D investments f_{RD} with $\varepsilon(0) = 1$ and $\varepsilon(\bar{f}_{RD}) = 0$:

$$\varepsilon(f_{RD}) = \frac{\bar{f}_{RD} - f_{RD}}{\bar{f}_{RD}} \quad (0 \leq f_{RD} \leq \bar{f}_{RD}).$$

We assume that the f_{RD} chosen has to be incurred each period, reflecting the idea that staying in the market requires a continuous R&D effort. In the absence of time discounting, the pre-entry expected value of a firm is therefore

$$v(\varphi) = \max \left\{ 0, \sum_{t=0}^{\infty} (1 - \delta)^t \pi(\varphi) \right\} = \max \left\{ 0, \frac{\pi(\varphi) - f_{RD}}{\zeta + \varepsilon(f_{RD}) - \zeta \varepsilon(f_{RD})} \right\}$$

where $\pi(\varphi)$ is a firm's profit level, which depends on its (yet unknown) productivity level $\varphi > 0$ (see below). A firm's optimal investment level is determined by taking the derivative of $v(\varphi)$ w.r.t. f_{RD} . Taking into account the specification of $\varepsilon(f_{RD})$, this yields:

¹⁰ The possibility that firms could invest to enhance their productivity in order to make it profitable to stay in the market has been investigated by Vannoorenberghe (2009). As he shows, however, this does not affect the cutoff productivity level of profitable entry. The optimal level of investment is strictly increasing in firm size, making the optimal investment of the cutoff firm independent of the level of cutoff productivity (propositions 1 and 2). If it becomes profitable for less productive firms to invest, it is even more profitable for more productive firms to invest.

$$\frac{dv(\varphi)}{df_{RD}} = \frac{(1 - \zeta)(\pi(\varphi) - \bar{f}_{RD})/\bar{f}_{RD} - \zeta}{[\zeta + \varepsilon(f_{RD}) - \zeta\varepsilon(f_{RD})]^2}. \quad (2)$$

The sign of (2) is independent of f_{RD} , implying that it is either optimal for a firm to fully invest in R&D or not to invest at all. If $(1 - \zeta)(\pi(\varphi) - \bar{f}_{RD})/\bar{f}_{RD} - \zeta$ is negative, a firm will decide not to invest in R&D and $f_{RD} = 0$. If it is positive or zero (the latter by assumption), a firm will decide to invest fully: $f_{RD} = \bar{f}_{RD}$.¹¹ However, the decision to invest in R&D clearly depends on a firm's productivity level. Equation (2) is positive if $\pi(\varphi) \geq \bar{f}_{RD}/(1 - \zeta)$, implying $\varphi_{RD}^* = \inf(\varphi | (1 - \zeta)\pi(\varphi) > \bar{f}_{RD})$ as the minimum required productivity level for a firm to invest in R&D. Only for the more profitable firms is it worthwhile to invest to stay in the market.¹²

If firms invest in R&D, they will reduce their firm-specific risk ε to zero, leaving the systemic risk ζ as the only exit risk they face. If a firm chooses not to invest in R&D, the firm-specific risk becomes such that it will have to leave the market after one period of (profitable) production: ε becomes one. Using a subscript RD to distinguish firms that invest in R&D and a subscript H to distinguish firms that do not invest, we get:

$$\delta_{RD} = \zeta \quad \text{and} \quad \delta_H = 1 \quad (3)$$

Henceforth we will refer to firms that invest as innovator firms and to firms that enter the market to make a one-time profit as hype-followers.¹³

¹¹ This binary choice feature of our set-up is a direct consequence of how we specified $\varepsilon(f_{RD})$ and the fact that there is a lower bound of zero to firm specific risk. A binary choice would also follow for the quadratic specification $\varepsilon(f_{RD}) = 1 - (\frac{1}{b})f_{RD}^a/\bar{f}_{RD}$ for all $b > 0$ and $a > 1$. When $a < 1$, however, different outcomes may occur. When $0 < a < 1$, the specification would yield $0 < f_{RD} < \bar{f}_{RD}$ as optimal outcome while for $a < 0$ $f_{RD} = 0$ results.

¹² This common feature of models with heterogeneous firms and R&D is related to the fact that in heterogeneous firms models with monopolistic competition productivity and profitability are positively related.

¹³ Our terminology is based on Golotto and Kim (2003) arguing that there are two types of dotcoms: hype followers who spend their money on marketing and do not have long-term viability; and firms who invest a lot in R&D and have a long-term vision. Furthermore, our terminology is related to the Hype Cycle concept developed by Gartner, Inc. in 1995, see Fenn & Raskino (2008) and Järvenpää & Mäkinen (2008). It is also related to Malerba and Orsenigo (2001 & 2002), who model the history of the pharmaceutical industry. They include two types of firms: imitators and innovators. Innovators try to research new drugs, while imitators only imitate the drugs already researched by others and do not execute any research themselves.

Applying (3) and the firm's investment decision (2), the value function of the firm can be rewritten to

$$v(\varphi) = \max \left\{ 0, \pi(\varphi), \frac{\pi(\varphi) - \bar{f}_{RD}}{\zeta} \right\} \quad (4)$$

Equation (4) defines two productivity cut-off points for market entry. The first cut-off point is the familiar Zero Profit Cut-off point φ^* which denotes the minimum productivity level for firms to have positive profits: $\varphi^* = \inf(\varphi | v(\varphi) > 0)$. This holds for all firms, irrespective of their type: since $\zeta > 0$ and $\bar{f}_{RD} > 0$, any firm with productivity $\varphi \leq \varphi^*$ will have negative profits when investing. The second cut-off point is the aforementioned productivity level φ_{RD}^* below which a firm finds it not profitable to invest in R&D. This threshold marks the difference between becoming a hype-follower or an innovator firm. As hype-follower a firm receives income for one period $\pi(\varphi)$; as an innovator firm it receives profits until it is forced to exit by a systemic shock, yielding a firm value of $(\pi(\varphi) - \bar{f}_{RD})/\zeta$. Consequently, and consistent with eq. (2), a firm would only want to become an R&D firm if $\pi(\varphi) \geq \bar{f}_{RD}/(1 - \zeta)$. In order for the investment to be profitable, the discounted profit in each period has to outweigh the fixed costs associated with staying in the market. Furthermore, as argued, $\varphi_{RD}^* > \varphi^*$.

The profits a firm derives from its operations are determined as in the Melitz model and we only repeat those equations that are useful for further reference. The demand side of the model is governed by a familiar Dixit-Stiglitz type of utility function with a constant elasticity of substitution $\sigma > 1$. Utility maximization defines demand q and revenue r for a firm producing variety ω :

$$q(\omega) = Q \left[\frac{p(\omega)}{P} \right]^{-\sigma} \quad (5)$$

$$r(\omega) = R \left[\frac{p(\omega)}{P} \right]^{1-\sigma} \quad (6)$$

where p denotes price and $R = PQ$ is aggregate expenditure with P denoting the aggregate price level

$$P = \left[\int_{\omega \in \Omega} p(\omega)^{1-\sigma} d\omega \right]^{\frac{1}{1-\sigma}} \quad (7)$$

These equations apply to hype followers and innovator firms alike. Furthermore, all firms produce their varieties using labor only, of which the total supply is completely inelastic and fixed at L . Production features increasing returns to scale, modeled by a fixed overhead cost $f > 0$, along with a marginal costs that depends on a firms productivity level $\varphi > 0$:

$$l = f + q/\varphi \quad (8)$$

Assuming a sufficiently large number of firms in industry, each firm faces a demand curve with constant elasticity σ . Profits for an individual variety (omitting indices) can therefore be written as:

$$\pi(\varphi) = q(\varphi)p(\varphi) - w \left(\frac{q}{\varphi} + f \right) \quad (9)$$

with w denoting the wage rate, which we normalize to 1. We will refer to this profit level as operational profits, as it only takes into account the fixed and variable costs associated with a firm's production activities. For hype followers these operational profits correspond to the overall profit level, but innovator firms' overall profits would also have to include the R&D investment costs.

Standard profit maximization gives a firm's optimal price and quantity:

$$p(\varphi) = \frac{1}{\rho\varphi} \quad (10)$$

$$q(\varphi) = RP^{\sigma-1}[\rho\varphi]^\sigma \quad (11)$$

with $0 < \rho \equiv \left(\frac{\sigma-1}{\sigma} \right) < 1$ as the familiar mark-up over marginal cost. This implies operational profits of

$$\pi(\varphi) = \frac{r(\varphi)}{\sigma} - f = \frac{R(P\rho\varphi)^{\sigma-1}}{\sigma} - f \quad (12)$$

As is well known, these formulas imply that a more productive firm will sell more products, charge a lower price and has higher revenues. The equations also allow for writing profits and revenue of firms relative to one another:

$$\frac{r(\varphi_1)}{r(\varphi_2)} = \left(\frac{\varphi_1}{\varphi_2}\right)^{\sigma-1} \quad (13)$$

$$\pi(\varphi_2) = \left(\frac{\varphi_2}{\varphi_1}\right)^{\sigma-1} \frac{r(\varphi_1)}{\sigma} - f \quad (14)$$

Consequently, we can write the cut-off point for R&D firms φ_{RD}^* relative to the cut off point for profitable entry φ^* :

$$(1 - \zeta)f \left(\left(\frac{\varphi_{RD}^*}{\varphi^*} \right)^{\sigma-1} - 1 \right) = \bar{f}_{RD} \quad (15)$$

where we applied (14), using that $\pi(\varphi^*) = 0$ and acknowledging that profitable entry as an R&D firm also implies payment of fixed R&D costs.

In equilibrium, there will be a large number of firms, each producing a distinct product variety. We can aggregate the previous results and solve the various variables for the entire economy. To do so, we must introduce two other variables: M , which denotes the total mass of producers in an economy and hence the number of varieties produced in an economy, and M_e , the mass of entering firms each period.

M firms have productivity levels drawn from an ex ante probability density function $g(\varphi)$ and associated cumulative distribution function $G(\varphi)$. It follows that the ex ante probability of successful entry is $p_e = 1 - G(\varphi^*)$ and that of entering as an innovator $p_{rd} = 1 - G(\varphi_{RD}^*)$. Taking into account that the distribution changes due to the exit of firms, the ex post probability distributions of productivities become:

$$\mu(\varphi_H) = \frac{g(\varphi)}{G(\varphi_{RD}^*) - G(\varphi^*)} \quad \text{and} \quad \mu(\varphi_{RD}) = \frac{g(\varphi)}{1 - G(\varphi_{RD}^*)} \quad (16)$$

The average productivity level in the market for each type of firm becomes:

$$\tilde{\varphi}_H(\varphi^*, \varphi_{RD}^*) = \left(\frac{1}{G(\varphi_{RD}^*) - G(\varphi^*)} \int_{\varphi^*}^{\varphi_{RD}^*} \varphi^{\sigma-1} g(\varphi) d\varphi \right)^{\frac{1}{\sigma-1}} \quad (17)$$

$$\tilde{\varphi}_{RD}(\varphi_{RD}^*) = \left(\frac{1}{1 - G(\varphi_{RD}^*)} \int_{\varphi_{RD}^*}^{\infty} \varphi^{\sigma-1} g(\varphi) d\varphi \right)^{\frac{1}{\sigma-1}} \quad (18)$$

implying average overall profits of

$$\bar{\pi}_H = \pi(\tilde{\varphi}_H) = \left(\left(\frac{\tilde{\varphi}_H}{\varphi^*} \right)^{\sigma-1} - 1 \right) f \quad (19)$$

$$\bar{\pi}_{RD} = \pi(\tilde{\varphi}_{RD}) - f_{RD} = \left(\left(\frac{\tilde{\varphi}_{RD}}{\varphi^*} \right)^{\sigma-1} - 1 \right) f - \bar{f}_{RD} \quad (20)$$

where we have applied (12) and $\pi(\varphi^*) = 0$. Note that the average profit levels refer to the overall profit levels of firms as it includes R&D costs.

Equations (19) and (20) establish an equilibrium relationship between average profits and the cut-off productivity level of profitable entry φ^* . As in the original Melitz model, these are downward sloping curves. Furthermore, in equilibrium, the expected value of entering the market must be zero:

$$\begin{aligned} v_e &= (p_e - p_{rd})\bar{v}_H + p_{rd}\bar{v}_{RD} \\ &= (G(\varphi_{RD}^*) - G(\varphi^*))\bar{\pi}_H + (1 - G(\varphi_{RD}^*))\frac{1}{\zeta}\bar{\pi}_{RD} - f_e = 0 \end{aligned} \quad (21)$$

where $f_e > 0$ denote the fixed entry cost that each firm will have to incur when entering the market. We require $\bar{f}_{RD} < f_e$ for otherwise the costs of (renewed) entry would be lower than the costs of reducing the firms-specific risk of exit. As in Melitz, we could use this formula to

calculate average profit levels and to generate a free entry condition. The average profit functions become:¹⁴

$$\bar{\pi}_H = \frac{f_e - (1 - G(\varphi_{RD}^*)) \frac{1}{\zeta} \left(f \left(\left(\frac{\tilde{\varphi}_{RD}}{\tilde{\varphi}_H} \right)^{\sigma-1} - 1 \right) \left(\frac{\tilde{\varphi}_H}{\varphi^*} \right)^{\sigma-1} \right) - \bar{f}_{RD}}{(G(\varphi_{RD}^*) - G(\varphi^*)) + (1 - G(\varphi_{RD}^*)) \frac{1}{\zeta}} \quad (22)$$

$$\bar{\pi}_{RD} = \frac{f_e + (G(\varphi_{RD}^*) - G(\varphi^*)) \left(f \left(\left(\frac{\tilde{\varphi}_{RD}}{\tilde{\varphi}_H} \right)^{\sigma-1} - 1 \right) \left(\frac{\tilde{\varphi}_H}{\varphi^*} \right)^{\sigma-1} \right) - \bar{f}_{RD}}{(G(\varphi_{RD}^*) - G(\varphi^*)) + (1 - G(\varphi_{RD}^*)) \frac{1}{\zeta}}. \quad (23)$$

To get the equilibrium values for both cut-off points φ^* and φ_{RD}^* , we set (21) to zero and use (19)-(20) to obtain $(G(\varphi_{RD}^*) - G(\varphi^*)) \left(\left(\frac{\tilde{\varphi}_H}{\varphi^*} \right)^{\sigma-1} - 1 \right) f + (1 - G(\varphi_{RD}^*)) \frac{1}{\zeta} \left[\left(\left(\frac{\tilde{\varphi}_{RD}}{\varphi^*} \right)^{\sigma-1} - 1 \right) f - \bar{f}_{RD} \right] = f_e$. Applying the expressions for average productivity (17) and (18) this reduces to

$$\begin{aligned} & \left(\left(\int_{\varphi^*}^{\varphi_{RD}^*} \left(\frac{\varphi}{\varphi^*} \right)^{\sigma-1} g(\varphi) d\varphi \right) - (G(\varphi_{RD}^*) - G(\varphi^*)) \right) f \\ & + \frac{1}{\zeta} \left(\left(\int_{\varphi_{RD}^*}^{\infty} \left(\frac{\varphi}{\varphi^*} \right)^{\sigma-1} g(\varphi) d\varphi \right) - (1 - G(\varphi_{RD}^*)) \right) f \\ & = (1 - G(\varphi_{RD}^*)) \frac{1}{\zeta} \bar{f}_{RD} + f_e \end{aligned}$$

¹⁴ These unwieldy functions could be used to prove existence of equilibrium. Their derivation follow from rewriting (19) to $\bar{\pi}_{RD} = \left(\left(\frac{\tilde{\varphi}_{RD}}{\tilde{\varphi}_H} \right)^{\sigma-1} \left(\frac{\tilde{\varphi}_H}{\varphi^*} \right)^{\sigma-1} - 1 \right) f - \bar{f}_{RD}$, so that $\bar{\pi}_{RD} = \bar{\pi}_H + f \left(\left(\frac{\tilde{\varphi}_{RD}}{\tilde{\varphi}_H} \right)^{\sigma-1} - 1 \right) \left(\frac{\tilde{\varphi}_H}{\varphi^*} \right)^{\sigma-1} - \bar{f}_{RD}$. Substituting this in the expression for v_e , and rearranging, gives the expression for $\bar{\pi}_H$ in (22). Equation (23) is obtained by substituting $\bar{\pi}_H$ in the expression for $\bar{\pi}_{RD}$ above.

This free-entry condition relates the cut-off points to exogenous variables only. Together with equation (14) this could in principle be solved to obtain equilibrium values for the two cut-off points in our analysis.

What is left is to determine the equilibrium mass of entrants into the industry each period, M_e . Each period, ζ innovators and all hype followers leave the market. To have constant levels of all aggregate variables over time (steady-state equilibrium), the mass of exiting firms $\zeta M_{RD} + M_H$ needs to be equal to the mass of entering firms M_e . Taking into account the probability of successful entry, this implies

$$p_e M_e = \zeta M_{RD} + M_H \quad (24)$$

Furthermore, the division across types of firms must remain constant in steady state. Because the total mass of firms is variable, deriving the mass of innovators does not amount to deriving their share. Defining $P_{rd} \equiv \frac{p_{rd}}{p_e}$ as the probability of becoming an innovator firm after successful entry, the number of innovator firms is¹⁵

$$M_{RD} = \left(\frac{P_{rd}}{P_{rd} + (1 - P_{rd})\zeta} \right) M \quad \text{and} \quad M_H = M - M_{RD} \quad (25)$$

If this condition is satisfied, the shares of the two types of firms are stable over time. Each combination of P_{rd} and ζ gives rise to a single unique equilibrium M_{RD} .¹⁶

The main implication of Equation (25) is that it shows that the reduced chance of death for the more productive R&D firms begets a selection effect: in the equilibrium distribution of firms, the high-productivity R&D firms are better represented compared to a situation where the distribution of firms equals the distribution of random entrants. To see this, we remove ε from the model, reducing the model to the original Melitz model with no exit divide between

¹⁵ Equation (25) has been derived by applying that M_{RD} and M remain constant over time. Using that $M_{RD} + M_H = M$, each period $\delta_{RD} M_{RD} + \delta_H (M - M_{RD})$ firms leave the market, of which a percentage P_{RD} re-enter as innovative firms. Equating this to the $\delta_{RD} M_{RD}$ innovator firms that leave leads to $M_{RD} = \delta_H P_{RD} / (\delta_H P_{RD} + (1 - P_{RD})\delta_{RD})$. Implementing (3) gives (25).

¹⁶ Except for the corner case $P_{rd} = \zeta = 0$. However, with finite innovation costs $P_{rd} \neq 0$ so that we can safely ignore this possibility.

innovators and hype followers. Then, in each period ζ firms, chosen randomly, exit the market, and $\frac{\zeta}{p_e}$ firms enter the market to take their place. Because both exit and entry are random, the same average distribution persists and there is an equilibrium with $M_{RD} = P_{rd}M$ and $M_H = (1 - P_{rd})M$. Now let us suppose that this is also the average distribution in the model as we specified it before ($\varepsilon_{RD} = 0, \varepsilon_H = 1$). At the end of the first period, $\zeta M_{RD} + M_H$ firms leave the market. The remaining firms then have a different distribution than we had in the model with random exiting. Since all hype followers leave the market at the end of each period, the remaining firms have a distribution of M_{RD} . By contrast, new entrants in the next period have the ‘normal’ average distribution $(1 - P_{rd})M_H + P_{rd}M_{RD}$. Hence, the distribution changes over time, converging to (25) in equilibrium. The overrepresentation of R&D firms in comparison to the fully random exit situation becomes clear when noting that $P_{rd} + (1 - P_{rd})\zeta < 1$ in (25).

We are now in the position to calculate average productivity, our main variable of interest in this paper. Given the equilibrium distribution of firms, we can calculate average productivity as:

$$\tilde{\varphi} = \left(1 - \frac{P_{rd}}{P_{rd} + (1 - P_{rd})\zeta}\right) \tilde{\varphi}_H + \frac{P_{rd}}{P_{rd} + (1 - P_{rd})\zeta} \tilde{\varphi}_{RD} \quad (26)$$

Equation (26) describes $\tilde{\varphi}$ for all $1 \geq P_{rd} > 0$ and for all $1 \geq \zeta > 0$.¹⁷ If $\zeta = 1$, we are again in a situation where death does not discriminate between the two types of firms and the selection effect disappears, implying we have the same results as in the Melitz model. If $P_{rd} = 1$, there will never be hype followers and we do not find any different results either. However, our parameters always satisfy $1 > P_{rd} > 0$ and $1 > \zeta > 0$ which means that the denominator is always smaller than one and that there is always a selection effect.

The conclusion of our benchmark model is therefore that, through a selection effect, average productivity increases when firms must engage in R&D to remain in the market. In the Melitz model only average productivity matters for the value of aggregate variables (Melitz, 2003: 1700). If such would be the case in our model as well, the selection effect would have a

¹⁷If P_{rd} and ζ are both zero the starting distribution of firms would remain in place, a possibility we will exclude.

similar impact on aggregate variables as a general increase in productivity for all firms. However, in our set-up also the distribution of firm productivities affects aggregate variables through its impact on the total mass of firms.

To see this, we consider the expressions for the aggregate variables, which remain the same as in Melitz (2003):

$$P = M^{\frac{1}{1-\sigma}} p(\tilde{\varphi}) \quad (27)$$

$$Q = M^{1-\rho} q(\tilde{\varphi}) \quad (28)$$

$$R = Mr(\tilde{\varphi}) \quad (29)$$

$$\Pi = M\pi(\tilde{\varphi}) \quad (30)$$

Following the same procedure as in Melitz (2003: 1705), we derive

$$M = \frac{L}{\sigma(\pi(\tilde{\varphi}) + f) + \bar{f}_{RD} \frac{P_{rd}}{P_{rd} + (1-P_{rd})\zeta}} \quad (31)$$

The denominator of (31) comprises average operational profits, excluding what is paid to R&D labor. Payments to R&D labor have been accounted for in the derivation of (31) and explain the second term in the denominator.¹⁸ It is clear from the formula that an increase in the probability of entering as an innovator firm decreases the total mass of firms in the industry. Since the introduction of R&D implies an increase in that probability from zero to some positive number, the addition of R&D costs begets a lower total mass of firms¹⁹.

¹⁸ The free entry of firms implies that labor involved with the entry of firms must receive overall profits, which in our notation amounts to Π . Production labor (excluding labor involved in R&D) receives $R - (\Pi + \bar{f}_{RD} M_{RD})$ and aggregate payments to R&D labor are $\bar{f}_{RD} M_{RD}$. Hence, the total wage bill becomes $L = R + \bar{f}_{RD} M_{RD}$. Using (25) and (29), noting that $r(\tilde{\varphi}) = \sigma(\pi(\tilde{\varphi}) + f)$, gives (31).

¹⁹ Unlike the total mass of firms, the mass of innovator firms is always increasing in P_{rd} . There is no point at which the effect on the total mass of firms starts to outweigh the effect on the probability of becoming an innovator.

Including firm-specific R&D thus affects aggregate variables through an increase in average productivity and through a decline in the mass of firms. We will refer to the former effect as the ‘class effect’ and to the latter effect as the ‘mass effect’.

The mass effect has an adverse effect on aggregate variables: prices increase; quantities, revenues and profits decrease when P_{rd} increases. The class effect is less clear, however:

$$\frac{\partial \tilde{\varphi}}{\partial P_{rd}} = \left(\frac{\zeta}{(P_{rd} + (1 - P_{rd})\zeta)^2} \right) (\tilde{\varphi}_{RD} - \tilde{\varphi}_H) + \frac{M_H}{M} \frac{\partial \tilde{\varphi}_H}{\partial P_{rd}} + \frac{M_{RD}}{M} \frac{\partial \tilde{\varphi}_{RD}}{\partial P_{rd}}.$$

The first term on the right-hand-side of this equation indicates a composition effect and is positive due to $\tilde{\varphi}_{RD} > \tilde{\varphi}_H$. The second and third terms on the right-hand-side are both negative: when the chance of becoming an innovator firm increases, the most productive hype-followers will become innovator firms, lowering the average productivity of both the hype-followers as well as the innovator firms. The overall effect on average productivity is positive when P_{rd} is low and small or even negative when P_{rd} is high. For low P_{rd} the composition effect dominates as its weight approaches $1/\zeta \gg 1$. When P_{rd} is high, by contrast, the composition effect’s weight approaches $\zeta \ll 1$, implying an overall effect close to zero or even negative.²⁰

Taking the mass effect and class effect together, this implies that average productivity is likely to go up in an economy in which a minority of firms becomes innovator firms. If there would be only innovator firms, there would be no class effect but there would be a significant negative mass effect due to the R&D spending that all firms are doing.

The question is then, how does P_{rd} change? P_{rd} is a function of both φ^* and φ_{RD}^* , $P_{rd} = (1 - G(\varphi_{RD}^*)) / (1 - G(\varphi^*))$. From (15) we know that φ^* and φ_{RD}^* are in a fixed relation. Applying $d\varphi^* / \varphi^* = d\varphi_{RD}^* / \varphi_{RD}^*$, we get

$$\frac{dP_{RD}}{P_{RD}} = - \left[\frac{G'(\varphi_{RD}^*)}{1 - G(\varphi_{RD}^*)} \frac{\varphi_{RD}^*}{\varphi^*} - \frac{G'(\varphi^*)}{1 - G(\varphi^*)} \right] d\varphi^*.$$

²⁰ The division of firms in hype-followers and innovator firm does not matter as they both pertain to negative terms of equal magnitude.

Hence, $\frac{dP_{RD}}{d\varphi^*} < 0$ if $G'(\varphi_{RD}^*) \geq G'(\varphi^*)$. When $G'(\varphi_{RD}^*) < G'(\varphi^*)$, the sign is unclear but even then is likely that $\frac{dP_{RD}}{d\varphi^*} < 0$. For $1 - G(\varphi_{RD}^*) < 1 - G(\varphi^*)$ and $\frac{\varphi_{RD}^*}{\varphi^*} > 1$. Hence, a reduction in φ^* goes in tandem with an increase in P_{RD} . This is a logical consequence of $d\varphi^*/\varphi^* = d\varphi_{RD}^*/\varphi_{RD}^*$.

The equation implies that the relation between P_{RD} and φ^* is not independent of the particular distribution of productivities. A linear distribution (e.g. a uniform distribution) leads to different outcomes than, for instance, a Pareto-distribution. While for a linear distribution we derive $G'(\varphi_{RD}^*) = G'(\varphi^*)$ and hence $\frac{dP_{RD}}{d\varphi^*} < 0$, a Pareto-distribution featuring $G(\varphi) = 1 - \left(\frac{b_m}{\varphi}\right)^\alpha$ and $G'(\varphi) = \alpha \frac{b_m^\alpha}{\varphi^{\alpha+1}}$ leads to $\frac{dP_{RD}}{d\varphi^*} = 0$. How the productivities are distributed is therefore of crucial importance for understanding the effect of increased innovative activity on average productivity.

3. Effects of clustering

The previous section has established the groundwork for an analysis of clustering on average productivity. The consequence of clustering is that knowledge spillovers may occur. That is, firms not engaging in R&D themselves may nevertheless be able to reduce their chance of exit due to spillovers. The importance of geographical proximity for knowledge spillovers has been well-established, see Audretsch & Feldman (2004) for an overview. We will consider two alternative ways of modeling spillovers. As a first alternative we consider spillovers as the ability of non-R&D firms to gain part of the technology researched by R&D firms at no cost. This perspective sees knowledge spillovers as the direct imitation of R&D practices. We do assume however that these knowledge spillovers by direct imitation are imperfect, for instance because non-R&D firms lack the ability to fully absorb the knowledge that is inadvertently transmitted. The second alternative of modeling spillovers is that we take spillovers to imply that it lowers the costs of doing R&D. This perspective bears a relation with the idea that knowledge creation entails a positive externality, making R&D investments cheaper for all firms in the cluster.

As alluded to in the introduction, we keep the geography in our model exogenous and simply assume that the model we developed pertains to some geographic cluster of industrial activity. The exit and entry of firms could therefore also be seen as firms leaving or entering the cluster.²¹ We will also assume that within the cluster there are no distance decay effects of knowledge, so that the gains from R&D accrue to all other firms in the industry.

Spillovers by direct imitation

Spillovers imply that the chance of market exit does not only rely on one's own R&D investments, but also on the R&D investments of others. The systemic risk ζ remains the same, but firms are now able to imitate some of the technology researched by others at no cost. This reduces the firm-specific risk also for non-investing firms. Using a parameter $0 \leq \theta < 1$ to denote the extent of knowledge spillovers, this implies that the modeling of firm-specific risks becomes:

²¹ Demand is geographically concentrated in our set-up, which would be one reason for firms to enter the cluster.

$$\varepsilon = (1 - \theta) \frac{\bar{f}_{RD} - f_{RD}}{\bar{f}_{RD}} \quad (0 < f_{RD} < \bar{f}_{RD}).$$

We restrict θ to a value below 1 is required as with perfect spillovers no single firm would have an incentive to invest in R&D. For $\theta = 0$ we have our benchmark case with no spillovers. Our specification implies that only hype-followers will benefit from knowledge spillovers. This is a logical consequence of the R&D investment decision: a firm either invests fully, reducing the firms-specific market risk to zero, or it does not invest at all, becoming a hype-follower. By restricting θ to a value smaller than one, we assume that hype-followers cannot fully benefit from imitation. This reflects a notion of absorptive capacity - in order to benefit from R&D spillovers one should engage in R&D oneself – but also the possibility that the technology available does not match the firm’s requirements perfectly.²² We also assume that spillovers are independent of the number of R&D firms. Direct imitation implies that a hype follower imitates with the specific goal to stay in the market longer, improving its product in line with the incremental product innovations of R&D firms. Since all firms produce a slightly different product, it thus makes sense to only copy from that firm that is closest to what the hype follower produces itself. Furthermore, also for reason of (lack of) absorptive capacity hype followers will stand a better chance imitating from an R&D firm that produces a variety close to that of the hype follower.²³

Accordingly, with direct imitation the consolidated chances of exit become:

$$\delta_H = 1 - (1 - \zeta)\theta \quad (32)$$

$$\delta_{RD} = \zeta \quad (33)$$

changing the value function into:

²² For R&D firms these issues are irrelevant since also with knowledge spillovers the optimal investment decision implies $f_{RD} = \bar{f}_{RD}$. The relevant first-order-condition of optimal investment is $dv(\varphi)/df_{RD} = [(1 - \zeta)(1 - \theta)(\pi(\varphi) - \bar{f}_{RD})/\bar{f}_{RD} - \zeta]/[\zeta + \varepsilon(f_{RD}) - \zeta\varepsilon(f_{RD})]^2 = 0$.

²³ While we acknowledge that a greater number of firms in the cluster could be helpful in this respect – after all a more crowded product space would lower average distance between product specifications – it would also make it more difficult for a hype follower to pick the right R&D firm to imitate from.

$$v(\varphi) = \max \left\{ 0, \frac{1}{1 - (1 - \zeta)\theta} \pi(\varphi), \frac{\pi(\varphi) - \bar{f}_{RD}}{\zeta} \right\} \quad (34)$$

Eq. (34) defines a new φ^* as well as a new φ_{RD}^* . Again it holds that any firm with productivity $\varphi \leq \varphi^*$ will also have negative profits when investing.²⁴ Note that spillovers do not affect φ^* directly. Ceteris paribus aggregate variables, the profit level of a firm that does not invest is not affected by spillovers. This is different for the cutoff point for R&D investment, which is now implicit in:

$$\pi(\varphi) \geq \left[\frac{1 - (1 - \zeta)\theta}{(1 - \theta)} \right] \bar{f}_{RD} / (1 - \zeta)$$

The requirement on φ is consistent with the FOC on optimal investment.²⁵ Applying (14) and using $\pi(\varphi^*) = 0$, we get

$$\left(\frac{1 - \theta}{1 - (1 - \zeta)\theta} \right) (1 - \zeta) f \left(\left(\frac{\varphi_{RD}^*}{\varphi^*} \right)^{\sigma-1} - 1 \right) = \bar{f}_{RD} \quad (35)$$

For $\theta = 0$ (no spillovers), this yields the same φ_{RD}^* as in the benchmark model. For $\theta = 1$ (perfect spillovers), no productivity would be high enough to render R&D investment profitable. For any value $0 < \theta < 1$, the required φ for profitable R&D investment exceeds that of the benchmark model: the first term in (35) is smaller than was the case in the benchmark configuration, cf. (15). Hence, spillovers by imitation imply that the cutoff point for R&D investment, φ_{RD}^* , increases, ceteris paribus the aggregate variables. Because of knowledge spillovers, there is less incentive to invest compared to the benchmark model. As θ increases and knowledge spillovers become stronger, this effect on investment becomes stronger as well. Because part of the technology is released to all other firms, firms have less incentive to do research themselves.

In this paper we are interested in how spillovers affect the cluster's average productivity. Average cluster productivity can be written as

²⁴ The relevant comparison is $\pi(\varphi^*) / (1 - (1 - \zeta)\theta) - (\pi(\varphi^*) - \bar{f}_{RD}) / \zeta \leq 0$, which holds true for the designated values of ζ, θ and \bar{f}_{RD} .

²⁵ The FOC for optimal investment requires $\pi(\varphi) \geq [(1 - \zeta) + \zeta / (1 - \theta)] \bar{f}_{RD} / (1 - \zeta)$ for a firm to invest.

$$\tilde{\varphi} = \left(1 - \frac{M_{rd}}{M}\right) \tilde{\varphi}_H + \frac{M_{rd}}{M} \tilde{\varphi}_{RD}$$

with $\tilde{\varphi}_H$ and $\tilde{\varphi}_{RD}$ as defined in (17) and (18) respectively, and where M_{rd} and M are given by²⁶

$$M_{RD} = \frac{(1 - (1 - \zeta)\theta)P_{rd}}{(1 - P_{rd})\zeta + (1 - (1 - \zeta)\theta)P_{rd}} M,$$

$$M = \frac{L}{\sigma(\pi(\tilde{\varphi}) + f) + \bar{f}_{RD} \frac{(1 - (1 - \zeta)\theta)P_{rd}}{(1 - P_{rd})\zeta + (1 - (1 - \zeta)\theta)P_{rd}}}.$$

For understanding the impact on average cluster productivity we need to know how the average productivity of each of both firm categories changes, which in turn depend on φ^* and φ_{RD}^* changes. Since also with spillovers there is a direct and positive relation between φ^* and φ_{RD}^* (see further below), ultimately it suffices to determine the impact of spillovers on φ^* . Furthermore, we need insight in how spillovers affect the relative incidence of R&D firms in the cluster, M_{rd}/M . As we will see also here it is important to know the impact on φ^* .

We therefore first turn to how the inclusion of θ affects the cut-off point for profitable production φ^* . By (34) we know that there is no direct effect of θ on φ^* , but clearly there is one through the aggregate variables. This is governed by the free-entry condition. Using the same procedure as we did for the benchmark model, the free-entry condition with spillovers by direct imitation becomes:

²⁶ In deriving M_{rd} and M we used as before that in equilibrium the mass of entering and exiting firms needs to have the same productivity distribution.

$$\begin{aligned}
& \frac{\left(\int_{\varphi^*}^{\varphi_{RD}^*} \left(\frac{\varphi}{\varphi^*} \right)^{\sigma-1} g(\varphi) d\varphi \right) - (G(\varphi_{RD}^*) - G(\varphi^*))}{[1 - (1 - \zeta)\theta]} f \\
& + \frac{1}{\zeta} \left(\left(\int_{\varphi_{RD}^*}^{\infty} \left(\frac{\varphi}{\varphi^*} \right)^{\sigma-1} g(\varphi) d\varphi \right) \right. \\
& \left. - (1 - G(\varphi_{RD}^*)) \right) f = (1 - G(\varphi_{RD}^*)) \frac{1}{\zeta} \bar{f}_{RD} + f_e.
\end{aligned} \tag{36}$$

For $\theta = 0$ the equation reduces to the free-entry condition of the benchmark model. The impact of θ on φ^* is effectively determined by total differentiation of (36), recognizing from (35) that on φ^* and φ_{RD}^* stand in a fixed relation:

$$\hat{\varphi}_{RD}^* = \hat{\varphi}^* + \left(\frac{\zeta}{(1 - \theta)[1 - (1 - \zeta)\theta]} \right) \left(1 - \left(\frac{\varphi^*}{\varphi_{RD}^*} \right)^{\sigma-1} \right) d\theta \tag{37}$$

with a hat “^” denoting a proportional change, for instance $\hat{\varphi}_{RD}^* = d\varphi_{RD}^*/\varphi_{RD}^*$. In the presence of spillovers the proportional change of the threshold for profitable entry as an R&D firm exceeds that of the threshold for general entry: $\hat{\varphi}_{RD}^* > \hat{\varphi}^*$. Using this in the total differentiation of (36) and applying the Pareto-distribution yields:

$$\hat{\varphi}^* = Z d\theta$$

where Z is shorthand for an expression that is given in the appendix. As we show, Z is positive when both of the following two conditions are fulfilled:

$$(1 - \zeta C(\theta)) \left(\left(\frac{\varphi_{RD}^*}{\varphi^*} \right)^{\sigma-1} - 1 \right) > \bar{f}_{RD}/f$$

and

$$\left[\left(\frac{\alpha}{\sigma - \alpha - 1} \left(\frac{\varphi_{RD}^*}{\varphi^*} \right)^{\sigma-1} + 1 \right) \varphi_{RD}^{*\alpha} - \frac{\sigma - 1}{\sigma - \alpha - 1} \varphi^{*\alpha} \right] < 0.$$

The conditions indicate that higher relative productivity of R&D firms compared to hype firms $\left(\frac{\varphi_{RD}^*}{\varphi^*}\right)$ due to a ceteris paribus increase in φ_{RD}^* will make both conditions more likely to be satisfied. This follows from the model because an increase in spillovers implies that firms will move from being hype followers to being innovators. The higher the productivity difference between the two types of firms, the greater the gain for average productivity when the shift occurs. For φ^* , however, things are not so clear. A decrease in φ^* will make the first condition more likely to be satisfied, but the second condition will be less likely to be satisfied.. This is likely because of distribution effects: since it is hype followers that shift over to being sustainable innovators, lower average productivity for hype followers may imply a relatively bigger gain from becoming an innovator, but also that there will be a bigger decrease in average productivity of sustainable innovators because of the hype followers' switch to innovatorhood.

In that case, when spillovers by direct imitation are possible ($d\theta > 0$), the threshold for profitable entry in the market increases. Furthermore, φ_{RD}^* increases by more, also percentage wise. Spillovers imply that fewer hype-followers leave the market each period, implying all new entrants should be more productive than before to be able to overcome the more intense competition.

The effects of the changes in φ^* and φ_{RD}^* on average cluster productivity are not clear though. We can write the change in cluster productivity as:

$$d\tilde{\varphi} = (\tilde{\varphi}_{RD} - \tilde{\varphi}_H)d(M_{rd}/M) + \left(1 - \frac{M_{rd}}{M}\right) d\tilde{\varphi}_H + \frac{M_{rd}}{M} d\tilde{\varphi}_{RD}.$$

The change in M_{rd}/M is

$$d\left(\frac{M_{RD}}{M}\right) = \frac{1}{M} \frac{\partial P_{rd}}{\partial \theta} \left\{ \frac{\partial M_{RD}}{\partial P_{rd}} - \frac{M_{RD}}{M} \frac{\partial M}{\partial P_{rd}} \right\} d\theta < 0$$

for $d\theta > 0$ and when the Pareto-distributio is applied (see appendix). Hence, spillovers imply that the relative incidence of R&D firms in the cluster diminishes, exerting a negative impact on overall cluster productivity since $\tilde{\varphi}_{RD} > \tilde{\varphi}_H$.

However, spillovers by direct imitation also imply that the average productivity of both groups of firms changes. Starting with the effect on average productivity of hype-followers, we note that there are two effects that happen to the distribution of productivities of hype followers as φ^* increases. One is a rightward shift along the Pareto distribution due to the proportional changes in φ^* and φ_{RD}^* . The second is a rightward dilation because φ_{RD}^* increases proportionally more than φ^* . Both effects become clear when looking at the two variables that together determine average productivity, see (17): the volume of productivities $\int_{\varphi^*}^{\varphi_{RD}^*} \varphi^{\sigma-1} g(\varphi) d\varphi$ and the cumulative density of all firms between φ^* and φ_{RD}^* , $G(\varphi_{RD}^*) - G(\varphi^*)$, which is the base over which the volume is spread. For the Pareto distribution, a rightward shift of φ^* and φ_{RD}^* will cause the base to decrease. The Pareto distribution flattens out at the right end, which means that a fixed interval of φ^* until φ_{RD}^* will have a higher cumulative density the lower φ^* is. Also when φ^* and φ_{RD}^* experience a fixed proportional increase rather than a fixed absolute increase, the negative shift-effect on the base remains. However, also the volume will become smaller, so that the rightward shift means a smaller volume divided over a smaller base. The same holds true for dilation. Here, φ_{RD}^* increases, ceteris paribus. This means that the base becomes broader, but there is also some additional volume added to the integral, so that the effect on average productivity is again ambiguous. The combination of shift and dilation, or the combination of changes in volume and base, in the end determine the overall result on average productivity of hype-followers.

In the appendix we determine these opposing effects mathematically, showing that a sufficient condition for $\frac{d\tilde{\varphi}_H}{d\theta} > 0$ is that $[\alpha(1 - (\varphi_{RD}^*/\varphi^*)^{1-\sigma+\alpha}) < (1 - \sigma + \alpha)(1 - (\varphi_{RD}^*/\varphi^*)^\alpha)]$. The balance between dilation and shift effects obscures deriving an unambiguous effect however. By contrast, for the average productivity of R&D firms we find an unambiguous positive effect of spillovers: $\frac{d\tilde{\varphi}_{RD}}{d\theta} > 0$. In this case, there is neither a dilation nor a shift effects, and there is only a left truncation effect due to $d\varphi_{RD}^* > 0$. With the firms of lowest productivity leaving the distribution of R&D firms, it stands to reason that the average productivity of R&D firms increases.

Knowing the effects on the relative incidence of R&D firms and average productivity if hype-followers and R&D firms, the effect on overall cluster productivity is ambiguous. With the

share of the more productive R&D firms diminishing, spillovers may very well imply that average cluster productivity decreases if the productivity difference between hype followers and sustainable innovators is not large enough or hype followers are not productive enough to begin with.

Spillovers by indirect imitation

We now assume that spillovers imply that firms are not able to imitate each other's technologies directly, but that they can imitate the research done by others. This implies that spillovers reduce the fixed investment costs of R&D and have no consequences for the consolidated chance of survival of firms: δ_H and δ_{RD} remain as in the benchmark model. Hence, using $0 < \theta < 1$ to parameterize the reduction in research costs, a firm's value function becomes:

$$v(\varphi) = \max \left\{ 0, \pi(\varphi), \frac{\pi(\varphi) - (1 - \theta)\bar{f}_{RD}}{\zeta} \right\} \quad (38)$$

where we also retained our assumption that $\varepsilon_H = 1$ and $\varepsilon_{RD} = 0$. As before, as long as the investment costs are non-prohibitive, there will always be at least one firm productive enough to invest in R&D.

The new zero cutoff point for R&D investment becomes:

$$(1 - \zeta)f \left(\left(\frac{\varphi_{RD}^*}{\varphi^*} \right)^{\sigma-1} - 1 \right) = (1 - \theta)\bar{f}_{RD} \quad (39)$$

Because $0 < \theta < 1$, we know that the right-hand side is smaller than before, implying a lower cutoff-point for R&D investment. Therefore, there will be more firms investing in R&D than was the case in the benchmark equilibrium. However, average productivity has declined compared to the benchmark model. In our benchmark model, R&D investments and the uneven chance of survival provide an additional element to the selection process and skew the distribution of firm productivities towards the higher echelons. The cheaper R&D becomes, the smaller this effect becomes. If R&D investments do not cost anything, the effect

completely disappears and our model is reduced to the original Melitz model. Note that the formula for the percentage of innovator firms (equation 25) does not change. Instead, the change in average productivity arises endogenously through a change in P_{rd} . Despite the fact that indirect imitation causes there to be more innovators and fewer hype followers, the average productivity within industry declines.

To formalize these issues we follow the same procedure as before. Average cluster productivity can be written as

$$\tilde{\varphi} = \left(1 - \frac{M_{rd}}{M}\right) \tilde{\varphi}_H + \frac{M_{rd}}{M} \tilde{\varphi}_{RD}$$

with $\tilde{\varphi}_H$ and $\tilde{\varphi}_{RD}$ as defined in (17) and (18) respectively, and where M_{rd} and M are given by²⁷

$$M_{RD} = \frac{P_{rd}}{(1 - P_{rd})\zeta + P_{rd}} M,$$

$$M = \frac{L}{\sigma(\pi(\tilde{\varphi}) + f) + (1 - \theta)\bar{f}_{RD} \frac{P_{rd}}{(1 - P_{rd})\zeta + P_{rd}}}.$$

As before it is useful to first determine how the inclusion of θ affects the cut-off point for profitable production φ^* . This is governed by the free-entry condition. Using the same procedure as we did for the benchmark model, the free-entry condition with spillovers by direct imitation becomes:

²⁷ Both expressions are as in the benchmark model, except for the inclusion of $1 - \theta$ in the denominator of M .

$$\begin{aligned}
& \left(\left(\int_{\varphi^*}^{\varphi_{RD}^*} \left(\frac{\varphi}{\varphi^*} \right)^{\sigma-1} g(\varphi) d\varphi \right) - (G(\varphi_{RD}^*) - G(\varphi^*)) \right) f \\
& + \frac{1}{\zeta} \left(\left(\int_{\varphi_{RD}^*}^{\infty} \left(\frac{\varphi}{\varphi^*} \right)^{\sigma-1} g(\varphi) d\varphi \right) \right. \\
& \left. - (1 - G(\varphi_{RD}^*)) \right) f \\
& = \frac{(1 - \theta)}{\zeta} (1 - G(\varphi_{RD}^*)) \bar{f}_{RD} + f_e.
\end{aligned} \tag{40}$$

The impact of θ on φ^* is effectively determined by total differentiation of (40), recognizing from (39) that on φ^* and φ_{RD}^* stand in the following fixed relation:

$$\hat{\varphi}_{RD}^* = \hat{\varphi}^* + \left(\frac{\bar{f}_{RD}/f}{(1 - \zeta)(\sigma - 1)} \right) d\theta \tag{41}$$

Again we see that in the presence of spillovers the proportional change of the threshold for profitable entry as an R&D firm exceeds that of the threshold for general entry: $\hat{\varphi}_{RD}^* > \hat{\varphi}^*$. Using this in the total differentiation of (36) and applying the Pareto-distribution yields:

$$\hat{\varphi}^* = Z' d\theta$$

where Z' is a shorthand for an expression that is given in the appendix. As we show in the appendix, Z' is negative when both of the following two conditions are fulfilled:

$$(1 - \zeta) \left(\left(\frac{\varphi_{RD}^*}{\varphi^*} \right)^{\sigma-1} - 1 \right) > (1 - \theta) \bar{f}_{RD}/f$$

and

$$\left[1 - \frac{\alpha}{(1 - \zeta)(\sigma - 1)} \left((1 - \zeta) \left(\left(\frac{\varphi_{RD}^*}{\varphi^*} \right)^{\sigma-1} - 1 \right) - \frac{(1 - \theta) \bar{f}_{RD}}{f} \right) \right] < 0.$$

The first condition is similar to the one we derived for the direct imitation case. However, its implications are different, for together with the second condition it would imply a negative relationship between $\hat{\varphi}^*$ and θ . Furthermore, we note that if it holds, $Z' < 0$ follows if and only if $\frac{\alpha}{(1-\zeta)(\sigma-1)}$ is sufficiently large. This is accommodated by a sufficiently low ζ and/or σ (noting that $\alpha > \sigma - 1$ by assumption).

The conditions indicate that changes in average productivity of either group of firms have an unambiguous effect: an increase in relative productivity of innovators, effected either through an increase in φ_{RD}^* or a decrease in φ^* will make both conditions more likely to be satisfied. As before, this stems from the effects of spillovers. Since spillovers induce firms to become hype followers instead of innovators, a higher difference between the two implies that spillovers are more likely to have a negative effect on average productivity of the cluster as a whole. In that case, when spillovers by direct imitation are possible ($d\theta > 0$), the threshold for profitable entry in the market decreases. Furthermore, φ_{RD}^* decreases by more, also percentage wise.

The effects of the changes in φ^* and φ_{RD}^* on average cluster productivity are not clear though. We write the change in cluster productivity as:

$$d\tilde{\varphi} = (\tilde{\varphi}_{RD} - \tilde{\varphi}_H)d(M_{rd}/M) + \left(1 - \frac{M_{rd}}{M}\right) d\tilde{\varphi}_H + \frac{M_{rd}}{M} d\tilde{\varphi}_{RD}.$$

The change in M_{rd}/M is

$$d\left(\frac{M_{RD}}{M}\right) = \frac{1}{M} \frac{\partial P_{rd}}{\partial \theta} \left\{ \frac{\partial M_{RD}}{\partial P_{rd}} - \frac{M_{RD}}{M} \frac{\partial M}{\partial P_{rd}} \right\} d\theta < 0$$

for $d\theta > 0$ and when the Pareto-distribution is applied (see appendix). Hence, also with indirect imitation spillovers imply that the relative incidence of R&D firms in the cluster diminishes, exerting a negative impact on overall cluster productivity since $\tilde{\varphi}_{RD} > \tilde{\varphi}_H$.

However, as before, spillovers also imply that the average productivity of both groups of firms changes. The analysis is qualitatively the same as in the direct imitation case, see the appendix. Hence, a sufficient condition for $\frac{d\tilde{\varphi}_H}{d\theta} > 0$ is that $[\alpha(1 - (\varphi_{RD}^*/\varphi^*)^{1-\sigma+\alpha}) <$

$(1 - \sigma + \alpha)(1 - (\varphi_{RD}^*/\varphi^*)^\alpha)$]. For the average productivity of R&D firms we find an unambiguous positive effect of spillovers: $\frac{d\tilde{\varphi}_{RD}}{d\theta} > 0$.

Knowing the effects on the relative incidence of R&D firms and average productivity if hype-followers and R&D firms, the effect on overall cluster productivity is ambiguous. With the share of the more productive R&D firms diminishing, spillovers imply that average cluster productivity decreases if there is a large difference in average productivity between innovators and hype followers.

Our way of modeling direct and indirect knowledge spillovers could also be linked to the theory on absorptive capacity, arguing that the two kinds of imitation are based on two different views of absorptive capacity. If we model direct imitation, there is no difference in absorptive capacity between firms, as all firms benefit equally from the knowledge spillover. If we model indirect imitation, the specific characteristics of a firm become important, as only the most productive of the hype followers will be able to assimilate the technology. This is in line with the results of Cohen and Levinthal (1989), who argued that firms need to invest in R&D to be able to assimilate other technologies.

4. Extensions

We now drop the assumption that θ is constant. Before, we argued that if one firm discovered a particular technology, it would diffuse to all other firms immediately (though imperfectly). Now, we argue that the knowledge that other firms can absorb depends on the amount of firms which have this knowledge. The more firms know about a specific technology, the easier it will be for other firms to imitate it. In mathematical terms, we will assume that the extent of spillovers $1 - \theta$ equals M_{RD} .

For direct imitation, we must derive a new percentage of sustainable innovators: M_{RD} is now on both sides of the equation, showing a feedback loop through spillovers:

$$M_{RD} = \left(\frac{P_{rd}}{2P_{rd} + \zeta - 2P_{rd}\zeta - P_{rd}M_{RD} + \zeta P_{rd}M_{RD}} \right) M \quad (42)$$

Despite the feedback loop, there is still only one equilibrium percentage of sustainable innovators for all possible configurations, which is stable²⁸. The original distribution of firms is irrelevant for what the equilibrium looks like in the end and shocks that change the firm distribution are quickly repaired. As far as results go, this specification does not lead to great differences with the results we found earlier for direct imitation. Each set of P_{rd} and ζ has a unique equilibrium, as it did before. The only real difference is that θ has become endogenous. But as long as firms take θ as given, nothing really changes.

For indirect imitation, the formula for the percentage of innovator firms does not change compared to our benchmark model, as all changes are again endogenous. However, the percentage itself does change. We cannot be exactly sure what happens, but there will be a new equilibrium with a higher percentage of sustainable innovators, as before. It becomes interesting, at this point, to look at the chance of entering as an innovator, if entry success is guaranteed.

$$P_{rd} = \frac{1 - G(\varphi_{RD}^*)}{1 - G(\varphi^*)} \quad (43)$$

We can already discuss one case here. If this adds up to $P_{rd} = M_{RD}$, the only stable equilibrium will be one in which all firms are innovators²⁹. The unstable equilibrium is actually impossible to even start with, because of the same argument we used before: as long as R&D costs are finite, there will always be at least one firm engaging in it, because there is no upper limit on the possible productivity levels firms may have³⁰. We can be sure that $G(\varphi^*)$ does not change if the percentage of sustainable innovators changes, so we must only consider what happens to $G(\varphi_{RD}^*)$. What happens to this probability when the mass of firms change?

²⁸ As M_{RD} goes up, the left-hand side always increases while the right-hand side always decreases. This implies that there is only one equilibrium.

²⁹ As long as some firms engage in innovation, ζM_{RD} sustainable innovators exit each period whereas $M_{RD}(\zeta M_{RD} + 1 - M_{RD})$ new sustainable innovators enter the market. Comparing these two gives $\zeta M_{RD} \leq M_{RD}(\zeta M_{RD} + 1 - M_{RD})$, which simplifies to $(\zeta - 1)M_{RD} \leq (\zeta - 1)M_{RD}^2$. Since $(\zeta - 1)$ is negative and $0 < M_{RD} < 1$, we know that $(\zeta - 1)M_{RD} < (\zeta - 1)M_{RD}^2$ and the number of sustainable innovators is always increasing as long as there is at least one sustainable innovator to begin with. The only alternative equilibrium is $M_{RD} = 0$, but this is an unstable equilibrium because if even one firm starts innovating the economy will move away from it.

³⁰ This is the result for any equation for which $P_{rd} \geq M_{RD}$ for all possible M_{RD} .

A way of thinking about this is by sorting firms in the economy to productivity. Doing this, we can find the first firm which does not engage in R&D investment. To find out if an equilibrium is stable, we need only ask what happens if that firm accidentally starts doing R&D. If engaging in R&D causes the cutoff point to move by such a length that the firm is now below it, the original equilibrium was unstable. If this condition holds for all firms, we know that there is only one stable equilibrium, when all firms are innovators. It is impossible to state any other condition or law: everything depends on the way in which firm productivities are distributed, which is not specified by the Melitz model. If we observe a normal distribution with a large number of firms around a specific productivity level, there will probably be very many or very few sustainable innovators, as the break-off point will likely be in one of the tails of the distribution.

5. Concluding remarks

We have added the geographical component to our model by introducing spillovers. We operationalized spillovers and showed a number of results when these are included in the analysis. We noted that the effect of spillovers on innovation in an economy is strongly reliant on the way we choose to model them. If we assume that firms can imitate technology directly and thus remove part of the endogenous chance without doing anything, we find that fewer firms innovate. If firms can only imitate technology indirectly and the spillovers simply make research for other firms cheaper, we find that a selection effect causes more firms innovate than in the benchmark equilibrium. However, in both cases, average productivity will decline when comparing complete spillovers to none, though there may be some intermediate point where spillovers have a positive effect on average productivity. In the final section, we allowed the size of the spillover to vary and found that this does not strongly change our results.

References

- Alcácer, J. and W. Chung (2007), “Location Strategies and Knowledge Spillovers,” *Management Science*, 53(5): 760-776

- Almeida, P. and B. Kogut (1999), “Localization of Knowledge and the Mobility of Engineers in Regional Networks”, *Management Science*, 45(7): 905-17
- Asheim, B.T. and M. Gertler (2005), “The Geography of Innovation: Regional Innovation Systems”, in: J. Fagerberg, D. Mowery, and R. Nelson (eds.), *The Oxford Handbook of Innovation*, Oxford University Press: 291-317
- Atkeson, A. and A.T. Burstein (2010), “Innovation, Firm Dynamics, and International Trade”, *Journal of Political Economy*, 118(3): 433-484.
- Atkeson, A. and A.T. Burstein (2011), “Aggregate Implications of Innovation Policy”, *NBER Working Paper Series 17493*, National Bureau of Economic Research, Cambridge, MA.
- Audretsch, D.B. & M.P. Feldman (1996), “R&D Spillovers And The Geography of Innovation and Production”, *American Economic Review*, 86(3): 630-40
- Audretsch, D.B & M.P. Feldman (2004), “Chapter 61: Knowledge spillovers and the geography of innovation”, in: J. V. Henderson and J-F. Thisse (eds), *Handbook of Regional and Urban Economics, Volume 4*: 2713-2739, Elsevier
- Baptista, R. and P. Swann (1998), “Do firms in clusters innovate more?”, *Research Policy*, 27: 525-540
- Bell, G.G. (2005), “Clusters, Networks, and Firm Innovativeness,” *Strategic Management Journal*, 26: 287-295
- Bustos, P. (2011), “Trade Liberalization, Exports, and technology Upgrading: Evidence on the of MERCOSUR on Argentinian Firms”, *American Economic Review*, 101(1): 304-340
- Burstein, A. and M. Melitz (2011), “Trade Liberalization and firm Dynamics”, *NBER Working Paper Series 16960*, National Bureau of Economic Research, Cambridge, MA.
- Bailey, M.N., C. Hulten, D. Campbell, T. Bresnahan and R.E. Caves (1992), “Productivity Dynamics in Manufacturing Plants”, *Brookings Papers on Economic Activity. Microeconomics*, Vol. 1992: 187-267.
- Baldwin, R. and T. Okubo (2006), “Heterogeneous firms, agglomeration and economic geography: spatial selection and sorting”, *Journal of Economic Geography*, 6: 323-346.

- Boschma, R.A. (2005), “Proximity and Innovation: A Critical Assessment”, *Regional Studies*, 39(1): 61-74. Brakman, S., H. Garretsen, and C. van Marrewijk (2009), *The New Introduction to Geographical Economics*, second edition, Cambridge University Press, Cambridge.
- Braunerhjelm, P. and M.P. Feldman (2006), *Cluster Genesis. Technology-Based Industrial Development*, Oxford University Press, Oxford
- Caves, R.E. (1998), “Industrial Organization and New Findings on the Turnover and Mobility of Firms”, *Journal of Economic Literature*, 36(4): 1947-1982
- Cefis, Elena and Marsili, Orietta (2005), “A Matter of Life and Death: Innovation and Firm Survival”, *LEM Working Paper Series*
- Cohen W.M., Levinthal D.A. (1990), “Absorptive Capacity: A New Perspective on Learning and Innovation,” *Administrative Science Quarterly*, 35(1): 128-152.
- Cohen, W.M. and Levinthal, D.A. (1989), “Innovation and Learning: The Two Faces of R&D,” *The Economic Journal*, 99(397): 569-596.
- Combes, P.-P., Duranton, G., Gobillon, L., Puga, D. and S. Roux (2012), “The Productivity Advantages of Large Cities: Distinguishing Agglomeration from Firm Selection,” *Econometrica*, 80(6): 2543-2594
- Decarolis, D.M. and D.L. Deeds (1999), “The Impact of Stocks and Flows of Organizational Knowledge on Firm Performance: An Empirical Investigation of the Biotechnology Industry,” *Strategic Management Journal*, 20(10): 953-968
- European Commission (2003), *Thematic report: Cluster policies, European trend chart on innovation*. Brussels: Enterprise Directorate-General.
- Feldman, M., J. Francis and J. Bercovitz (2005), “Creating a Cluster While Building a Firm: Entrepreneurs and the Formation of Industrial Clusters”, *Regional Studies*, 39(1): 129-141.
- Fenn, J. and M. Raskino (2008), *Mastering the Hype Cycle: How to Choose the Right Innovation at the Right Time*, Gartner, Inc. & Harvard Business Press.
- Findlay, R. (1978), “Relative Backwardness, Direct Foreign Investment, and the Transfer of Technology: A Simple Dynamic Model,” *The Quarterly Journal of Economics*, 92(1): 1-16.

- Fujita, M., Krugman, P. and A.J. Venables (1999), *The Spatial Economy; Cities, Regions, and International Trade*, M.I.T. Press, Cambridge, MA.
- Garcia, R. and R. Calantone (2001), “A Critical Look At Technological Innovation Typology And Innovativeness Terminology: A Literature Review”, *Product Innovation Management*, 19: 110-132.
- Gollotto, J.C. and S. Kim (2003), “Market Valuation of Dot Com Companies; R&D versus Hype”, *Managerial Finance*, 29(11): 61-72
- Grossman, G.M. and E. Helpman (1991), “Quality Ladders in the Theory of Growth”, *The Review of Economic Studies*, 58(1): 43-61
- Iammarino, S. and P. McCann (2006), “The structure and evolution of industrial clusters: Transaction, technology and knowledge spillovers”, *Research Policy*, 35: 1018-1036.
- Järvenpää, H. and S.J. Mäkinen (2008), “Empirically Detecting the Hype Cycle with the Life Cycle Indicators: An Exploratory Analysis of Three Technologies”, *Proceeding of the 2008 International Conference on Industrial Engineering and Engineering Management (IEEM)*, Singapore: Institute of Electrical and Electronics Engineers (IEEE).
- Jaffe, A. B., M. Trajtenberg & M. S. Fogarty (2000), “Knowledge spillovers and Patent Citations: Evidence from a Survey of Inventors”, *American Economic Review*, 90 (2), 215-18
- Jaffe, A. B., M. Trajtenberg & R. Henderson (1993), “Geographic Localization Of Knowledge Spillovers As Evidenced By Patent Citations”, *Quarterly Journal Of Economics*, 108 (3), 577-98
- Jaffe, A. B. and M. Trajtenberg (2002), *Patents, Citations, and Innovations: A Window on the Knowledge Economy*, M.I.T. Press.
- Jensen, J Bradford & McGuckin, Robert H, 1997. "Firm Performance and Evolution: Empirical Regularities in the US Microdata," *Industrial and Corporate Change*, Oxford University Press, 6(1): 25-47.
- Keller, W. (2002), “Geographic localization of international technology diffusion”, *The American Economic Review*, 92(1): 120-142.

- Malerba, F. and L. Orsenigo (2001), “Towards a History Friendly Model of Innovation, Market Structure and Regulation in the Dynamics of the Pharmaceutical Industry: The Age of Random Screening,” CESPRI working paper
- Malerba, F. and L. Orsenigo (2002), “Innovation and Market Structure in the Dynamics of the Pharmaceutical Industry and Biotechnology: Towards a History Friendly Model,” *Industrial and Corporate Change*, 11(4): 667-703
- Maliranta, M., Mohnen, P. and P. Rouvinen (2009), “Is Inter-Firm Labor Mobility a Channel of Knowledge Spillovers? Evidence from a Linked Employer-Employee Panel”, *Industrial and Corporate Change*, 18(6): 1161-91
- McCann, P. and T.B. Folta (2009), Demand- and Supply-Side Agglomerations: Distinguishing between Fundamentally Different Manifestations of Geographic Concentration, *Journal of Management Studies* 46: 362-392.
- Melitz, M.J. (2003), “The impact of trade on intra-industry reallocations and aggregate industry productivity”, *Econometrica*, 71(6), November 2003, 1695-1725.
- Møen, J. (2005), “Is Mobility of Technical Personnel a Source of R&D Spillovers”, *Journal of Labor Economics*, 23(1): 81-114
- Molina-Morales, F.X. and M.T. Martínez-Fernández (2003), “The Impact of Industrial District Affiliation on Firm Value Creation,” *European Planning Studies*, 11(2): 155-170
- Molina-Morales, F.X. and M.T. Martínez-Fernández (2009), “Too much love in the neighborhood can hurt: How an excess of intensity and trust in relationships may produce negative effects on firms”, *Strategic Management Journal*, 30: 1013–1023.
- Nathan, M. and H. Overman (2013), “Agglomeration, clusters, and industrial policy”, *Oxford Review of Economic Policy*, 29(2): 383-404
- Pérez, Silvano Esteve, Llopis, Amparo Sanchis and Llopis, Juan Alberto Sanchis (2004), “The Determinants of Survival of Spanish Manufacturing Firms”, *Review of Industrial Organization*, 25: 251-273
- Pouder, R. and St. John, C. H. (1996). “Hot spots and blind spots: geographical clusters of firms and innovation”, *Academy of Management Review*, 21: 1192–1225.
- Rocha, H. (2004), “Entrepreneurship and Development: The Role of Clusters”, *Small Business Economics*, 23: 363-400
- Romer, P. (1986), “Increasing Returns and Long-Run Growth”. *Journal of Political Economy* 94(5): 1002-1037.

- Shaver, J.M. and F. Flyer (2000), “Agglomeration Economies, Firm Heterogeneity, and Foreign Direct Investment in the United States,” *Strategic Management Journal*, 21: 1175-1193
- Sölvell, Ö., Lindqvist, G., & Ketels, C. (2003). *The cluster initiative greenbook*. Stockholm: Ivory Tower.
- Swann, G.M.P. (2006), “Cluster and hinterland; When is a proactive cluster policy appropriate”, in: Asheim, B., P. Cooke and R. Martin (eds): *Clusters and Regional development: Critical Reflections and Explorations*”, Routledge, Abingdon.
- Van Long, N., H. Raff, and F. Stähler (2011), “Innovations and trade with heterogeneous firms”, *Journal of International Economics*, 84: 149-159.
- Vannoorenberghe, G. (2009), “Trade liberalization, heterogeneous firms and endogenous investment”, *CDSE Discussion Paper No. 44*, Center for Doctoral Studies in Economics, University of Mannheim.
- Wennberg, A. and G. Lindqvist (2010), “The effect of clusters on the survival and performance of new firms”, *Small Business Economics*, 34:221–241
- Yeaple, S.R. (2005), A Simple Model of Firm Heterogeneity, International Trade, and Wages”, *Journal of International Economics*, 65(1): 1-20.
- Zucker, L.G., M.R. Darby and J. Armstrong (1998), “Geographically Localized Knowledge: Spillovers or Markets?”, *Economic Inquiry*, 36(10): 65-86.

Appendix

A.1 Derivation of $\partial M / \partial P_{RD}$ and $\partial M_{RD} / \partial P_{RD}$

$$\bullet \quad \partial M / \partial P_{RD} = -\frac{L\bar{f}_{RD}}{[DenM]^2} \frac{\zeta(1-(1-\zeta)\theta)}{[Den]^2} < 0 \quad \text{and} \quad \partial M_{RD} / \partial P_{RD} = \frac{\zeta(1-(1-\zeta)\theta)}{[Den]^2} M > 0$$

A.2 Derivation of $\hat{\varphi}^* = Z d\theta$

Direct imitation

Rewrite the free-entry (36) to

$$\begin{aligned}
C(\theta) & \left(\left(\int_{\varphi^*}^{\varphi_{RD}^*} \left(\frac{\varphi}{\varphi^*} \right)^{\sigma-1} g(\varphi) d\varphi \right) - (G(\varphi_{RD}^*) - G(\varphi^*)) \right) + \frac{1}{\zeta} \left(\int_{\varphi_{RD}^*}^{\infty} \left(\frac{\varphi}{\varphi^*} \right)^{\sigma-1} g(\varphi) d\varphi \right) \\
& = \frac{1}{\zeta} (1 - G(\varphi_{RD}^*)) \left(1 + \frac{\bar{f}_{RD}}{f} \right) + \frac{f_e}{f}
\end{aligned}$$

with $C(\theta) = 1/[1 - (1 - \zeta)\theta]$.

Totally differentiate and rearrange:

$$\begin{aligned}
(1 - \zeta)C(\theta)^2 & \left[\left(\int_{\varphi^*}^{\varphi_{RD}^*} \left(\frac{\varphi}{\varphi^*} \right)^{\sigma-1} g(\varphi) d\varphi \right) - (G(\varphi_{RD}^*) - G(\varphi^*)) \right] d\theta \\
& + (1 - \sigma)\varphi^{*1-\sigma} \left[C(\theta) \left(\int_{\varphi^*}^{\varphi_{RD}^*} \varphi^{\sigma-1} g(\varphi) d\varphi \right) + \frac{1}{\zeta} \left(\int_{\varphi_{RD}^*}^{\infty} \varphi^{\sigma-1} g(\varphi) d\varphi \right) \right] \hat{\varphi}^* \\
& - C(\theta) [1 - (\varphi_{RD}^*/\varphi^*)^{\sigma-1}] dG(\varphi_{RD}^*) - \frac{1}{\zeta} \left(\frac{\varphi_{RD}^*}{\varphi^*} \right) dG(\varphi_{RD}^*) \\
& = -\frac{1}{\zeta} \left(1 + \frac{\bar{f}_{RD}}{f} \right) dG(\varphi_{RD}^*)
\end{aligned}$$

Apply $dG(\varphi_{RD}^*) = g(\varphi_{RD}^*)\varphi_{RD}^* \hat{\varphi}_{RD}^*$ and $dG(\varphi^*) = g(\varphi^*)\varphi^* \hat{\varphi}^*$, using $\hat{\varphi}_{RD}^* = \hat{\varphi}^* + X d\theta$ with $X = \left(\frac{\zeta C(\theta)}{(1-\theta)(\sigma-1)} \right) \left(1 - \left(\frac{\varphi^*}{\varphi_{RD}^*} \right)^{\sigma-1} \right) > 0$, we get,

$$\begin{aligned}
& \left[(1 - \zeta)C(\theta)^2 \left(\int_{\varphi^*}^{\varphi_{RD}^*} \left(\frac{\varphi}{\varphi^*}\right)^{\sigma-1} g(\varphi) d\varphi - (G(\varphi_{RD}^*) - G(\varphi^*)) \right) \right. \\
& \quad - \varphi_{RD}^* g(\varphi_{RD}^*) \frac{\zeta C(\theta)}{(1 - \theta)(\sigma - 1)} \left(1 - \left(\frac{\varphi^*}{\varphi_{RD}^*}\right)^{\sigma-1} \right) \left(C(\theta) \left(1 - \left(\frac{\varphi_{RD}^*}{\varphi^*}\right)^{\sigma-1} \right) \right. \\
& \quad \left. \left. + \frac{1}{\zeta} \left(\frac{\varphi_{RD}^*}{\varphi^*}\right)^{\sigma-1} - \frac{1}{\zeta} \left(1 + \frac{\bar{f}_{RD}}{f} \right) \right) \right] d\theta \\
& \quad + (1 - \sigma)\varphi^{*1-\sigma} \left[C(\theta) \left(\int_{\varphi^*}^{\varphi_{RD}^*} \varphi^{\sigma-1} g(\varphi) d\varphi \right) + \frac{1}{\zeta} \left(\int_{\varphi_{RD}^*}^{\infty} \varphi^{\sigma-1} g(\varphi) d\varphi \right) \right] \hat{\varphi}^* \\
& \quad - C(\theta) g(\varphi_{RD}^*) \varphi_{RD}^* \left(1 - \left(\frac{\varphi_{RD}^*}{\varphi^*}\right)^{\sigma-1} \right) \hat{\varphi}^* \\
& \quad - \frac{1}{\zeta} \left(\left(\frac{\varphi_{RD}^*}{\varphi^*}\right)^{\sigma-1} - \left(1 + \frac{\bar{f}_{RD}}{f} \right) \right) g(\varphi_{RD}^*) \varphi_{RD}^* \hat{\varphi}^* = 0
\end{aligned}$$

Hence,

$$(a - b)d\theta + (c - d - e)\hat{\varphi}^* = 0$$

The expression in the main text is obtained by defining $Z = (a-b)/(c-d-e)$, with:

$$\begin{aligned}
a &= (1 - \zeta)C(\theta)^2 \left(\int_{\varphi^*}^{\varphi_{RD}^*} \left(\frac{\varphi}{\varphi^*}\right)^{\sigma-1} g(\varphi) d\varphi - (G(\varphi_{RD}^*) - G(\varphi^*)) \right) \\
b &= \varphi_{RD}^* g(\varphi_{RD}^*) \frac{\zeta C(\theta)}{(1 - \theta)(\sigma - 1)} \left(1 - \left(\frac{\varphi^*}{\varphi_{RD}^*}\right)^{\sigma-1} \right) \left(C(\theta) \left(1 - \left(\frac{\varphi_{RD}^*}{\varphi^*}\right)^{\sigma-1} \right) + \frac{1}{\zeta} \left(\frac{\varphi_{RD}^*}{\varphi^*}\right)^{\sigma-1} \right. \\
& \quad \left. - \frac{1}{\zeta} \left(1 + \frac{\bar{f}_{RD}}{f} \right) \right) \\
c &= (1 - \sigma)\varphi^{*1-\sigma} \left[C(\theta) \left(\int_{\varphi^*}^{\varphi_{RD}^*} \varphi^{\sigma-1} g(\varphi) d\varphi \right) + \frac{1}{\zeta} \left(\int_{\varphi_{RD}^*}^{\infty} \varphi^{\sigma-1} g(\varphi) d\varphi \right) \right]
\end{aligned}$$

$$d = C(\theta)g(\varphi_{RD}^*)\varphi_{RD}^* \left(1 - \left(\left(\frac{\varphi_{RD}^*}{\varphi^*} \right)^{\sigma-1} \right) \right)$$

$$e = \frac{1}{\zeta} \left[\left(\left(\frac{\varphi_{RD}^*}{\varphi^*} \right)^{\sigma-1} - 1 \right) - \frac{\bar{f}_{RD}}{f} \right] g(\varphi_{RD}^*)\varphi_{RD}^*.$$

Applying the Pareto-distribution $G(\varphi) = 1 - \left(\frac{b_m}{\varphi}\right)^\alpha$ and $dG(\varphi) = \alpha \frac{b_m^\alpha}{\varphi^{\alpha+1}} d\varphi$, these expressions reduce to:

$$a = b_m^\alpha (1 - \zeta) C(\theta)^2 \left[\left(\frac{\alpha}{\sigma - \alpha - 1} \left(\frac{\varphi_{RD}^*}{\varphi^*} \right)^{\sigma-1} + 1 \right) \varphi_{RD}^{*\alpha} - \frac{\sigma - 1}{\sigma - \alpha - 1} \varphi^{*\alpha} \right] \leq 0$$

$$b = \alpha \left(\frac{b_m}{\varphi_{RD}^*} \right)^\alpha \frac{C(\theta)}{(1 - \theta)(\sigma - 1)} \left(1 - \left(\frac{\varphi_{RD}^*}{\varphi^*} \right)^{\sigma-1} \right) \left((1 - \zeta C(\theta)) \left(\left(\frac{\varphi_{RD}^*}{\varphi^*} \right)^{\sigma-1} - 1 \right) - \frac{\bar{f}_{RD}}{f} \right) \leq 0$$

$$c = \frac{(1 - \sigma)\varphi^{*1-\sigma} \alpha b_m^\alpha}{\sigma - \alpha - 1} \left[\left(C(\theta) - \frac{1}{\zeta} \right) \varphi_{RD}^{*\sigma-\alpha-1} - C(\theta) \varphi^{*\sigma-\alpha-1} \right] < 0$$

$$d = C(\theta) \alpha \left(\frac{b_m}{\varphi_{RD}^*} \right)^\alpha \left(1 - \left(\frac{\varphi_{RD}^*}{\varphi^*} \right)^{\sigma-1} \right) < 0$$

$$e = \frac{\alpha}{\zeta} \left(\frac{b_m}{\varphi_{RD}^*} \right)^\alpha \left[\left(\left(\frac{\varphi_{RD}^*}{\varphi^*} \right)^{\sigma-1} - 1 \right) - \frac{\bar{f}_{RD}}{f} \right] \leq 0.$$

From these expressions we infer that $b > (<) 0$ if $(1 - \zeta C(\theta)) \left(\left(\frac{\varphi_{RD}^*}{\varphi^*} \right)^{\sigma-1} - 1 \right) > (<) \bar{f}_{RD}/f$.

Furthermore, noting that

$$d + e = \frac{\alpha}{\zeta} \left(\frac{b_m}{\varphi_{RD}^*} \right)^\alpha \left[(1 - \zeta C(\theta)) \left(\left(\frac{\varphi_{RD}^*}{\varphi^*} \right)^{\sigma-1} - 1 \right) - \frac{\bar{f}_{RD}}{f} \right],$$

it follows $d+e >0 (<0)$ if this condition holds.

Since $c < 0$, this implies that the denominator of Z has a clear negative sign provided $(1 - \zeta C(\theta)) \left(\left(\frac{\varphi_{RD}^*}{\varphi^*} \right)^{\sigma-1} - 1 \right) > \bar{f}_{RD}/f$. This implies that if this latter condition holds, to get a definite sign for Z requires $a < 0$. In that case $Z > 0$. A necessary condition for $a < 0$ is:

$$\left[\left(\frac{\alpha}{\sigma - \alpha - 1} \left(\frac{\varphi_{RD}^*}{\varphi^*} \right)^{\sigma-1} + 1 \right) \varphi_{RD}^{*-\alpha} - \frac{\sigma - 1}{\sigma - \alpha - 1} \varphi^{*-\alpha} \right] < 0.$$

For $\theta=1$ we get ($C(1) = \frac{1}{\zeta}$ and $\frac{\varphi_{RD}^*}{\varphi^*} \rightarrow 1$):

$$a \rightarrow 0$$

$$b \rightarrow -0$$

$$c = -\frac{\alpha}{\zeta} \frac{(1 - \sigma)}{\sigma - \alpha - 1} \left(\frac{b_m}{\varphi^*} \right)^\alpha < 0$$

$$d = 0$$

$$e = -\frac{\alpha}{\zeta} \left(\frac{b_m}{\varphi_{RD}^*} \right)^\alpha \left(1 + \frac{\bar{f}_{RD}}{f} \right) < 0.$$

Hence $Z = (a-b)/(c-d-e)$ reduces to $Z \rightarrow 0$.

For $\theta=0$ we get ($C(0) = 1$):

$$a = b_m^\alpha (1 - \zeta) \left[\left(\frac{\alpha}{\sigma - \alpha - 1} \left(\frac{\varphi_{RD}^*}{\varphi^*} \right)^{\sigma-1} + 1 \right) \varphi_{RD}^{*-\alpha} - \frac{\sigma - 1}{\sigma - \alpha - 1} \varphi^{*-\alpha} \right] \leq 0$$

$$b = \alpha \left(\frac{b_m}{\varphi_{RD}^*} \right)^\alpha \frac{1}{(1 - \theta)(\sigma - 1)} \left(1 - \left(\frac{\varphi^*}{\varphi_{RD}^*} \right)^{\sigma-1} \right) \left((1 - \zeta) \left(\left(\frac{\varphi_{RD}^*}{\varphi^*} \right)^{\sigma-1} - 1 \right) - \frac{\bar{f}_{RD}}{f} \right)$$

$$c = \frac{(1-\sigma)\varphi^{*1-\sigma}\alpha b_m^\alpha}{\sigma-\alpha-1} \left[\left(1 - \frac{1}{\zeta}\right) \varphi_{RD}^{*\sigma-\alpha-1} - \varphi^{*\sigma-\alpha-1} \right] < 0$$

$$d = \alpha \left(\frac{b_m}{\varphi_{RD}^*} \right)^\alpha \left(1 - \left(\frac{\varphi_{RD}^*}{\varphi^*} \right)^{\sigma-1} \right) < 0$$

$$e = \frac{\alpha}{\zeta} \left(\frac{b_m}{\varphi_{RD}^*} \right)^\alpha \left[\left(\left(\frac{\varphi_{RD}^*}{\varphi^*} \right)^{\sigma-1} - 1 \right) - \frac{\bar{f}_{RD}}{f} \right] \leq 0.$$

Indirect imitation

Rewrite the free-entry (40) to

$$\begin{aligned} & \left(\left(\int_{\varphi^*}^{\varphi_{RD}^*} \left(\frac{\varphi}{\varphi^*} \right)^{\sigma-1} g(\varphi) d\varphi \right) - (G(\varphi_{RD}^*) - G(\varphi^*)) \right) + \frac{1}{\zeta} \left(\left(\int_{\varphi_{RD}^*}^{\infty} \left(\frac{\varphi}{\varphi^*} \right)^{\sigma-1} g(\varphi) d\varphi \right) \right) \\ & = \frac{1}{\zeta} (1 - G(\varphi_{RD}^*)) \left(1 + \frac{(1-\theta)\bar{f}_{RD}}{f} \right) + \frac{f_e}{f}. \end{aligned}$$

Totally differentiate and rearrange:

$$\begin{aligned} & (1-\sigma)\varphi^{*1-\sigma} \left[\left(\int_{\varphi^*}^{\varphi_{RD}^*} \varphi^{\sigma-1} g(\varphi) d\varphi \right) + \frac{1}{\zeta} \left(\int_{\varphi_{RD}^*}^{\infty} \varphi^{\sigma-1} g(\varphi) d\varphi \right) \right] \hat{\varphi}^* \\ & - [1 - (\varphi_{RD}^*/\varphi^*)^{\sigma-1}] dG(\varphi_{RD}^*) - \frac{1}{\zeta} \left(\frac{\varphi_{RD}^*}{\varphi^*} \right) dG(\varphi_{RD}^*) \\ & = -\frac{1}{\zeta} \left(1 + \frac{(1-\theta)\bar{f}_{RD}}{f} \right) dG(\varphi_{RD}^*) - \frac{1}{\zeta} (1 - G(\varphi_{RD}^*)) \frac{\bar{f}_{RD}}{f} d\theta \end{aligned}$$

Apply $dG(\varphi_{RD}^*) = g(\varphi_{RD}^*)\varphi_{RD}^*\hat{\varphi}_{RD}^*$ and $dG(\varphi^*) = g(\varphi^*)\varphi^*\hat{\varphi}^*$, using $\hat{\varphi}_{RD}^* = \hat{\varphi}^* + \left(\frac{\bar{f}_{RD}/f}{(1-\zeta)(\sigma-1)} \right) d\theta$ we get,

$$\begin{aligned}
& \left[\frac{1}{\zeta} (1 - G(\varphi_{RD}^*)) \frac{\bar{f}_{RD}}{f} - \varphi_{RD}^* g(\varphi_{RD}^*) \left(\frac{\bar{f}_{RD}/f}{(1-\zeta)(\sigma-1)} \right) \times \right. \\
& \left. \left[\left(\left(1 - \left(\frac{\varphi_{RD}^*}{\varphi^*} \right)^{\sigma-1} \right) + \frac{1}{\zeta} \left(\frac{\varphi_{RD}^*}{\varphi^*} \right)^{\sigma-1} - \frac{1}{\zeta} \left(1 + \frac{(1-\theta)\bar{f}_{RD}}{f} \right) \right] d\theta \right. \\
& + (1-\sigma)\varphi^{*1-\sigma} \left[\left(\int_{\varphi^*}^{\varphi_{RD}^*} \varphi^{\sigma-1} g(\varphi) d\varphi \right) + \frac{1}{\zeta} \left(\int_{\varphi_{RD}^*}^{\infty} \varphi^{\sigma-1} g(\varphi) d\varphi \right) \right] \hat{\varphi}^* \\
& - \varphi_{RD}^* g(\varphi_{RD}^*) \left(1 - \left(\left(\frac{\varphi_{RD}^*}{\varphi^*} \right)^{\sigma-1} \right) \right) \hat{\varphi}^* \\
& \left. - \frac{1}{\zeta} \left(\left(\left(\frac{\varphi_{RD}^*}{\varphi^*} \right)^{\sigma-1} - 1 \right) - \frac{(1-\theta)\bar{f}_{RD}}{f} \right) g(\varphi_{RD}^*) \varphi_{RD}^* \hat{\varphi}^* = 0
\end{aligned}$$

Hence,

$$(a-b)d\theta + (c-d-e)\hat{\varphi}^* = 0$$

The expression in the main text is obtained by defining $Z' = (a-b)/(c-d-e)$, with:

$$\begin{aligned}
a &= \frac{1}{\zeta} (1 - G(\varphi_{RD}^*)) \frac{\bar{f}_{RD}}{f} > 0 \\
b &= \varphi_{RD}^* g(\varphi_{RD}^*) \frac{1}{\zeta} \left(\frac{\bar{f}_{RD}/f}{(1-\zeta)(\sigma-1)} \right) \left((1-\zeta) \left(\left(\frac{\varphi_{RD}^*}{\varphi^*} \right)^{\sigma-1} - 1 \right) - \frac{(1-\theta)\bar{f}_{RD}}{f} \right) \\
c &= (1-\sigma)\varphi^{*1-\sigma} \left[\left(\int_{\varphi^*}^{\varphi_{RD}^*} \varphi^{\sigma-1} g(\varphi) d\varphi \right) + \frac{1}{\zeta} \left(\int_{\varphi_{RD}^*}^{\infty} \varphi^{\sigma-1} g(\varphi) d\varphi \right) \right] \\
d &= g(\varphi_{RD}^*) \varphi_{RD}^* \left(1 - \left(\left(\frac{\varphi_{RD}^*}{\varphi^*} \right)^{\sigma-1} \right) \right) \\
e &= \frac{1}{\zeta} \left(\left(\left(\frac{\varphi_{RD}^*}{\varphi^*} \right)^{\sigma-1} - 1 \right) - \frac{(1-\theta)\bar{f}_{RD}}{f} \right) g(\varphi_{RD}^*) \varphi_{RD}^* \hat{\varphi}^*.
\end{aligned}$$

Applying the Pareto-distribution $G(\varphi) = 1 - \left(\frac{b_m}{\varphi}\right)^\alpha$ and $dG(\varphi) = \alpha \frac{b_m^\alpha}{\varphi^{\alpha+1}} d\varphi$, these expressions reduce to:

$$a = \frac{1}{\zeta} \left(\frac{b_m}{\varphi_{RD}^*}\right)^\alpha \frac{\bar{f}_{RD}}{f} > 0$$

$$b = \frac{\alpha}{\zeta} \left(\frac{b_m}{\varphi_{RD}^*}\right)^\alpha \left(\frac{\bar{f}_{RD}/f}{(1-\zeta)(\sigma-1)}\right) \left((1-\zeta) \left(\left(\frac{\varphi_{RD}^*}{\varphi^*}\right)^{\sigma-1} - 1\right) - \frac{(1-\theta)\bar{f}_{RD}}{f} \right)$$

$$c = \frac{(1-\sigma)\varphi^{*1-\sigma} \alpha b_m^\alpha}{\sigma - \alpha - 1} \left[\left(1 - \frac{1}{\zeta}\right) \varphi_{RD}^{*\sigma-\alpha-1} - \varphi^{*\sigma-\alpha-1} \right] < 0$$

$$d = \alpha \left(\frac{b_m}{\varphi_{RD}^*}\right)^\alpha \left(1 - \left(\frac{\varphi_{RD}^*}{\varphi^*}\right)^{\sigma-1}\right) < 0$$

$$e = \frac{\alpha}{\zeta} \left(\frac{b_m}{\varphi_{RD}^*}\right)^\alpha \left(\left(\left(\frac{\varphi_{RD}^*}{\varphi^*}\right)^{\sigma-1} - 1\right) - \frac{(1-\theta)\bar{f}_{RD}}{f} \right) \leq 0.$$

From these expressions we infer that $b > (<) 0$ if $(1-\zeta) \left(\left(\frac{\varphi_{RD}^*}{\varphi^*}\right)^{\sigma-1} - 1\right) > (<)(1-\theta)\bar{f}_{RD}/f$.

Furthermore, noting that

$$d + e = \frac{\alpha}{\zeta} \left(\frac{b_m}{\varphi_{RD}^*}\right)^\alpha \left[(1-\zeta) \left(\left(\frac{\varphi_{RD}^*}{\varphi^*}\right)^{\sigma-1} - 1\right) - \frac{(1-\theta)\bar{f}_{RD}}{f} \right],$$

it follows $d+e > 0 (< 0)$ if this condition holds.

Since $c < 0$, this implies that the denominator of Z' has a clear negative sign provided $(1-\zeta) \left(\left(\frac{\varphi_{RD}^*}{\varphi^*}\right)^{\sigma-1} - 1\right) > (1-\theta)\bar{f}_{RD}/f$. This implies that if this latter condition holds, to get a definite sign for Z requires $a-b < 0$. In that case $Z > 0$.

$$a - b = \frac{1}{\zeta} \left(\frac{b_m}{\varphi_{RD}^*} \right)^\alpha \frac{\bar{f}_{RD}}{f} \left[1 - \frac{\alpha}{(1-\zeta)(\sigma-1)} \left((1-\zeta) \left(\left(\frac{\varphi_{RD}^*}{\varphi^*} \right)^{\sigma-1} - 1 \right) - \frac{(1-\theta)\bar{f}_{RD}}{f} \right) \right].$$

We note that if $\left((1-\zeta) \left(\left(\frac{\varphi_{RD}^*}{\varphi^*} \right)^{\sigma-1} - 1 \right) - \frac{(1-\theta)\bar{f}_{RD}}{f} \right) > 0$, it follows that $(a-b) < 0$ if and only if $\frac{\alpha}{(1-\zeta)(\sigma-1)}$ is sufficiently large. This is accomodated by a sufficiently low ζ and/or σ (noting that $\alpha > \sigma - 1$ by assumption).

For $\theta=1$ we get $\left(\frac{\varphi_{RD}^*}{\varphi^*} \rightarrow 1 \right)$: $a = \frac{1}{\zeta} \left(\frac{b_m}{\varphi_{RD}^*} \right)^\alpha \frac{\bar{f}_{RD}}{f} > 0$, $b = 0$, $c =$

$-\frac{1}{\zeta} \frac{(1-\sigma)\varphi^{*1-\sigma} \alpha b_m^\alpha}{\sigma-\alpha-1} \varphi_{RD}^{*\sigma-\alpha-1} < 0$, $d = 0$, and $e = 0$. Hence $Z = (a-b)/(c-d-e)$ reduces to a/c ,

which equals:

$$-\frac{(\sigma - \alpha - 1) \bar{f}_{RD}/f}{\alpha(1-\sigma)\varphi^{*1-\sigma} \varphi_{RD}^{*\sigma-1}} < 0.$$

For $\theta=0$ we get:

$$a = \frac{1}{\zeta} \left(\frac{b_m}{\varphi_{RD}^*} \right)^\alpha \frac{\bar{f}_{RD}}{f} > 0$$

$$b = \alpha \left(\frac{b_m}{\varphi_{RD}^*} \right)^\alpha \left(\frac{\bar{f}_{RD}/f}{(1-\zeta)(\sigma-1)} \right) \left((1-\zeta) \left(\left(\frac{\varphi_{RD}^*}{\varphi^*} \right)^{\sigma-1} - 1 \right) - \frac{(1-\theta)\bar{f}_{RD}}{f} \right)$$

$$c = \frac{(1-\sigma)\varphi^{*1-\sigma} \alpha b_m^\alpha}{\sigma-\alpha-1} \left[\left(1 - \frac{1}{\zeta} \right) \varphi_{RD}^{*\sigma-\alpha-1} - \varphi^{*\sigma-\alpha-1} \right] < 0$$

$$d = \alpha \left(\frac{b_m}{\varphi_{RD}^*} \right)^\alpha \left(1 - \left(\frac{\varphi_{RD}^*}{\varphi^*} \right)^{\sigma-1} \right) < 0$$

$$e = \frac{\alpha}{\zeta} \left(\frac{b_m}{\varphi_{RD}^*} \right)^\alpha \left(\left(\frac{\varphi_{RD}^*}{\varphi^*} \right)^{\sigma-1} - \left(1 + \frac{\bar{f}_{RD}}{f} \right) \right) \leq 0.$$

A.3 Derivation $d\left(\frac{M_{RD}}{M}\right) = \frac{1}{M} \frac{\partial P_{rd}}{\partial \theta} \left\{ \frac{\partial M_{RD}}{\partial P_{rd}} - \frac{M_{RD}}{M} \frac{\partial M}{\partial P_{rd}} \right\} d\theta < 0$ for $d\theta > 0$

Direct imitation: $d\left(\frac{M_{RD}}{M}\right) = \frac{1}{M} \frac{\partial P_{rd}}{\partial \theta} \left\{ \frac{\partial M_{RD}}{\partial P_{rd}} - \frac{M_{RD}}{M} \frac{\partial M}{\partial P_{rd}} \right\} d\theta < 0$ for $d\theta > 0$

- $\partial M / \partial P_{RD} = -\frac{\zeta(1-(1-\zeta)\theta)\bar{f}_{RD}}{L[(1-(1-\zeta)\theta)P_{rd}/MRD]^2} < 0$
- $\partial M_{RD} / \partial P_{RD} = \frac{\zeta(1-(1-\zeta)\theta)}{M[(1-(1-\zeta)\theta)P_{rd}/Mrd]^2} > 0$
- $\frac{\partial P_{rd}}{\partial \theta} < 0$ through its dependence with $\hat{\varphi}^*$.
 - $\frac{dP_{RD}}{P_{RD}} = -\left[\frac{G'(\varphi_{RD}^*)}{1-G(\varphi_{RD}^*)} d\varphi_{RD}^* - \frac{G'(\varphi^*)}{1-G(\varphi^*)} d\varphi^* \right]$
 - $\frac{dP_{RD}}{P_{RD}} = -\left[\frac{G'(\varphi_{RD}^*)}{1-G(\varphi_{RD}^*)} \frac{\varphi_{RD}^*}{\varphi^*} - \frac{G'(\varphi^*)}{1-G(\varphi^*)} \right] d\varphi^* - \varphi_{RD}^* \frac{G'(\varphi^*)}{1-G(\varphi^*)} X d\theta$

with $X \equiv \left(\frac{\zeta}{(1-\theta)[1-(1-\zeta)\theta]} \right) \left(1 - \left(\frac{\varphi^*}{\varphi_{RD}^*} \right)^{\sigma-1} \right) > 0$. For the Pareto-distribution the term in brackets becomes zero ($G'/(1-G) = \alpha/\varphi$), implying that $\frac{dP_{RD}}{P_{RD}} = -\alpha X d\theta < 0$.

Indirect imitation: $d\left(\frac{M_{RD}}{M}\right) = \frac{1}{M} \frac{\partial P_{rd}}{\partial \theta} \left\{ \frac{\partial M_{RD}}{\partial P_{rd}} - \frac{M_{RD}}{M} \frac{\partial M}{\partial P_{rd}} \right\} d\theta < 0$ for $d\theta > 0$

- $\partial M / \partial P_{RD} = -\frac{\zeta\bar{f}_{RD}}{L[P_{rd}/MRD]^2} < 0$
- $\partial M_{RD} / \partial P_{RD} = \frac{\zeta}{M[P_{rd}/Mrd]^2} > 0$
- $\frac{\partial P_{rd}}{\partial \theta} < 0$ through its dependence with $\hat{\varphi}^*$.
 - $\frac{dP_{RD}}{P_{RD}} = -\left[\frac{G'(\varphi_{RD}^*)}{1-G(\varphi_{RD}^*)} d\varphi_{RD}^* - \frac{G'(\varphi^*)}{1-G(\varphi^*)} d\varphi^* \right]$
 - $\frac{dP_{RD}}{P_{RD}} = -\left[\frac{G'(\varphi_{RD}^*)}{1-G(\varphi_{RD}^*)} \frac{\varphi_{RD}^*}{\varphi^*} - \frac{G'(\varphi^*)}{1-G(\varphi^*)} \right] d\varphi^* - \varphi_{RD}^* \frac{G'(\varphi^*)}{1-G(\varphi^*)} X d\theta$

with $X \equiv \left(\frac{\bar{f}_{RD}/f}{(1-\zeta)(\sigma-1)} \right) > 0$. For the Pareto-distribution the term in brackets becomes zero; ($G'/(1-G) = \alpha/\varphi$), implying that $\frac{dP_{RD}}{P_{RD}} = -\alpha X d\theta < 0$.

A.4 Effect on M_{rd} and M , keeping $\tilde{\varphi}$ constant (direct imitation case):

$$\begin{aligned} \circ \quad \frac{dM_{RD}}{d\theta} &= \frac{(1-\zeta)P_{rd}(1-P_{rd})\zeta}{[D_{RD}]^2} M \left\{ \frac{M_{RD}\bar{f}_{RD}}{L} - 1 \right\} + \frac{\zeta(1-(1-\zeta)\theta)}{[D_{RD}]^2} M \frac{\partial P_{RD}}{\partial \theta} \\ \circ \quad \frac{dM}{d\theta} &= -\frac{L\bar{f}_{RD}}{[D_M]^2[D_{RD}]^2} \left\{ \zeta(1-(1-\zeta)\theta) \frac{\partial P_{RD}}{\partial \theta} - (1-\zeta)P_{rd}(1-P_{rd})\zeta \right\} \end{aligned}$$

where D_{RD} and D_M are short-hand notations for, respectively, the denominators of the expressions for M_{RD} and M .

As long as $\frac{\partial P_{RD}}{\partial \theta} \leq 0$, it follows that the number of R&D firms in the cluster falls ($dM_{RD}/d\theta < 0$)³¹, whereas the total number of firms M goes up ($dM/d\theta > 0$). Ceteris paribus $\tilde{\varphi}$, the possibility of spillovers by direct imitation then lowers the number of R&D firms in the cluster.

A.5 Derivation of $\frac{d\tilde{\varphi}_H}{d\theta}$ and $\frac{d\tilde{\varphi}_{RD}}{d\theta}$.

Recall from (16) that $\tilde{\varphi}_H(\varphi^*, \varphi_{RD}^*) = \left(\frac{\int_{\varphi^*}^{\varphi_{RD}^*} \varphi^{\sigma-1} g(\varphi) d\varphi}{G(\varphi_{RD}^*) - G(\varphi^*)} \right)^{\frac{1}{\sigma-1}}$. Hence, we can write

$$X_H \hat{\tilde{\varphi}}_H = \left(\varphi_{RD}^*{}^{\sigma-1} - \frac{\int_{\varphi^*}^{\varphi_{RD}^*} \varphi^{\sigma-1} g(\varphi) d\varphi}{G(\varphi_{RD}^*) - G(\varphi^*)} \right) dG(\varphi_{RD}^*) - \left(\varphi^*{}^{\sigma-1} - \frac{\int_{\varphi^*}^{\varphi_{RD}^*} \varphi^{\sigma-1} g(\varphi) d\varphi}{G(\varphi_{RD}^*) - G(\varphi^*)} \right) dG(\varphi^*)$$

with $X_H \equiv (\sigma - 1)\tilde{\varphi}_H^{\sigma-1} [G(\varphi_{RD}^*) - G(\varphi^*)] > 0$.

Applying the Pareto-distribution $G(\varphi) = 1 - \left(\frac{b_m}{\varphi}\right)^\alpha$ and $dG(\varphi) = \alpha \frac{b_m^\alpha}{\varphi^{\alpha+1}} d\varphi$ we get, after rearranging,

$$\frac{X_H}{\alpha} \hat{\tilde{\varphi}}_H = \frac{\alpha b_m^\alpha}{\sigma - \alpha - 1} \left(\frac{\varphi_{RD}^{*\alpha} \hat{\varphi}_{RD}^* - \varphi^{*\alpha} \hat{\varphi}^*}{\varphi_{RD}^{*\alpha} - \varphi^{*\alpha}} \right) (\varphi_{RD}^*{}^{\sigma-\alpha-1} - \varphi^*{}^{\sigma-\alpha-1})$$

³¹In the expression for $dM_{RD}/d\theta$ the term $M_{RD}\bar{f}_{RD}/L - 1 < 0$ since $L=R+M_{RD}\bar{f}_{RD}$.

$$+(\varphi_{RD}^* \sigma^{-\alpha-1} \hat{\varphi}_{RD}^* - \varphi^* \sigma^{-\alpha-1} \hat{\varphi}^*) b_m^\alpha .$$

Direct imitation case:

From (37) we know that

$$\hat{\varphi}_{RD}^* = \hat{\varphi}^* + X d\theta$$

with $X \equiv \left(\frac{\zeta}{(1-\theta)[1-(1-\zeta)\theta]} \right) \left(1 - \left(\frac{\varphi^*}{\varphi_{RD}^*} \right)^{\sigma-1} \right) > 0$ as before. Hence,

$$\frac{X_H}{\alpha} \hat{\varphi}_H = \frac{\alpha b_m^\alpha}{\sigma - \alpha - 1} \left(\hat{\varphi}^* + \frac{\varphi_{RD}^{*\alpha} X d\theta}{\varphi_{RD}^{*\alpha} - \varphi^{*\alpha}} \right) (\varphi_{RD}^{*\sigma-\alpha-1} - \varphi^{*\sigma-\alpha-1})$$

$$+(\varphi_{RD}^{*\sigma-\alpha-1} - \varphi^{*\sigma-\alpha-1}) b_m^\alpha \hat{\varphi}^* + b_m^\alpha \varphi_{RD}^{*\sigma-\alpha-1} X d\theta.$$

Collecting terms, we arrive at

$$\begin{aligned} \frac{X_H}{\alpha} \hat{\varphi}_H &= \frac{b_m^\alpha (\sigma - 1)}{\sigma - \alpha - 1} (\varphi_{RD}^{*\sigma-\alpha-1} - \varphi^{*\sigma-\alpha-1}) \hat{\varphi}^* \\ &\quad - b_m^\alpha \varphi_{RD}^{*\sigma-\alpha-1} \left[\frac{\alpha}{1 - \sigma + \alpha} \frac{1 - (\varphi_{RD}^*/\varphi^*)^{1-\sigma+\alpha}}{1 - (\varphi_{RD}^*/\varphi^*)^\alpha} - 1 \right] X d\theta. \end{aligned}$$

In this equation the first term on the RHS is the effect of the proportional change of φ^*

and φ_{RD}^* , the shift effect. The second term on the RHS marks the dilation effect. Since $\sigma - \alpha - 1 < 0$ and $\varphi_{RD}^* > \varphi^*$ the shift effect is positive. The dilation effect is ambiguous. It is negative once we restrict $\alpha > 1 - \sigma + \alpha > 0$ such that $(\varphi_{RD}^*/\varphi^*)^{1-\sigma+\alpha} < (\varphi_{RD}^*/\varphi^*)^\alpha < 1$. It is positive when $[\alpha(1 - (\varphi_{RD}^*/\varphi^*)^{1-\sigma+\alpha}) < (1 - \sigma + \alpha)(1 - (\varphi_{RD}^*/\varphi^*)^\alpha)]$, which constitutes a sufficient condition for $\hat{\varphi}_H > 0$.

Similarly, we can derive that $\hat{\varphi}_{RD}/d\theta > 0$. Recalling that

$$\tilde{\varphi}_{RD}(\varphi_{RD}^*) = \left(\frac{1}{1-G(\varphi_{RD}^*)} \int_{\varphi_{RD}^*}^{\infty} \varphi^{\sigma-1} g(\varphi) d\varphi \right)^{\frac{1}{\sigma-1}}, \text{ we can write:}$$

$$X_{RD}\hat{\varphi}_{RD} = \left(\frac{\int_{\varphi_{RD}^*}^{\infty} \varphi^{\sigma-1} g(\varphi) d\varphi}{1 - G(\varphi_{RD}^*)} - \varphi_{RD}^*{}^{\sigma-1} \right) dG(\varphi_{RD}^*)$$

with $X_{RD} \equiv (\sigma - 1)\tilde{\varphi}_{RD}^{\sigma-1} [1 - G(\varphi_{RD}^*)] > 0$. Applying the Pareto-distribution and rearranging gives:

$$X_{RD}\hat{\varphi}_H = -\frac{\alpha b_m^\alpha (\sigma - 1)}{\sigma - \alpha - 1} \varphi_{RD}^*{}^{\sigma-\alpha-1} \hat{\varphi}_{RD}^*.$$

Indirect imitation case:

From (41) we know that

$$\hat{\varphi}_{RD}^* = \hat{\varphi}^* + X d\theta$$

with $X \equiv \left(\frac{\bar{f}_{RD}/f}{(1-\zeta)(\sigma-1)} \right) > 0$. Hence,

$$\frac{X_H}{\alpha} \hat{\varphi}_H = \frac{\alpha b_m^\alpha}{\sigma - \alpha - 1} \left(\hat{\varphi}^* + \frac{\varphi_{RD}^*{}^{-\alpha} X d\theta}{\varphi_{RD}^*{}^{-\alpha} - \varphi^*{}^{-\alpha}} \right) (\varphi_{RD}^*{}^{\sigma-\alpha-1} - \varphi^*{}^{\sigma-\alpha-1})$$

$$+ (\varphi_{RD}^*{}^{\sigma-\alpha-1} - \varphi^*{}^{\sigma-\alpha-1}) b_m^\alpha \hat{\varphi}^* + b_m^\alpha \varphi_{RD}^*{}^{\sigma-\alpha-1} X d\theta.$$

This expression is exactly the same as in the direct imitation case, implying that all results there carry over to the indirect imitation case. This also applies to the conditions we derived there that would ascertain $\hat{\varphi}_H > 0$.

Similarly, we can derive that $\hat{\varphi}_{RD}/d\theta > 0$:

$$X_{RD}\hat{\varphi}_{RD} = \left(\frac{\int_{\varphi_{RD}^*}^{\infty} \varphi^{\sigma-1} g(\varphi) d\varphi}{1 - G(\varphi_{RD}^*)} - \varphi_{RD}^*{}^{\sigma-1} \right) dG(\varphi_{RD}^*)$$

with $X_{RD} \equiv (\sigma - 1)\tilde{\varphi}_{RD}^{\sigma-1} [1 - G(\varphi_{RD}^*)] > 0$. Applying the Pareto-distribution and rearranging gives:

$$X_{RD}\hat{\varphi}_H = -\frac{\alpha b_m^\alpha (\sigma - 1)}{\sigma - \alpha - 1} \varphi_{RD}^*{}^{\sigma-\alpha-1} \hat{\varphi}_{RD}^*.$$