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)

Preliminary version prepared for presentation at the ETSG 2012 meetings in Leuven.
--- Please do not quote.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

¹ Address: GIGA Institute of Asian Studies, Rothenbaumchaussee 32, 20148 Hamburg, Germany. E-mail: haasnoot@giga-hamburg.de.
² Corresponding 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

The geographical clustering of industrial activity has received a lot of academic attention. Research on this topic comprises economic, business and geography literatures, including detailed accounts on determinants of specific success stories such as Silicon Valley and the Detroit automobile industry (e.g. Saxenian, 1994; Klepper, 2010), more general advancements that try to establish the broader economic and geographical reasons for clustering (Fujita et al., 1999; Brakman et al., 2009), 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 and vice versa (e.g. Wennberg & Lindqvist, 2010; Molina-Morales and Martínez-Fernández, 2009). 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 vast interest in geographical clustering dates back to Marshall (1920), who argued that geographical concentration of economic activity leads to pecuniary and technological externalities that will benefit individual firms and the industry they are part of. There is evidence that firms indeed benefit from being part of a cluster, though that is not always true (Wennberg & Lindqvist, 2010). Economic clustering may also involve negative externalities for firms. The literature has forwarded reasons for diseconomies of agglomeration that may come when the cluster ages or when the clusters becomes too dense (e.g. Pouder and St. John, 1996; Molina-Morales and Martínez-Fernández, 2009; Boschma, 2005; McCann & Folta, 2009). Nevertheless, the vast majority of industries are concentrated in space, indicating that it pays off for firms to cluster. However, Ellison and Glaeser (1997) conclude for the U.S. that ‘while there are a number of industries that look like Silicon Valley of the auto industry, it is much more common for industries to be only slightly concentrated” (p. 891). Apparently there are positive net effects of clustering, but clearly these should not be overestimated. Furthermore, industries tend to coagglomerate (Ellison et al., 2010), giving rise to clusters of economic activity in a broader sense than just the industry level. As such, the benefits of geographical clustering clearly transcend to the region they are part of. Patterns have also been discerned for the EU (e.g. Combes and Overman, 2004). Clustering of economic activity is rule rather than exception and can be discerned at different levels of aggregation, including the regional level (Krugman, 1991; Brakman et al., 2009).
We argue that geographical clustering may have an unexpected negative effect for regional output and growth through an inefficient composition of firms in the cluster. By generating positive agglomeration effects for firms, a neglected externality of clustering is that it helps the least productive firms to survive. Firms differ by their intrinsic productivity, leading some firms to be larger and more profitable than others in the industry. Clustering will be relatively more beneficial for those firms at the low end of the productivity distribution, lowering the productivity threshold for firms to survive. This will reduce the overall level of efficiency in the economy: clustering might be individually rational but collectively irrational. This argument has its antecedents in the seminal paper by Melitz (2003), who provides a formal framework for studying the effects of productivity heterogeneity amongst firms in a model of international trade. Baldwin and Okubo (2006) use a similar argument 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).

To make this neglected externality explicit we set up a model of heterogeneous firms à la Melitz (2003), incorporating knowledge spillovers as a proxy for the geographical proximity of firms. Knowledge spillovers are seen as one of the Marshallian raisons d’être of clusters. Note therefore that we do not explicitly model geographical proximity, but instead choose to model knowledge spillovers in a Melitz-type heterogeneous firm framework. By varying the extent of spillovers we proxy differences in geographical proximity within the cluster. This helps us make our main point without entangling ourselves in the analytical complexity of Krugman-like models of geographical economics. Our main interest is to unravel a neglected externality due to the heterogeneity of firms within clusters and not the determinants of clustering itself. For our purposes, clustering must be exogenous rather than endogenous.

We assume that firms may invest in R&D as a way to reduce the risk of leaving the market. In innovative environments, firms need to constantly rethink their strategy and adapt to their surroundings in order to remain competitive. Our approach is that R&D is needed to increase the probability of surviving. In particular, we assume $\zeta > 0$ as some form of strong systematic

---

3 This is based on the work by Jaffe and Trajtenberg (2002), who looked at patent citations and geographical proximity and found that patents are more likely to build upon previous patents if they were filed near to each other.
risk comparable to the exogenous chance of firm exit $\delta$ in the Melitz model. In addition, we assume a firm-specific risk $\varepsilon > 0$ that can be reduced by doing R&D. This can be seen as the constant innovation that is required for a firm to remain a strong player in the market. The benefits of R&D in this respect depend on the firm’s productivity level. Firms that draw a high productivity level have a stronger incentive to stay in the market than firms drawing a low productivity level. R&D is a binary choice: either a firm invests or it does not invest.

The outcome is that only the most productive firms will invest in R&D and that their increased chance of survival in the market has a positive effect on the average productivity of firms. This outcome would reflect a situation without knowledge spillovers at all, i.e. a situation where firms are not geographically proximate.

Having established this baseline result, we then assume clustering of firms by allowing the benefits from firms’ R&D investments to spill over to other firms. We model this in two ways: either the benefits go directly to all other firms in the cluster or the benefits reach other firms more indirectly, by making R&D investments cheaper. The effects on innovation are different in each of these cases: direct spillovers reduce innovation while indirect spillovers increase it. However, the effect on average productivity within the region is clearly negative, irrespective of the particular way spillovers take place. Proximity helps less efficient firms to survive, reducing the efficiency of the cluster at large.

Geographical clustering thus gives rise to a negative externality that has been neglected by the literature so far. While having positive effects for individual firms, the effects on cluster performance are negative. For governmental policy this implies that stimulating geographic clusters may have less positive effects than previously thought.

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. In innovative environments, firms need to constantly rethink their strategy and adapt to their surroundings. In this case, R&D does not so much enhance productivity as much as it enhances the probability of surviving. In the original Melitz model, the probability of survival is denoted by $\delta$ and is considered to be fixed and exogenous. In the light of the literature on firm survival (see, for example, Caves, 1998, Cefis and Marsili, 2005 and Pérez et al, 2004), this is a limited assumption. In the way it is modelled, $\delta$ only reflects an industry-specific risk. However, firms suffer not only from industry-specific risk, but also from firm-specific risk. To reflect this in the model, we assume that the chance on firm exit consists of two different and independent factors. 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 $\zeta$. Note that $\zeta$ is comparable to the exogenous chance of firm exit in the Melitz model. On the other hand, we assume a firm-specific risk $\epsilon$, 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 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). However, firms also face a risk to exit the market – the aforementioned exogenous systemic risk $\zeta$ and the endogenous firm-specific risk $\epsilon$. The consolidated chance of survival in the market thus becomes $(1 - \zeta)(1 - \epsilon)$, implying a chance of post-entry exit $\delta$ of:

$$\delta \equiv \zeta + \epsilon - \zeta \cdot \epsilon \quad (1)$$
We assume that firms can invest in R&D to lower their firm-specific risk. In the absence of time discounting, the pre-entry expected value of a firm is therefore

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

where \( \pi(\varphi) \) is a firm’s profit level, which depends on its (yet unknown) productivity level \( \varphi > 0 \) (see below), and where the subscripts in the denominators on the right-hand-side distinguish the chance of exit of firms that invest (subscript RD) and those that do not (subscript H).

The decision to invest depends on a comparison of the additional profit reached and the additional investment cost. Suppose that \( \varepsilon \) is a linearly declining function in R&D investments \( f_{RD} \), such that \( \varepsilon(0) = 1 \) and \( \varepsilon(\tilde{f}_{RD}) = 0 \):

\[
\varepsilon(f_{RD}) = \frac{\tilde{f}_{RD} - f_{RD}}{\tilde{f}_{RD}} \quad (0 < f_{RD} < \tilde{f}_{RD}).
\]

Assuming that the \( f_{RD} \) chosen has to be incurred each period, a firm invests if and only if:

\[
g(f_{RD}) = \pi(\varphi) - f_{RD} \geq 0.
\]

The optimal level of investment is determined by taking the derivative w.r.t. \( f_{RD} \). This yields (after rearranging):

\[
\frac{dg(f_{RD})}{df_{RD}} = \frac{(1 - \zeta)(\pi(\varphi) - \tilde{f}_{RD})/\tilde{f}_{RD} - \zeta}{[\zeta + \varepsilon(f_{RD}) - \zeta \varepsilon(f_{RD})]^2}
\]

which is positive or negative, depending on the sign of \( (1 - \zeta)(\pi(\varphi) - \tilde{f}_{RD})/\tilde{f}_{RD} - \zeta \). If it is negative, a firm will decide not to invest in R&D at all: \( f_{RD} = 0 \). If it is positive, a firm will decide to invest fully: \( f_{RD} = \tilde{f}_{RD} \). Hence, a firm will invest only if \( \pi(\varphi) \geq \tilde{f}_{RD}/(1 - \zeta) \).

Hence, investing in R&D is a binary choice: either a firm invests or it does not invest. If a firm invests in R&D, it will reduce firm-specific risk \( \varepsilon \) to zero. Consequently, firms that invest only face the systemic risk of exiting \( \zeta \). If a firm chooses not to invest in R&D, we assume the firm-specific risk becomes such that it will have to leave the market after one
period of (profitable) production: $\epsilon$ 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:

$$\delta_{RD} = \zeta \quad \text{and} \quad \delta_{H} = 1 \quad (2)$$

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.\(^4\)

Applying (2), the value function of the firm reduces to

$$v(\varphi) = \max \left\{ 0, \pi(\varphi), \frac{\pi(\varphi) - \tilde{f}_{RD}}{\zeta} \right\} \quad (3)$$

A firm can either choose to exit immediately, receiving nothing, or it can choose to stay in the market and produce. If it chooses to produce, it must then decide if it wishes to receive income for one period $\pi(\varphi)$ as a hype-follower or to invest in R&D and receive profits until it is forced to exit by a systemic shock. In this case, its value is $(\pi(\varphi) - \tilde{f}_{RD})/\zeta$.

Equation (3) defines two productivity cut-off points for market entry. The first cut-off point is the familiar Zero Profit Cut-off point $\varphi^*$ 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 $\tilde{f}_{RD} > 0$, any firm with productivity $\varphi \leq \varphi^*$ will have negative profits when investing. The second cut-off point defines a productivity level where the profits that are gained from investing in R&D are exactly equal to the fixed costs of R&D investment, and thus firms are indifferent between becoming a hype-follower or an innovator. As we derived earlier, this implies $\varphi_{RD}^* = \inf (\varphi | (1 - \zeta)\pi(\varphi) > \tilde{f}_{RD})$. In order for the investment to be productive, the discounted profit in each period has to outweigh the fixed costs associated with staying in the market. Furthermore, as argued, $\varphi_{RD}^* > \varphi^*$.

\(^4\) Our terminology is based on Gollotto 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.
The profits a firm derives from its operations is 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 \( \omega \):

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

(4)

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

(5)

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

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

(6)

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 firm’s productivity level \( \varphi > 0 \):

\[
l = f + q/\varphi 
\]

(7)

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

\[
\pi(\varphi) = q(\varphi)p(\varphi) - w\left(\frac{q}{\varphi} + f\right) 
\]

(8)

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:

\[ p(\varphi) = \frac{1}{\rho \varphi} \]  

(9)

\[ q(\varphi) = R P^{\sigma-1} [\rho \varphi]^\sigma \]  

(10)

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

\[ \pi(\varphi) = \frac{r(\varphi)}{\sigma} - f = \frac{R(P\rho \varphi)^{\sigma-1}}{\sigma} - f \]  

(11)

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:

\[ \frac{r(\varphi_1)}{r(\varphi_2)} = \left( \frac{\varphi_1}{\varphi_2} \right)^{\sigma-1} \]  

(12)

\[ \pi(\varphi_2) = \left( \frac{\varphi_2}{\varphi_1} \right)^{\sigma-1} \frac{r(\varphi_1)}{\sigma} - f \]  

(13)

Consequently, we can write the cut-off point for R&D firms \( \varphi_{RD}^* \) relative to the cut off point for profitable entry \( \varphi^* \):

\[ (1 - \zeta) f \left( \left( \frac{\varphi_{RD}^*}{\varphi^*} \right)^{\sigma-1} - 1 \right) = \tilde{f}_{RD} \]  

(14)

where we applied (13), 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. These \( M \) 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 \( p_{rd} = 1 - \)
Taking into account that the distribution changes due to the exit of firms, the ex post probability distributions of productivities become:

\[
\mu(\varphi_H) = \frac{g(\varphi)}{G(\varphi_{RD}) - G(\varphi^*)} \quad \text{and} \quad \mu(\varphi_{RD}) = \frac{g(\varphi)}{1 - G(\varphi_{RD})}
\]  \hfill (15)

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

\[
\bar{\varphi}_H(\varphi^*, \varphi_{RD}) = \left( \frac{1}{G(\varphi_{RD}) - G(\varphi^*)} \int_{\varphi^*}^{\varphi_{RD}} \varphi^{\sigma-1} g(\varphi) d\varphi \right)^{\frac{1}{\sigma-1}}
\]  \hfill (16)

\[
\bar{\varphi}_{RD}(\varphi_{RD}) = \left( \frac{1}{1 - G(\varphi_{RD})} \int_{\varphi_{RD}}^{\infty} \varphi^{\sigma-1} g(\varphi) d\varphi \right)^{\frac{1}{\sigma-1}}
\]  \hfill (17)

and, consequently, average overall profits are

\[
\bar{\pi}_H = \pi(\bar{\varphi}_H) = \left( \left( \frac{\bar{\varphi}_H}{\varphi^*} \right)^{\sigma-1} - 1 \right) f
\]  \hfill (18)

\[
\bar{\pi}_{RD} = \pi(\bar{\varphi}_{RD}) - f_{RD} = \left( \left( \frac{\bar{\varphi}_{RD}}{\varphi^*} \right)^{\sigma-1} - 1 \right) f - \bar{f}_{RD}
\]  \hfill (19)

where we have applied (11) 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 (18) and (19) establish an equilibrium relationship between average profits and the cut-off productivity level of profitable entry \( \varphi^* \). As in the original Melitz model, these are downward sloping curves. Furthermore, in equilibrium, the expected value of entering the market must be zero:

\[
v_e = (p_e - p_{rd})\bar{\pi}_H + p_{rd}\bar{\pi}_{RD}
\]

\[
= \left( G(\varphi_{RD}) - G(\varphi^*) \right) \bar{\pi}_H + \left( 1 - G(\varphi_{RD}^*) \right) \frac{1}{\zeta} \bar{\pi}_{RD} - f_e = 0
\]  \hfill (20)
As in Melitz, we could use this formula to calculate average profits and generate a free entry condition. However, because we have two types of firms, this leads to more complicated expressions:  

\[ \bar{\pi}_H = \frac{f_e - (1 - G(\varphi^*) \frac{1}{\zeta} \left( f \left( \left( \frac{\varphi_{RD}}{\varphi_H} \right)^{\sigma-1} - 1 \right) \left( \frac{\varphi_H}{\varphi^*} \right)^{\sigma-1} \right) - \bar{f}_{RD} \right)}{(G(\varphi^*_{RD}) - G(\varphi^*)) + (1 - G(\varphi^*_RD)) \frac{1}{\zeta}} \]  

(21)

\[ \bar{\pi}_{RD} = \frac{f_e + (G(\varphi^*_{RD}) - G(\varphi^*)) \left( f \left( \left( \frac{\varphi_{RD}}{\varphi_H} \right)^{\sigma-1} - 1 \right) \left( \frac{\varphi_H}{\varphi^*} \right)^{\sigma-1} \right) - \bar{f}_{RD} \right)}{(G(\varphi^*_{RD}) - G(\varphi^*)) + (1 - G(\varphi^*_RD)) \frac{1}{\zeta}} \]  

(22)

These are very unwieldy free entry conditions, which only serve a purpose of proving existence of equilibrium. However, such can also be accomplished as follows. Suppose first that firms are not aware of the effect of innovation before entering the market, assuming that they will simply enter and produce for one period before leaving. This reduces the free entry condition to

\[ \bar{\pi}_e = \frac{f_e}{1 - G(\varphi^*)} \]  

(23)

with \( \bar{\pi}_e \) denoting expected lifetime profits. Combined with (18), this leads to a unique equilibrium as in Melitz, and mathematically speaking the analysis is almost exactly the same. Note, however, that the fact that firms are unaware of post-entry options affects the equilibrium in a serious way. Real profits are always higher than expected profits, so we would expect under-entry and an equilibrium with a lower mass of firms, amount of varieties, competition and consumer utility than otherwise. However, we can show that we can safely drop this assumption. To do this, we take the other extreme. What if firms think they will all

---

5To derive these average profit levels rewrite (19) to \( \bar{\pi}_{RD} = \left( \left( \frac{\varphi_{RD}}{\varphi_H} \right)^{\sigma-1} \left( \frac{\varphi_H}{\varphi^*} \right)^{\sigma-1} - 1 \right) f - \bar{f}_{RD} \), so that \( \bar{\pi}_{RD} = \bar{\pi}_H + f \left( \left( \frac{\varphi_{RD}}{\varphi_H} \right)^{\sigma-1} - 1 \right) \left( \frac{\varphi_H}{\varphi^*} \right)^{\sigma-1} - \bar{f}_{RD} \). Substituting this in the expression for \( \nu_e \) leads, after rearranging to the expression for \( \bar{\pi}_H \) in (21). Equation (22) is obtained by substituting this in the modified expression for \( \bar{\pi}_{RD} \).
become sustainable innovators post-entry? That would lead to the following free entry condition:

\[ \bar{\pi}_e = \frac{\zeta (f_e + \bar{f}_{RD})}{1 - G(\varphi^*)} \]  

(24)

As before, it is clear that the difference with Melitz’s work is minor and that this equation will combine with the earlier mentioned Zero Profit Condition to form a unique equilibrium. In this case, however, it will lead to over-entry.

In reality, neither of these two free entry conditions will be the one used by entering firms. Rather, it will be a combination of the two. Because the two individual free entry conditions enable an equilibrium, any simple percentile combination of the two must also lead to an equilibrium.\(^6\)

To get the equilibrium values both cut-off points \( \varphi^* \) and \( \varphi_{RD}^* \), we set (20) to zero and use (18)-(19) to obtain \( (G(\varphi_{RD}^*) - G(\varphi^*)) \left( \left( \frac{\varphi_{RD}}{\varphi^*} \right)^{\sigma-1} - 1 \right) f + (1 - G(\varphi_{RD}^*)) \frac{1}{\zeta} \left( \left( \frac{\varphi_{RD}}{\varphi^*} \right)^{\sigma-1} - 1 \right) f = f_{RD} = f_e \). Applying the expressions for average productivity (16) and (17) this reduces to

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

\(^6\) Admittedly, it is not as simple as this. \( G(\varphi^*) \) is different between the two conditions because of the extra costs of R&D investment. Entrants that know of both types of firms will also know that they will not become a hype follower if they draw a high productivity, so the correct term in (18) is \( (G(\varphi_{RD}^*) - G(\varphi^*)) \). However, none of these changes will disable a possible equilibrium, and our main point here is to prove that such an equilibrium exists.
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 point in our analysis.

What is left is to determine the equilibrium mass of entrants into the industry each period. Each period, $\zeta$ 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

$$p_e M_e = \zeta M_{RD} + M_H \quad (25)$$

Furthermore, the division across types of firms must remain constant in steady state. Because $M_{RD} + M_H = M$, deriving the mass of innovators amounts to deriving their percentage. Defining $P_{rd} \equiv \frac{p_{rd}}{p_e}$ as the probability of becoming an innovator firm after successful entry, the percentage of innovator firms is

$$M_{RD} = \left( \frac{P_{rd}}{P_{rd} + (1 - P_{rd}) \zeta} \right) M \quad \text{and} \quad M_H = M - M_{RD} \quad (26)$$

If this condition is satisfied, the percentages of the two types of firms are stable over time. Each combination of $P_{rd}$ and $\zeta$ gives rise to a single unique equilibrium $M_{RD}$, except for one corner case: if $P_{rd} = 0$ and $\zeta = 0$, the original distribution of firms will never change and the equilibrium fully depends on this distribution. Given the fact that innovation costs are not infinite, however, we already know that $P_{rd} \neq 0$, so this corner case is impossible and will be ignored henceforth.

We are now in the position to calculate average productivity, our main variable of interest in this paper. The main result is that the reduced chance of death for more productive firms begets a selection effect: in the equilibrium distribution of firms, the more productive firms are overrepresented compared to the productivity distribution of random entrants. We can

---

7 Equation (26) has been derived by applying that $M_{RD}$ and $M$ remain constant over time. Using that $M_{RD} + M_H = M$, each period $\zeta M_{RD} + M - M_{RD}$ firms are leaving the market, of which a percentage $P_{RD}$ re-enter as innovative firms. Equating this to the $\zeta M_{RD}$ innovator firms that are leaving leads to (26).
show this mathematically. If we remove $\varepsilon$ from the model so that it does not discriminate between innovators and hype followers, in essence reducing the model to the original Melitz model, we know that the mass of firms exiting and entering have the exact same productivity distribution. Consequently, we can write average productivity as $\bar{\phi} = (1 - P_{rd})\bar{\phi}_H + P_{rd}\bar{\phi}_{RD}$, which is equal to the average productivity of a random successful entrant. In each period, $\zeta$ firms, chosen randomly, exit the market, and $\frac{\xi}{P_e}$ firms enter the market to take their place. Because both exit and entry are random, the same average productivity persists and there is an equilibrium. Now let us suppose that this is also the average productivity 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 average productivity 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 an average productivity of $\bar{\phi}_{RD}$. By contrast, new entrants have the ‘normal’ average productivity $\bar{\phi} = (1 - P_{rd})\bar{\phi}_H + P_{rd}\bar{\phi}_{RD}$, which implies that average productivity in the second period can be described as $\bar{\phi} = (1 - \zeta)P_{rd}\bar{\phi}_{RD} + (1 - P_{rd} + \zeta P_{rd})((1 - P_{rd})\bar{\phi}_H + P_{rd}\bar{\phi}_{RD})$. This showcases the selection effect described before: innovator firms are the more productive firms, increasing average productivity in our example. However, in equilibrium average productivity must remain the same. Continuing this exercise until we arrive at an equilibrium, we can calculate the average productivity of our model:

$$\bar{\phi} = \left(1 - \frac{P_{rd}}{P_{rd} + (1 - P_{rd})\xi}\right)\bar{\phi}_H + \frac{P_{rd}}{P_{rd} + (1 - P_{rd})\xi}\bar{\phi}_{RD} \quad (27)$$

This equation describes $\bar{\phi}$ for all $1 \geq P_{rd} > 0$ and for all $1 \geq \zeta > 0$. There is only a problem if both $P_{rd}$ and $\xi$ are zero, in which case the starting distribution of firms will remain in place. 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. Since the actual distribution of productivities does not matter for the aggregate variables, this means that, as far as the aggregate variables are concerned, we can argue that the selection effect causes a new equilibrium where all firms have a higher productivity (Melitz, 2003: 1700).

The expressions for the aggregate variables such as prices, quantities, revenues and profits remain the same as in Meltiz (2003) and are repeated here for convenience.

\[
P = M^{1-\sigma} p(\bar{\phi}) \quad (28)
\]

\[
Q = M^{1/v} q(\bar{\phi}) \quad (29)
\]

\[
R = Mr(\bar{\phi}) \quad (30)
\]

\[
\Pi = M\pi(\bar{\phi}) \quad (31)
\]

At first sight, the increased productivity due to the selection effect implies that prices decline and that revenues, prices and quantities sold increase. However, this is not the entire story, as the mass of firms also changes. Following the same procedure as in Melitz (2003: 1704), we derive

\[
M = \frac{L}{\sigma(\pi(\bar{\phi}) + f) + \bar{r}_R \frac{P_{rd}}{Prd}\frac{P_{rd}}{(1-P_{rd})\zeta}} \quad (32)
\]

We note that the denominator of (32) 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 (32) and explains the second term in the denominator of (32).\(^8\)

Due to the selection effect, average productivity, and hence average revenues, increase. Because the term for R&D expenditures is also positive, the denominator unambiguously increases due to the selection effect, which then implies a smaller mass of firms. This presents

---

\(^8\) The free entry of firms implies that labor involved with the entry of firms must receive overall profits, which in our notation amounts to \(\Pi+\bar{f}_{RD}M_{RD}\). 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 (26) and (30), noting that \(r(\bar{\phi}) = \sigma(\pi(\bar{\phi}) + f)\), gives (32).
a counter-balancing effect to the increased average productivity. The end result is that the selection effect has an ambiguous effect on prices, quantities, profits and revenues.

3. Effects of clustering

The previous section has established groundwork for an analysis of clustering on average productivity. The consequence of clustering is that technological 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. We will forward two alternative ways of modeling these spillovers. First, we assume that firms are able to imitate other firms, thus being able to gain part of the technology researched by others at no cost. Second, we assume that firms are not able to directly copy another firm’s technology, but that spillovers imply that it lowers the cost of doing the research themselves.

As said, 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. 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 death does not only rely on one’s own decision, but also on the decisions of others. Clearly, the systemic risk $\zeta$ remains the same, but firms are now able to imitate some of the technology researched by others, so that the firm-specific risk is reduced for all other firms. Keeping firm-specific risks as before ($\varepsilon_H = 1, \varepsilon_{RD} = 0$), spillovers will only affect the chance of death for hype followers. Specifically, we assume that spillovers imply that hype-followers obtain a share $1 - \theta$ of the technology at no cost. Furthermore, we assume that $\theta$ is independent of the number of R&D firms, an assumption that will be relaxed later.

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

---

9 Demand is geographically concentrated in our set-up, which would be one reason for firms to enter the cluster.
\[
\delta_H = 1 - (1 - \zeta)(1 - \theta) \quad (33)
\]
\[
\delta_{RD} = \zeta \quad (34)
\]

We restrict the value of \(\theta\) to values between zero and one. If \(\theta = 1\), the resulting model would be exactly the same as in the previous section. If \(\theta = 0\), our model would become explicitly strategic: only one firm needs to invest for everyone to receive the spillover and the model turns into something like a public good game: each individual firm has an incentive not to invest and wait for someone else to do that. If \(0 < \theta < 1\), the strategy becomes a lot simpler. Because \(g(\varphi) \sim F(0, \infty)\), there is always a firm which invests for any \(\theta\).

This changes the value function into:

\[
u(\varphi) = \max \left\{ 0, \frac{1}{\zeta + \theta - \zeta \theta} \pi(\varphi), \frac{1}{\zeta} \pi(\varphi) - \frac{\tilde{f}_{RD}}{\zeta} \right\} \quad (35)\]

which allows us to calculate a new zero cutoff point for R&D investment. Following the same procedure as before, we get:

\[
\left(1 - \frac{\zeta}{\zeta + (1 - \zeta)\theta}\right) f\left(\left(\frac{\varphi_{RD}^*}{\varphi^*}\right)^{\sigma - 1} - 1\right) = \tilde{f}_{RD} \quad (36)
\]

Note that unless \(\theta\) or \(\zeta\) equal 1, \(\theta + (1 - \theta)\zeta < 1\). Hence, compared with the benchmark zero cutoff point derived in (12), the first term becomes smaller. Because the fixed overhead costs and fixed investment costs do not change, spillovers by imitation imply that the cutoff point for R&D investment, \(\varphi_{RD}^*\), increases. Because of knowledge spillovers, there is less incentive to invest compared to the benchmark model, ceteris paribus. As \(\theta\) becomes smaller 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.

Making use of the equilibrium condition that the mass of entering and exiting firms needs to have the same productivity distribution, the percentage of innovator firms in equilibrium is

\[
M_{RD} = \left(\frac{P_{rd} - (1 - \zeta)(1 - \theta)P_{rd}}{P_{ra} + (1 - P_{rd})\zeta - (1 - \zeta)(1 - \theta)P_{rd}}\right) M \quad (37)
\]
which is lower than in the benchmark model. Spillovers partly transfer the lower chance of
death to hype followers, reducing the selection effect. Furthermore, the probability of entering
as an innovator firm decreases due to the higher cutoff point for R&D investments. For the
hype followers we can say that the smaller the knowledge spillovers, the more important \( \zeta \)
becomes: if there are no knowledge spillovers (\( \theta = 1 \)), then hype followers exit the market at
the end of each period regardless of \( \zeta \). If, by contrast, spillovers are extremely large (\( \theta = 0 \)),
hype followers have the same chance of exit as sustainable innovators, \( \zeta \). This allows us to
make a prediction about knowledge spillovers and industry characteristics: the greater the
importance of R&D for the continuation of a firm, expressed by the difference between \( \zeta \) and
\( \varepsilon \), the greater the effect spillovers have on reducing the extra advantage that innovators get.

**Spillovers by indirect imitation**
We now assume that spillovers imply that firms are not able to imitate each others’
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: \( \delta_H \) and \( \delta_{RD} \) remain as in the benchmark model. Hence,
using \( 0 < \theta < 1 \) to parameterize the reduction in research costs, a firm’s value function
becomes:

\[
\nu(\varphi) = \max \left\{ 0, \pi(\varphi), \frac{1}{\zeta} \pi(\varphi) - \frac{\theta \bar{f}_{RD}}{\zeta} \right\}
\]  

(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:

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

(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 26) 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.

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 $\theta$ 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:

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

(40)
Despite the feedback loop, there is still only one equilibrium percentage of sustainable innovators for all possible configurations, which is stable\(^{10}\). 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 \( \zeta \) has a unique equilibrium, as it did before. The only real difference is that \( \theta \) has become endogenous. But as long as firms take \( \theta \) 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.

\[
P_{rd} = \frac{1 - G(\phi_{RD})}{1 - G(\phi^*)}
\]  

(41)

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 innovators\(^{11}\). 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 have\(^{12}\). We can be sure that \( G(\phi^*) \) does not change if the percentage of sustainable innovators changes, so we must only consider what happens to \( G(\phi_{RD}^*) \). What happens to this probability when the mass of firms change?

\(^{10}\) As \( M_{RD} \) goes up, the left-hand side always increases while the left-hand side always decreases. This implies that there is only one equilibrium.

\(^{11}\) As long as some firms engage in innovation, \( \zeta M_{RD} \) 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 \( \zeta - 1 \leq (\zeta - 1)M_{RD} \). Since \( \zeta - 1 \) is negative and \( 0 < M_{RD} < 1 \), we know that \( \zeta M_{RD} \) 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.

\(^{12}\) 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 less 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. In the final section, we allowed the size of the spillover to vary and found that this does not strongly change our results.

References

• 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.
• Krugman, P.R. (1991), Geography and Trade, MIT Press.


