Heterogeneous firms and cluster externalities: how asymmetric effects at the firm level affect cluster productivity

Haasnoot, C.W.; de Vaal, A.

DOI
10.1016/j.respol.2022.104529
Publication date
2022
Document Version
Final published version
Published in
Research Policy
License
CC BY

Link to publication

Citation for published version (APA):
Haasnoot, C. W., & de Vaal, A. (2022). Heterogeneous firms and cluster externalities: how asymmetric effects at the firm level affect cluster productivity. *Research Policy*, *51*(6), [104529]. https://doi.org/10.1016/j.respol.2022.104529

Research Policy 51 (2022) 104529

Contents lists available at ScienceDirect

Research Policy

journal homepage: www.elsevier.com/locate/respol

Heterogeneous firms and cluster externalities: how asymmetric effects at the firm level affect cluster productivity

Cornelis W. Haasnoot[a,#], Albert de Vaal[b,*]

[a] University of Amsterdam, Netherlands & GIGA German Institute of Global and Area Studies, Germany
[b] Institute for Management Research, Radboud University, The Netherlands

ARTICLE INFO

JEL-codes:
L11
O33
R11

Keywords:
Knowledge spillovers
Industry clustering
Government policy
Heterogeneous firms

ABSTRACT

When firms are heterogeneous, externalities within clusters can affect firms asymmetrically. These asymmetries at the firm level lead to a productivity effect at the cluster level that has been overlooked thus far. We develop a heterogeneous firm model where firms with different productivity levels decide how much to invest in market survival. With this model, we find a differentiation between high-productivity firms investing in market survival and low-productivity firms not investing in market survival. Cluster externalities alter the optimal market survival investment of firms, which in turn affects both cluster composition and cluster-level outcomes. By focusing on cluster productivity and assuming that cluster externalities take the form of knowledge spillovers, we find that the effect on the cluster depends on the particular type of knowledge spillovers. Using modelling outcomes and an extensive numerical simulation, we show that knowledge spillovers that reduce the cost of investment benefit investing, high-productivity firms and increase cluster productivity. By contrast, knowledge spillovers that imply that non-investing, low-productivity firms can free ride on the efforts of investing firms tend to reduce cluster productivity. We discuss ramifications for research on clusters and cluster policy, highlighting the importance of industry and knowledge spillover characteristics.

1. Introduction

The literature on externalities within industry clusters typically emphasizes the positive effects of externalities for individual firms within the cluster. Clustering helps firms through higher productivity or decreased costs (Marshall, 1920; McCann and Folta, 2008; Porter, 1990), higher sales (Visser, 1999), more innovation (Baptista and Swann, 1998), or better performance (Decarolis and Deeds, 1999). More recently, there has been an interest in understanding better whether firms benefit from cluster externalities differently. Empirical results indeed show that firms benefit asymmetrically from cluster externalities (e.g., Knoben et al., 2016; McCann and Folta, 2011; Rigby and Brown, 2015). However, it remains unclear whether clusters are in turn affected by firms benefitting asymmetrically. We argue that a negative cluster-level effect arises if weaker performing firms benefit the most from cluster externalities. In contrast, we argue that a positive cluster-level effect arises if stronger-performing firms benefit the most from cluster externalities. Essentially, we argue that cluster externalities affecting firms asymmetrically cause a feedback effect at the cluster level. This feedback effect has been overlooked in the literature thus far.

In this paper, we propose a mechanism that explains how cluster externalities affecting firms asymmetrically can impact cluster-level outcomes. The mechanism we propose revolves around the fact that firms' intrinsic productivity levels are heterogeneous, and such heterogeneity leads to differences in how firms benefit from cluster externalities. The foundation of our analysis is the Melitz (2003) model, where firms are heterogeneous in productivity. The context we use to illustrate this mechanism is an industry cluster where firms may innovate to enhance their market survival. The cluster externalities take the form of knowledge spillovers meaning that the knowledge a firm gains due to its innovation may spill over to other firms within the cluster. These knowledge spillovers may affect firms of different productivity differently, which begets a potentially negative feedback effect on cluster-level outcomes. We reveal how this feedback mechanism is related to firm asymmetries and how its impact on the cluster depends on the particular way knowledge spillovers take place. We also identify

* Corresponding author at: Radboud University, Department of Economics, P.O. Box 9108, 6500 HK Nijmegen, The Netherlands.
E-mail addresses: c.w.haasnoot@uva.nl (C.W. Haasnoot), albert.devaal@ru.nl (A. de Vaal).
URL: http://www.ru.nl/english/people/vaal-a-de/ (A. de Vaal).
Address: University of Amsterdam, Faculty of Economics and Business, P.O. Box 15867, 1001 NJ Amsterdam.

https://doi.org/10.1016/j.respol.2022.104529
Received 5 December 2019; Received in revised form 2 March 2022; Accepted 28 March 2022
Available online 15 April 2022

policy-relevant parameter configurations where our mechanism has a stronger effect.

In our model, firms differing in productivity produce imperfect substitutes of an otherwise identical product in a monopolistic competition setting. All firms face a fixed, exogenous systemic risk of market exit; but in contrast to Melitz (2003), all firms also face a firm-specific risk that can be controlled through investment. This way of modeling is grounded in (empirical) literature on innovation and firm survival, which emphasizes that firms need to invest to adapt to market changes (Caves, 1998; Cefis and Marsili, 2005; Pérez et al., 2004; Porter, 2000). In our model, a firm's benefits of investing for sustaining market presence depend on the firm's productivity level. To exemplify, firms with high productivity have a stronger incentive (as well as more funds) to invest in their market survival, as they receive higher profits each year they survive.[1] This implies that, in addition to the familiar Melitz-model productivity threshold for profitable entry, our model also determines a productivity threshold for market survival R&D. This additional threshold causes a differentiation in the cluster between high-productivity firms that invest in R&D (innovative firms) and low-productivity firms that do not (non-innovative firms). This differentiation implies that knowledge spillovers, when they occur, will affect firms asymmetrically. Accordingly, the firm composition of the cluster changes, affecting cluster-level outcomes. We focus on cluster productivity as cluster-level outcome, in line with Melitz (2003) and its progeny. Cluster productivity is defined as the average labor productivity of all firms in the cluster. Cluster productivity thus depends on the relative importance of innovative and non-innovative firms in the cluster (cluster composition) and the average productivity of these two groups of firms. Since both cluster composition and the productivity distribution within the groups of firms are endogenous, knowledge spillovers have a non-straightforward effect on cluster-level outcomes.

The cluster externalities in our set-up take the form of knowledge spillovers. The literature has identified many ways to model knowledge spillovers as well as ways in which firms might benefit from these spillovers. We will consider knowledge spillovers from R&D investments in firm survival rates. Moreover, we stylize the analysis by considering two extreme cases by which knowledge spillovers can materialize. First, we consider a situation where knowledge spillovers materialize through the direct imitation of investment practices. This situation results in a reduction of the firm-specific exit risk for firms that do not invest themselves. Second, we consider a situation where knowledge spillovers imply lower investment cost for firms. This situation favors firms that invest in market survival. These firms indirectly benefit from the research practices of their peers by having lower investment costs. We refer to the first situation as direct imitation and to the second situation as indirect imitation. This distinction aligns well with the literature arguing either that non-innovative firms benefit most since they have the most to learn (e.g., Alcácer and Chung, 2007; Amir and Wooders, 1999 & 2000; Ebert et al., 2019) or that innovative firms benefit the most because they have the absorptive capacity required to benefit from knowledge spillovers (Cohen and Levinthal, 1990; Grünfeld, 2003).[2] In reality, both situations may occur at the same time (Nooteboom, 2000; Nooteboom et al., 2007, Petruzzelli, 2011), a possibility we give due consideration in our simulation. Nonetheless, focusing on two extreme cases enhances tractability and allows us to demonstrate the new mechanism.

The juxtaposition of two distinct cases of knowledge spillovers allows us to find a novel effect on cluster-level outcomes. In our set-up, this effect occurs when there are externalities in clusters with firms of different productivity. With a standard Pareto-distribution of firm productivity levels, we come to analytical and clear-cut results for (1) the effects of knowledge spillovers on average productivity of innovative firms; (2) average productivity of non-innovative firms; and (3) the incidence of innovative firms in the cluster for either of the two extreme cases. We summarize these outcomes in several propositions, showing divergent results of both types of knowledge spillovers. To determine the effect of our mechanism on cluster-level productivity we use an extensive simulation. We analyze thousands of iterations of our model with different parameter values, within which we identify configurations matching country- and sector-specific data that we obtain from the OECD and the literature.

Our findings can be summarized as follows. First, when firms are heterogeneous in productivity, the type of knowledge spillovers will determine how the firm composition in the cluster will change. Knowledge spillovers that imply that non-innovative firms can imitate R&D practices of innovative firms will decrease the proportion of innovative firms in the cluster. Knowledge spillovers that imply a reduction of R&D costs, benefitting innovative firms, will increase the proportion of innovative firms. Second, knowledge spillovers affect the average productivity of innovative and non-innovative firms in the cluster. Knowledge spillovers benefitting non-innovative firms increase the average productivity of both types of firms. Knowledge spillovers that benefit innovative firms decrease the average productivity of innovative firms, but the impact on average productivity of non-innovative firms is unclear. Third, the effect of knowledge spillovers on overall cluster productivity is analytically unclear. However, our numerical results show that knowledge spillovers that benefit non-innovative firms will exert a negative effect on cluster productivity. Knowledge spillovers that benefit innovative firms will exert a positive cluster productivity effect. Furthermore, we find that if both types of knowledge spillovers are active at the same time, the type of spillover that is stronger tends to dominate. If the two types of spillovers are equally strong, the effect of these spillovers on cluster productivity is negative. This is explained by the reduction in average productivity of innovative firms becoming the main effect.

Though our paper is mainly theoretical, our analysis also has important implications for policy initiatives regarding clusters. First and foremost, our analysis shows that firm heterogeneity in a cluster may affect cluster outcomes by changing the composition of the cluster. We have shown this by considering the impact of knowledge spillovers on cluster productivity in a model where firm heterogeneity consists of productivity differences between firms. But the feedback mechanism we disclose will hold for any type of firm heterogeneity in clusters where externalities affect firms asymmetrically. For policy design this implies that policy makers should take into account the feedback mechanism when targeting a specific cluster outcome. Second, our analysis clearly shows that there is no one-size-fits-all cluster policy. We have shown that in the context of knowledge spillovers, understanding the type of

[1] We will refer to this type of investment as R&D, although we realize that this may be a too general term for the specific type of innovation we have in mind. In the terminology of the literature on product innovation our concept of innovation comes closest to what is known as 'incremental product and process changes'. In their review of definitions of product innovation, Garcia and Calantone (2002: 123) define incremental innovations as "products that provide new features, benefits, or improvements to the existing technology in the existing market," while De Bondt (1996: 3) argues that "These and other incremental changes thus typically consolidate and stretch life cycles of existing technologies and do not stimulate drastic changes in industry structure." An example is the Apple or Samsung business model, with a new version of the same product coming out almost every year. The other firms in the market who follow this then engage in what they describe as 'imitative innovation.' Our approach to R&D is also related to Grossman and Helpman's (1991) quality ladders: continuous progress is required to remain in the market, and those firms that do not climb to the next rung will inevitably fall off.

[2] Clearly, this literature refers to a more general form of innovation, and thus also to a different type of knowledge spillovers. However, papers that have analyzed imitation strategies that come close to the way we model knowledge spillovers also emphasize the importance of the technological capacity of the imitator (see, for example, Fischer, 1978; Schewe, 1996).

knowledge spillover is crucial to determining which type of cluster policy would be appropriate. Furthermore, our analysis shows that the cluster-level effects of knowledge spillovers also depend on the specific situation within the cluster (extent of spillovers, extent of firm heterogeneity, industry-specific exit risk, and so on). For policy makers this implies that they should consider specific measures targeting the aspects of the clusters they wish to address.

Our paper contributes to the scientific literature on clusters in four important ways. First, we contribute to the literature that considers the effect of firm heterogeneity on cluster-level outcomes. Acknowledging firm heterogeneity in research on industry clusters was named by McCann and Folta (2008) as a major area in which clustering research needs to move forward. The empirical literature has made significant steps since, though evidence is still mixed. Empirically, cluster externalities have been found to favor young firms (McCann and Folta, 2011), old firms (Rigby and Brown, 2015), small firms (Chung and Kalnins, 2001; Knoben et al., 2016; Shaver and Flyer, 2000), large firms (Drucker and Feser, 2012; Knoben et al., 2016), innovative firms (Cohen and Levinthal, 1990), and non-innovative firms (Ebert et al., 2019). The theoretical side of the literature, however, remains underdeveloped. By considering firm productivity differences as source of firm heterogeneity within clusters, we contribute to the literature by showing how this gives rise to a feedback effect at the cluster level. This feedback effect is relatively new to the literature. The literature on the clustering of industrial activity typically investigates how the composition of a cluster moderates the impact of cluster externalities on firm performance, taking the composition of a cluster as given (e.g., Boschma, 2005; Braunerhjelm and Feldman, 2006; McCann and Folta, 2008; Molina-Morales and Martínez-Fernández, 2009; Pouder and St. John, 1996; Wennberg and Lindqvist, 2010). Instead, we consider the impact of cluster externalities (knowledge spillovers) on the composition of clusters, analyzing how these externality-induced changes in cluster composition feed back into the cluster level.

Second, by formulating a micro-founded model where firms of different productivity decide on R&D investment within industrial clusters, we contribute to the (theoretical) literature on knowledge spillovers within clusters.[3] Specifically, we identify a potential asymmetry in how knowledge spillovers affect firms within clusters. This provides a theoretical underpinning to some of the literature on knowledge spillovers within clusters. An example of this is Nooteboom et al. (2007), who find that firms benefit the most from spillovers when they are either very close in cognitive distance, or very far away. Likewise, letting R&D affect the chance of market exit rather than the productivity of a firm is a novelty in the literature on Melitz-type models (cf. Atkeson & Burstein, 2010).

Third, our model has important consequences for the estimation of clustering benefits that have not been addressed in the literature thus far. Specifically, in order to measure the full firm-level benefits from clustering, we show that not only the effect of cluster externalities on firm productivity should be considered, but also that on firm survival. When there is an ex-ante adverse selection problem through geographical sorting, one may correct for this by using firm-level fixed effects (Gaubert, 2018; Henderson, 2003) or other methods such as GMM (Martin et al., 2011).[4] Our paper argues that, in addition to this *ex-ante* selection problem, there is also an *ex-post* selection problem, with less productive firms seeing a reduced chance of exit. As such, while firm-level fixed effects will help resolving the ex-ante selection problem, to help correct for the ex-post selection problem, future empirical research will have to include data on firms that exit the market. Beyond that, the results from our model inform the methodology that is used to determine the cluster-level effects of externalities. To find these effects, more needs to be done than only aggregating firm-level benefits. Because these externalities beget a change in exit rates for firms with different productivity levels, papers attempting to find cluster-level effects should also consider the different survival rates of firms within the cluster.

Fourth, this paper presents an alternative way to quantitatively analyze shocks. Melitz-type models are not always analytically solvable, so the numerical analysis we perform in this paper is not unusual. Most numerical analyses tend to limit themselves to a single scenario, as in Ghironi and Melitz (2005), or a small number of scenarios, as in Atkeson and Burstein (2010). In contrast, we consider a wider variety of scenarios. By incorporating a wider variety of scenarios, our numerical analysis helps quantify the effects of the contextual parameters of our model. This is useful for policy makers, as it helps them tailor cluster policy to their specific needs. Rather than prescribing a one-size-fits-all cluster policy, our numerical analysis prescribes context-specific policies. This does justice to the empirical literature that finds differential effects of clusters on firm performance (e.g., Rocha, 2004).

The structure of this paper is as follows: Section 2 describes the benchmark Melitz model with R&D investments and derives results on cluster-level outcomes in the absence of knowledge spillovers. Section 3 applies the model to industry clusters by including knowledge spillovers by direct imitation and knowledge spillovers by indirect imitation. Section 4 provides an extensive numerical analysis of the model to clarify the complex results of the model. Section 5 concludes and discusses policy implications.

2. The benchmark model

We adapt the standard Melitz (2003) model of heterogeneous firms to include firm-level investments to sustain their presence in the market. In this section, we explain and discuss how we include market survival R&D investments in the Melitz-model and what this implies for cluster composition. In the next section, we will use the model to investigate the consequences of asymmetric knowledge spillover benefits for cluster-level outcomes. A more detailed account of the model, its derivations, and equilibrium is included in Appendix A.

In the Melitz model, firms with different productivity levels produce distinct varieties of an otherwise identical product using only labor. Production features increasing returns to scale, modeled by a fixed overhead cost $f > 0$, and marginal costs that are inversely related to a firm's productivity level $\varphi > 0$. On the demand side, the Melitz-model features a Dixit-Stiglitz type of utility function with a constant elasticity of substitution between product varieties of $\sigma > 1$. Consumers maximize utility, which determines demand for each particular product variety as a function of prices and total demand for the product varieties in this industry. Profit maximization determines a firm's optimal price and quantity, where price is the familiar mark-up over marginal cost.

[3] Many models, such as Callois (2008) and Leppälä (2018) assume ex-ante homogeneity of firms. In the wake of the work of Melitz and Ottaviano (2008), a number of papers included ex-ante heterogeneity of firms. Examples of this include Atkeson and Burstein (2010), Baldwin and Robert-Nicoud (2008), Behrens et al. (2014) and Combes et al. (2012). Many of these papers identify a selection effect, where clustering causes a truncation in the distribution of productivities. However, the surviving firms in these models are always influenced symmetrically.

[4] This method does come at a cost, as the usage of firm-level fixed effects makes it impossible to measure the effect of time-invariant variables that influence the industrial structure. Martin et al. (2011) argue that the results from their GMM analysis are best interpreted as the short-run gains from agglomeration, and that this method is not the right one for assessing the long-run gains from agglomeration.

The inverse relation of marginal cost to firm productivity produces the standard outcome of the Melitz-model that more productive firms charge lower prices, sell more output, and have higher profits.

A key issue in the Melitz-model is the entry and exit of firms in industry. Firms are ex ante uncertain about their productivity level and base their entry decision on a comparison of the one-time market entry costs and the expected profits of post-entry production. After entering, firms find out their actual productivity level and decide whether to stay in the market (in case of positive profits) or to exit without producing (in case of negative profits). Specifically, the Melitz-model determines a minimum productivity level φ^* for firms to have positive profits. However, even firms with positive profits may exit again: in the Melitz-model each firm faces a fixed and exogenous risk of being forced to exit the market.

Our main diversion from this basic set-up in the Melitz-model is that we endogenize the probability of firm exit by making it dependent on an optimal R&D decision at the firm level. Specifically, we model the chance of firm exit to consist of two separate and independent components. First, as in the Melitz-model, firms face a systemic risk: a risk of going out of business for reasons they cannot control. We parameterize this systemic risk by ζ. Second, in addition to the Melitz-model, we model a firm-specific risk ε, which can be controlled by doing R&D. The firm-specific risk signifies the continuous product innovation that is required for a firm to remain a relevant player in the market (e.g., Cefis and Marsili, 2005; Pérez et al., 2004).[5] Taking both risks of market exit together and assuming $0 \le \zeta \le 1$ and $0 \le \varepsilon \le 1$, the consolidated chance of survival in the market is $(1-\zeta)(1-\varepsilon) \le 1$. Put differently, the chance of post-entry exit is $\delta \equiv \zeta + \varepsilon - \zeta \cdot \varepsilon \ge 0$.

Firms can invest in R&D to lower their firm-specific risk. The decision to invest depends on a comparison between the additional profits from the longer life expectancy and the additional investment cost. Suppose that ε is a linearly declining function in R&D investments (f_{RD}), with $\varepsilon(0) = 1$ and $\varepsilon(\overline{f}_{RD}) = 0$:[6]

$$\varepsilon(f_{RD}) = \frac{\overline{f}_{RD} - f_{RD}}{\overline{f}_{RD}} \qquad (0 \le f_{RD} \le \overline{f}_{RD}).$$

We assume that staying in the market requires a continuous R&D effort, such that the chosen f_{RD} has to be incurred each period. In the absence of time discounting, this implies that the pre-entry expected value of a firm is a function of profits as a function of its productivity level, $\pi(\varphi)$, and the R&D investment cost f_{RD}:

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

A firm's optimal investment level is determined by taking the derivative of $v(\varphi)$ with respect to f_{RD}. Taking into account the specification of $\varepsilon(f_{RD})$, this yields:

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

The sign of this expression is independent of f_{RD}, implying that it is optimal for a firm to either fully invest in market survival or not to invest at all. If the numerator is negative, a firm will decide not to invest and $f_{RD} = 0$. If it is positive or zero (the latter by assumption), a firm will decide to invest fully: $f_{RD} = \overline{f}_{RD}$. This binary choice outcome of firms engaging in R&D is not uncommon, as noted in various accounts of R&D for the pharmaceutical industry (Grabowski and Vernon, 1992 & 1994; Khanna et al., 2015). In his paper on Chinese firm strategies, Zhou (2006) also makes a binary distinction between imitators and innovators.[7]

The expression in (1) also shows that the decision to invest clearly depends on a firm's productivity level. Particularly, it implies a minimum required productivity level for a firm to invest of $\varphi^*_{RD} = inf(\varphi|(1-\zeta)\pi(\varphi) > \overline{f}_{RD})$. Only the more profitable firms find it worthwhile to invest to stay in the market. We will use a subscript RD to distinguish firms that invest and a subscript H to distinguish firms that do not invest. For RD-firms the firm-specific risk ε will be reduced to zero ($\varepsilon_{RD} = 0$), making the systemic risk ζ the only exit risk they face. For H-firms, $\varepsilon_H = 1$ implying that these firms will have to leave the market after one period of profitable production. Accordingly, the consolidated chances of exit become:

$$\delta_{RD} = \zeta \quad \text{and} \quad \delta_H = 1 \tag{2}$$

Henceforth we will refer to firms that invest as innovative firms or innovators and to firms that enter the market to leave after one period as non-innovative firms or non-innovators. In line with other literature, non-innovative firms can also be seen as hype followers (Gollotto and Kim, 2003) or imitators (Malerba and Orsenigo, 2001, 2002).[8]

Each firm will compare its pre-entry expected value of non-entrance into the market ($v(\varphi) = 0$) to that of entering, either as a non-innovator ($v(\varphi) = \pi(\varphi)/\delta_H$) or as an innovator ($v(\varphi) = (\pi(\varphi) - \overline{f}_{RD})/\delta_{RD}$). Hence,

$$v(\varphi) = \max\left\{0,\ \frac{\pi(\varphi)}{\delta_H}, \frac{\pi(\varphi) - \overline{f}_{RD}}{\delta_{RD}}\right\} \tag{3}$$

Eq. (3) defines two productivity cut-off points for market entry. The first cut-off point is the Zero Profit Cut-off point φ^*, as in Melitz (2003). This cut-off point 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, and it implies that any firm with productivity level $\varphi \le \varphi^*$ will not enter the market. The second cut-off point is new and yields the productivity threshold φ^*_{RD} that we established above. This threshold marks the difference between non-innovative and innovative firms: below it, a firm finds it unprofitable to invest in market survival. Note that $\varphi^*_{RD} > \varphi^*$ must hold, as no firm will otherwise invest in R&D. Furthermore, the cut-off point for innovative firms φ^*_{RD} can be expressed in relation to the cut-off point for profitable entry φ^* (see Appendix A):

$$\frac{\delta_H - \delta_{RD}}{\delta_H} f\left(\left(\frac{\varphi^*_{RD}}{\varphi^*}\right)^{\sigma-1} - 1\right) = \overline{f}_{RD} \tag{4}$$

Eq. (4) governs how both cut-off points relate to each other as a

[5] Atkeson and Burstein (2010) provide a different way of incorporating product innovation into the Melitz model. They model product innovation as an R&D investment that creates new firms. This is in essence quite close to our implementation: both are a way for a firm to continue to exist. Ours is direct, as a specific firm sees its chance to exit reduce. Theirs is indirect, as a specific firm creates a new firm embodying their product.

[6] We use $\varepsilon(0) = 1$ and $\varepsilon(\overline{f}_{RD}) = 0$ to facilitate exposition, noting that our analytical results also hold for $0 < \varepsilon(\overline{f}_{RD}) < \varepsilon(0) < 1$. In the appendix we denote all results in this general form.

[7] The binary choice feature is not limited to our specific modelling of $\varepsilon(f_{RD})$. For instance, it also follows when using a quadratic specification with increasing returns: $\varepsilon(f_{RD}) = 1 - \left(\frac{1}{b}\right) f_{RD}{}^a / \overline{f}_{RD}$ for all b>0 and a>1. For decreasing returns ($0 < a < 1$), the quadratic specification yields $0 < f_{RD} < \overline{f}_{RD}$ as optimal outcome though. The reason is that the marginal cost of investing is constant, while the marginal benefit in terms of $\varepsilon(f_{RD})$ is increasing whenever $a > 1$ and decreasing in case $0 < a < 1$. When the marginal benefit is zero or negative ($a \le 0$), $f_{RD} = 0$ results.

[8] Gollotto and Kim (2003) argue 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. The Hype Cycle concept developed by Gartner, Inc. in 1995 also relates to this, see Fenn and Raskino (2008) or Järvenpää and Mäkinen (2008). Malerba and Orsenigo (2001, 2002), modelling the history of the pharmaceutical industry, talk of imitators and innovators. While innovators try to research new drugs, imitators only imitate the drugs already researched by others and do not perform any research themselves.

function of key parameters of the model. It will play a central role in determining how knowledge spillovers within clusters affect cluster composition and cluster productivity.

Equilibrium in the Melitz-model is governed by free entry and exit of firms in industry. As said, all firms are uncertain about their specific productivity level prior to entry: productivity levels are drawn from an ex-ante probability density function $g(\varphi)$ and associated cumulative distribution function $G(\varphi)$. In view of the cut-off points φ^* and φ^*_{RD}, this implies that the ex-ante probability of successful entry for firms is $p_e = 1 - G(\varphi^*)$ while that of entering as an innovator is $p_{RD} = 1 - G(\varphi^*_{RD})$.[9]

These probability distribution functions help firms to determine expected profit levels in the cluster, which they will use to make their entry decision. For a firm, the expected value of market entry must be zero in equilibrium:

$$v_e = (p_e - p_{RD})\tilde{v}_H + p_{RD}\tilde{v}_{RD} = \left(G(\varphi^*_{RD}) - G(\varphi^*)\right)\frac{\tilde{\pi}_H}{\delta_H} + \left(1 - G(\varphi^*_{RD})\right)\frac{\tilde{\pi}_{RD}}{\delta_{RD}} - f_e = 0 \tag{5}$$

where a '~' above a variable indicates an average value and where $f_e > 0$ denotes the fixed entry cost that each firm incurs when entering the market. In Appendix A we show how Eq. (5) establishes a free entry condition that relates the productivity thresholds φ^* and φ^*_{RD} to the ex-ante distribution functions, exit probabilities and the fixed costs of entry and operations. Together with Eq. (4), the free entry condition solves for the equilibrium values for the two cut-off points in our analysis.

The model is closed by determining the equilibrium mass of entrants into the industry cluster each period, M_e. Each period, a fraction δ_{RD} of innovators and a fraction δ_H of non-innovators leave the market. Denoting the number of innovative and non-innovative firms in the cluster by M_{RD} and M_H, respectively, steady-state equilibrium requires the mass of exiting firms $\delta_{RD}M_{RD} + \delta_H M_H$ to be equal to the mass of entering firms M_e. With an ex-ante probability of successful entry p_e, this implies

$$p_e M_e = \delta_{RD}M_{RD} + \delta_H M_H \tag{6}$$

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 innovative firm after successful entry, the share of innovators is[10]

$$M_{RD} \Big/ M = \left(\frac{\delta_H P_{RD}}{\delta_H P_{RD} + (1 - P_{RD})\delta_{RD}}\right) \quad \text{and} \quad M_H = M - M_{RD} \tag{7}$$

with $M = M_{RD} + M_H$ indicating the total number of producers in the cluster. If this condition is satisfied, the shares of the two types of firms are stable over time. Each combination of P_{RD}, δ_H and δ_{RD} gives rise to a single unique equilibrium M_{RD}. The main implication of (7) is that it shows that the reduced chance of exit for the more productive innovators begets a *selection effect*: in the equilibrium distribution of firms, there are more high-productivity innovative firms compared to a situation where the distribution of firms equals the distribution of random entrants.[11,12]

Finally, we turn to the main variable of interest in this paper: cluster productivity. Given the equilibrium distribution of firms, cluster productivity $\tilde{\varphi}$ is a weighted average of the average productivity levels of non-innovative and innovative firms:

$$\tilde{\varphi} = \left(1 - \frac{M_{RD}}{M}\right)\tilde{\varphi}_H + \frac{M_{RD}}{M}\tilde{\varphi}_{RD} \tag{8}$$

where the specific expressions for $\tilde{\varphi}_H$ and $\tilde{\varphi}_{RD}$ are given in Appendix A. Eqs. (8) and (7) together establish an important relation between cluster productivity and the probability of entering as an innovative firm:

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

The three terms on the right-hand-side of this expression highlight three distinctive effects that will be at the core of our analysis in later sections. The first term on the right-hand-side indicates a composition effect on cluster productivity, stemming from the increased probability of innovative firm entry. This effect is positive. A higher incidence of innovative firms is good for cluster productivity since innovative firms are more productive than non-innovative firms ($\tilde{\varphi}_{RD} > \tilde{\varphi}_H$). The second and third terms on the right-hand-side show the effects on cluster productivity due to group-level productivity effects. These are both negative. The most productive non-innovative firms will become innovators, lowering average productivity of the remaining group of non-innovators and of the (enlarged) group of innovators.

The relative importance of each of these effects depends primarily on the value of P_{RD}.[13] For P_{RD} approaching zero (one) the composition effect dominates as its weight becomes much greater (smaller) than one. Hence, the effect on cluster productivity of increased entry of innovative firms will be positive when P_{RD} is sufficiently small. This also implies that clusters will vary regarding the extent by which cluster productivity is affected by knowledge spillovers in clusters, an issue we turn to next.

3. Effects of knowledge spillovers

A well-established aspect of clusters is that knowledge spillovers occur.[14] In a cluster where a firm must innovate to sustain its presence in the industry, the benefits of innovation can spill over to non-innovative firms. This spillover will reduce the firm-specific exit risk for the non-innovative firms. We analyze the consequences of these knowledge spillovers by considering two ways in which firms can inadvertently benefit from the market survival R&D of other firms in the cluster.

[9] The *ex-post* probability distributions of productivities are different, since the distribution changes due to the exit of firms. Specifically, these become $\mu(\varphi_H) = \frac{g(\varphi)}{G(\varphi^*_{RD}) - G(\varphi^*)}$ and $\mu(\varphi_{RD}) = \frac{g(\varphi)}{1 - G(\varphi^*_{RD})}$.

[10] Eq. (7) is derived by imposing that in steady-state equilibrium M_{RD} and M are to remain constant over time. Using $M_{RD} + M_H = M$, 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 $\delta_{RD}M_{RD}$ innovative firms that leave leads to $M_{RD}/M = \delta_H P_{RD}/(\delta_H P_{RD} + (1 - P_{RD})\delta_{RD})$.

[11] If P_{RD} and ζ are both zero the starting distribution of firms remains in place, a possibility we exclude. If $= 1$ $P_{RD} = 1$, there is no difference between the two types of firms and the results are as in Melitz. However, our parameter choice always satisfies $0 < P_{RD},\ \zeta < 1$, which means that there is always a selection effect.

[12] The selection effect that we find is subtly different from selection effects found in previous Melitz models, such as Melitz (2003) or Behrens et al. (2014). In their models, the selection effect is an increase in the zero profit cut-off point, rather than a change in the mix of firms, as in our model.

[13] The relative amounts of M_H and M_{RD} do not matter as the terms in which they occur are both negative and of equal order of magnitude.

[14] Spillovers are an important aspect of the economics of clustering. Porter (2000) defines the boundaries of clusters by the extent of spillovers, and knowledge spillovers are seen as one of the Marshallian raisons d'être of clusters. Examples showcasing the importance of geographical proximity for knowledge spillovers are Jaffe et al. (1993), Audretsch and Feldman (1996), Zucker et al. (1998a), Jaffe and Trajtenberg (2002), Keller (2002), Asheim and Gertler (2005), and Figueiredo et al. (2015). Audretsch and Feldman (2004) give an overview of the literature.

However, we first consider the degree to which firms can use knowledge spillovers. For this, the degree to which the knowledge or skills of the innovative firm and the non-innovative firm are different is of prime importance. This degree of difference is called cognitive distance (Wuyts et al., 2005). When the cognitive distance is large, a firm may benefit from knowledge spillovers because the spillover presents a significant novelty compared to its own technology (Nooteboom et al., 2007). In that vein, knowledge from firms with weaker technology may not be helpful for firms with stronger technology (Amir and Wooders, 1999 & 2000). This perspective is consistent with empirical evidence suggesting that non-innovative firms benefit more from knowledge spillovers because they have the most to learn (Alcácer and Chung, 2007; Ebert et al., 2019; Pouder and St. John, 1996; Shaver and Flyer, 2000). Another perspective is that it may be beneficial for firms for cognitive distance to be small. In that case, firms are better able to use the knowledge accumulated by their rivals (Nooteboom et al., 2007). This is referred to as a firm's absorptive capacity (e.g., Cohen and Levinthal, 1989 & 1990; Findlay, 1978). In the literature, it is shown that firms need to invest in R&D themselves in order to be able to profit from spillovers, both theoretically (e.g., Grünfeld, 2003) and empirically (e.g., Cassiman and Veugelers, 2002 & 2006).

The two ways we model spillovers reflect two extreme implementations of these two perspectives. First, we consider a situation where knowledge spillovers occur by direct imitation. In this situation, knowledge spillovers take place through direct and costless imitation by non-innovative firms of the R&D practices of innovative firms. This imitation lowers the firm-specific exit rate for non-innovative firms. Firms are not required to invest into their own absorptive capacity, and firms that develop their own technology do not benefit from the work of others. This incentivizes firms to not innovate but instead piggyback on the investments done by other firms.

Second, we consider a situation of knowledge spillovers by indirect imitation. In this situation, knowledge spillovers only favor firms that invest in market survival by lowering R&D investment cost. In this case, firms need to invest in their own absorptive capacity in order to benefit from the spillover. This incentivizes firms to use the knowledge available in the cluster to generate their own innovation.

Even though both types of knowledge spillovers may occur simultaneously, focusing on these two extreme cases greatly improves tractability. Therefore, this section focuses on these two extremes. For completeness, we allow both types of knowledge spillovers to operate at the same time in the numerical simulations in Section 4.

In both cases we will parameterize the knowledge spillovers that occur by a fraction θ that lies between zero and one. This reflects the idea that knowledge spillovers will never be perfect.[15] This may be because of the aforementioned lack of absorptive capacity, but also because the technology available in the cluster does not match the firm's requirements. This is in line with other recent models that contain knowledge spillovers, such as Luttmer (2007) and Qiao et al. (2019). Knowledge spillovers are likely to occur to some degree, however. All firms are in the same industry and they only differ regarding the particular type of product they produce and their productivity level. Additionally, we assume no distance decay effects within the industry cluster for both cases of knowledge spillovers, meaning the gains from knowledge spillovers accrue to all firms in the cluster alike.

In our treatment of knowledge spillovers, we also assume that they are independent of the percentage of innovative firms in the cluster. There are two reasons for this assumption. Firstly, we are primarily interested in how (the extent of) knowledge spillovers affect cluster composition, not vice versa. Secondly, as argued in Vestal and Danneels (2018), it is not clear that increased inventive concentration in a cluster actually leads to more innovation. They find that in very innovative clusters, firms are more likely to source their knowledge from outside the cluster. This suggests that θ is not linearly increasing in the percentage of innovative firms, and may in fact decrease if the percentage of innovative firms becomes large enough. Therefore, we keep the extent of knowledge spillovers as an exogenous parameter.

Finally, we want to give attention to some possible alternative interpretations of our modeling set-up. One such interpretation is that knowledge spillovers cause a shirking problem. Indeed, there is a similarity between the results of our model and previous papers that identify a shirking or moral hazard problem, such as d'Aspremont et al. (1998). However, this interpretation only partly covers what the model entails. While one can view the results of knowledge spillovers by direct imitation through the lens of shirking, this cannot be done for knowledge spillovers by indirect imitation. In that case, firms may actually do the opposite of shirking: because they piggyback on others' investments, they actually invest themselves as well, doing more work. Furthermore, the choice by an individual firm not to innovate does not impact any other firm, as in our model knowledge is not a good that firms contribute to together.

Another interpretation is related to the seminal contribution of Cassiman and Veugelers (2002). They distinguish between *incoming spillovers* and *appropriability* when assessing the overall effects of spillovers. Incoming spillovers refer to the extent to which firms can profit from the knowledge of other firms. It can be increased by investing in absorptive capacity. Appropriability reflects the extent to which firms can reap the benefits of their own investments by controlling their outgoing knowledge flows (Cassiman and Veugelers, 2002: 1169). Applying their perspective to our framework, knowledge spillovers by indirect imitation represents a situation where appropriability is relatively high and firms must invest to increase their incoming spillovers. By contrast, knowledge spillovers by direct imitation represents a situation of low appropriability. Innovating firms cannot control the outward flow of knowledge and other firms can benefit from this without investing themselves. This provides an additional disincentive for firms to innovate. Knowledge spillovers then lead to a lower number of innovators in the cluster. This is similar to our interpretation in terms of the effects of spillovers, even though in our model firms do not take outgoing knowledge flows into account when deciding to innovate. Cassiman and Veugelers (2002) thus provide an interesting alternative perspective on the types of spillovers, leading to implications similar to ours.

3.1. Knowledge spillovers by direct imitation

Knowledge spillovers by direct imitation imply that the chance of market exit does not only rely on the firm's own R&D investments, but also on the R&D investments of other firms. 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 kind of modelling is similar to D'Aspremont & Jacquemin (1988), where firms directly benefit from the R&D activity of closely located firms (see Leppälä, 2018, and Qiao et al., 2019, for recent contributions in the same vein), or to models in which spillovers flow in one direction from the more innovative to the less innovative firms (e.g., Amir and Wooders, 1999 & 2000). This reduces the firm-specific risk for non-investing firms as well. Accordingly, we assume that in the presence of knowledge spillovers firm-specific risk becomes:

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

where $0 \le \theta < 1$ denotes the extent of knowledge spillovers. A value of $\theta = 0$ gives the benchmark case with no knowledge spillovers of Section 2.

The implication of knowledge spillovers by direct imitation is that only non-innovative firms will benefit. This is a logical consequence of

[15] Furthermore, with perfect knowledge spillovers no single firm will have an incentive to invest in R&D.

the R&D investment decision: a firm either invests fully, reducing the firm-specific market risk to zero, or it does not invest at all, becoming a non-innovative firm. For innovative firms knowledge spillovers are thus *de facto* irrelevant[16], while non-innovative firms witness a reduction of their firm-specific exit rate. Accordingly, the consolidated chances of exit become (for $\varepsilon(0) = 1$ and $\varepsilon(\bar{f}_{RD}) = 0$):

$$\delta_H = 1 - (1-\zeta)\theta \tag{9}$$

$$\delta_{RD} = \zeta \tag{10}$$

It is worth noting that while θ defines the extent of knowledge spillovers, the level of systemic exit risk ζ determines the extent by which non-innovators will profit from this. This makes sense as the level of ζ also constitutes what can be gained from innovation. A low ζ means that much can be gained from investing in R&D, implying that many firms will be inclined to invest in R&D, but also that knowledge spillovers yield a great benefit for non-innovators. Consequently, non-innovative firms stand to gain more from knowledge spillovers the lower ζ, ceteris paribus.

The value function remains as in the benchmark model, but the inclusion of θ in (9) defines new values for φ^* and φ^*_{RD}. Again, it holds that any firm with productivity level $\varphi \leq \varphi^*$ will also have negative profits when investing.[17] Note that the presence of knowledge spillovers does not affect φ^* directly. The reason for this is that the profit level of a firm not investing in R&D is not directly affected by knowledge spillovers. This is not the case for the cut-off point for R&D investment, which now also depends on θ. For $\varepsilon(0) = 1$ and $\varepsilon(\bar{f}_{RD}) = 0$ we get:[18]

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

However, the relation between the cut-off points for innovative firms φ^*_{RD} and profitable entry φ^* remains as in (4). Substituting the relevant values for δ_H and δ_{RD}, we get

$$\left(\frac{1-\theta}{1-(1-\zeta)\theta}\right)(1-\zeta)f\left(\left(\frac{\varphi^*_{RD}}{\varphi^*}\right)^{\sigma-1} - 1\right) = \bar{f}_{RD} \tag{11}$$

For $\theta = 0$ (no knowledge spillovers), this yields the same φ^*_{RD} as in the benchmark model. For $\theta = 1$ (perfect knowledge spillovers), no productivity level will be high enough to render market survival investment profitable. For any value $0 < \theta < 1$ the required φ for profitable R&D exceeds that of the benchmark model: the first term in (11) is smaller than in the benchmark model, cf. (4). Hence, knowledge spillovers by direct imitation imply that the cut-off point for R&D investment φ^*_{RD} increases. Because of knowledge spillovers, there are fewer incentives to invest compared to the benchmark model. As θ increases and knowledge spillovers become stronger, this effect on investment becomes stronger as well. Because all firms receive part of the technology, firms have less incentive to do research themselves.

However, to fully appreciate the effects of knowledge spillovers on φ^* and φ^*_{RD} we must also consider the effects that occur through changes in the aggregate variables. The impact of θ on φ^* is effectively determined by total differentiation of the free-entry condition, recognizing from (11) that changes in φ^* and φ^*_{RD} stand in a fixed relation:

$$\widehat{\varphi}^*_{RD} = \widehat{\varphi}^* + \left(\frac{\zeta}{(1-\theta)(\sigma-1)\delta_H}\right)\left(1 - \left(\frac{\varphi^*}{\varphi^*_{RD}}\right)^{\sigma-1}\right)d\theta \tag{12}$$

with a hat "^" denoting a proportional change, for instance $\widehat{\varphi}^*_{RD} = d\varphi^*_{RD}/\varphi^*_{RD}$. Using this in the total differentiation of the free entry condition and applying the Pareto-distribution[19] yields:

$$\widehat{\varphi}^* = -Zd\theta > 0$$

where $Z < 0$ is shorthand for an expression that is given in Appendix B.1.[20] Knowledge spillovers by direct imitation increase the threshold of profitable entry φ^*. When non-innovative firms can profit from the efforts of innovative firms, it becomes less attractive for firms to enter industry.

With $\widehat{\varphi}^* > 0$, it is immediate from (12) that also $\widehat{\varphi}^*_{RD} > 0$. In the presence of knowledge spillovers, the proportional change of the threshold for profitable entry as an innovator equals the effect on $\widehat{\varphi}^*$ that is due to the changes in aggregate variables (the first part of the equation) plus the direct effect knowledge spillovers have on $\widehat{\varphi}^*_{RD}$ (the second part of the equation). With $\varphi^* < \varphi^*_{RD}$, the direct impact of knowledge spillovers on $\widehat{\varphi}^*_{RD}$ is always positive, ensuring $\widehat{\varphi}^*_{RD} > \widehat{\varphi}^* > 0$. When non-innovative firms can profit from the efforts of innovative firms, it not only becomes less attractive for new firms to enter the industry, but it also becomes less attractive to innovate.

Proposition 1. (Threshold effects): When firm productivity levels within an industry cluster are Pareto-distributed, knowledge spillovers that benefit non-innovative firms (direct imitation) increase the productivity threshold for entering the industry as well as the productivity threshold for becoming an innovative firm.

The effects of these changes in φ^* and φ^*_{RD} on cluster productivity are not clear though. From (8), the change in cluster productivity can be written as:

$$d\widetilde{\varphi} = (\widetilde{\varphi}_{RD} - \widetilde{\varphi}_H)d(M_{RD}/M) + \left(1 - \frac{M_{RD}}{M}\right)d\widetilde{\varphi}_H + \frac{M_{RD}}{M}d\widetilde{\varphi}_{RD}. \tag{13}$$

This equation says that the overall effect on cluster productivity is the result of a change in the relative incidence of innovators in the cluster ($d(M_{RD}/M)$), an effect on the average productivity of non-innovators ($d\widetilde{\varphi}_H$) and an effect on the average productivity of innovators ($d\widetilde{\varphi}_{RD}$). We will refer to these effects as 'composition effect', 'non-innovator effect', and 'innovator effect', respectively.

With direct imitation, the cluster's composition M_{RD}/M is given by:

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

where we applied $\varepsilon(0) = 1$ and $\varepsilon(\bar{f}_{RD}) = 0$ for expositional reasons. Applying the Pareto-distribution so that $dP_{RD} = -\alpha XP_{RD}d\theta$, the composition effect becomes[21]

[16] Their optimal investment decision still implies $f_{RD} = \bar{f}_{RD}$. The relevant first-order-condition of optimal investment is $\mathrm{d}v(\varphi)/df_{RD} = [(1-\zeta)(1-\theta)(\pi(\varphi) - \bar{f}_{RD})/\bar{f}_{RD} - \zeta]/[\zeta + \varepsilon(f_{RD}) - \zeta\varepsilon(f_{RD})]^2 = 0$.

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

[18] The requirement on φ is consistent with the FOC on optimal investment: the FOC for optimal investment requires $\pi(\varphi) \geq [(1-\zeta) + \zeta/(1-\theta)]\bar{f}_{RD}/(1-\zeta)$ for a firm to invest.

[19] Although the Pareto-distribution is commonly used in models that deal with productivity differences between firms (e.g., Helpman et al., 2004), we acknowledge that the assumed distribution of productivities is not inconsequential for the results we derive. In view of the abundance of possible alternative distributions it is practically impossible, however, to check and report how the results will vary per distribution. Furthermore, the literature is unclear regarding which distribution reflects the distribution of firm productivities best, suggesting even that combinations of distributions may be warranted (Nigai, 2017). To avoid any misunderstanding, when presenting our results we therefore explicitly incorporate reference to the fact that a Pareto-distribution is assumed.

[20] We thank Jiangying Qiu for helping us prove that Z<0.

[21] In Appendix B.2 we show that the sign remains for the general case of $0 < \varepsilon(\bar{f}_{RD}) < \varepsilon(0) < 1$.

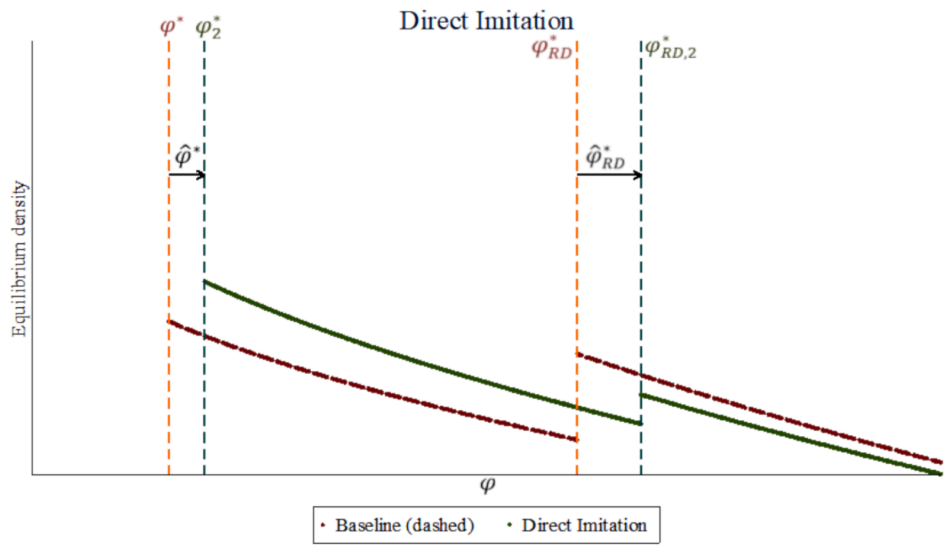


Fig. 1. The effect of spillovers on the composition of the cluster (direct imitation).

$$d\left(\frac{M_{RD}}{M}\right) = -\frac{\zeta P_{RD}[(1-(1-\zeta)\theta)\alpha X + (1-\zeta)(1-P_{RD})]}{[(1-P_{RD})\zeta + (1-(1-\zeta)\theta)P_{RD}]^2} d\theta < 0,$$

with $X \equiv \left(\frac{\zeta}{(1-\theta)(\sigma-1)\delta_H}\right)\left(1-\left(\frac{\varphi^*}{\varphi^*_{RD}}\right)^{\sigma-1}\right) > 0$. Hence, knowledge spillovers imply that the relative number of innovative firms in the cluster diminishes, exerting a negative composition effect on cluster productivity (since $\widetilde{\varphi}_{RD} > \widetilde{\varphi}_H$).

Proposition 2. (Composition effect): When firm productivity levels within an industry cluster are Pareto-distributed, knowledge spillovers that benefit non-innovative firms (direct imitation) reduce the relative number of innovative firms in the cluster. This decreases cluster productivity.

The composition effect can also be seen in the density functions, which we illustrate in Fig. 1. There are three effects at work here. Firstly, the increase in φ^* leads to a truncation on the left-hand side of the distribution, which increases the density of all other productivity levels in the cluster. Secondly, the increase in φ^*_{RD} causes the least productive of the innovative firms to stop innovating. The density of the distribution at these productivity levels then becomes smaller, as the selection effect starts to work against them. The third effect runs through the strength of the selection effect. Because knowledge spillovers reduce the difference between δ_H and δ_{RD}, the selection effect becomes weaker.

In Appendix D we show mathematically how the densities of various productivity levels change. Firms with a productivity level of $\varphi^* \leq \varphi < \varphi^*_2$ become unprofitable and exit the market, being removed from the distribution. Firms with a productivity level of $\varphi^*_2 \leq \varphi < \varphi^*_{RD}$ unambiguously increase in density: both the left-hand truncation of the distribution and the weakening of the selection effect increases their density. Firms with a productivity level of $\varphi^*_{RD} \leq \varphi < \varphi^*_{RD,2}$ see two effects: the truncation effect increases their density, while the fact that they switch to being non-innovators decreases their density. The overall effect cannot be determined mathematically, but our numerical simulations show that the density of these firms always decreases. Innovative firms, with a productivity level of $\varphi^*_{RD,2} \leq \varphi$, also see two effects: the truncation effect increases their density, while the weakening of the selection effect decreases their density. The simulation results show that the weakening of the selection effect is dominant for all realistic values of the various parameters and that the density of these firms decreases.[22] The changed cluster composition towards non-innovators makes the cluster less efficient.

We now turn to the effect of knowledge spillovers on productivity of the two types of firms. Regarding the non-innovator effect, there are multiple effects determining the impact on $d\widetilde{\varphi}_H$. One effect is a truncation of the distribution due to the increase in φ^*, which increases the average productivity of non-innovative firms, as it eliminates the least productive of them. The increase in φ^*_{RD} also has a positive effect on the average productivity of non-innovative firms, as the least productive innovators become the most productive non-innovative firms. In Appendix B.3 we prove that both the innovator and the non-innovator effect are positive:[23]

$$\widehat{\widetilde{\varphi}}_H = \widehat{\varphi}^* + \frac{X}{(\sigma-1)}\left[\frac{\alpha}{1-(\varphi^*_{RD}/\varphi^*)^{\alpha}} - \frac{\alpha-(\sigma-1)}{1-(\varphi^*_{RD}/\varphi^*)^{\alpha-(\sigma-1)}}\right] d\theta > 0,$$

$$\widehat{\widetilde{\varphi}}_{RD} = \widehat{\varphi}^*_{RD} > 0.$$

Proposition 3. (Non-innovator effect): When firm productivity levels within an industry cluster are Pareto-distributed, knowledge spillovers that benefit non-innovative firms (direct imitation) increase the average productivity of non-innovative firms in the cluster.

Proposition 4. (Innovator effect): When firm productivity levels within an industry cluster are Pareto-distributed, knowledge spillovers that benefit non-innovative firms (direct imitation) increase the average productivity of innovative firms in the cluster.

The effects on composition and average productivity effects in

[22] Running 15,000 iterations with extreme values only, we find one case where the effect on the density of innovative firms is positive. However, in this case the values of σ and α are unrealistically large (>800).

[23] Crucial to the proof we provide is that $\alpha > (\sigma - 1)$. This inequality is satisfied when the distribution of firm revenues has finite variance as in Helpman et al. (2004), requiring a shape parameter $\alpha - (\sigma - 1) > 2$. The inequality we impose is less strict for it does not affect analytical outcomes. It is met when there is a sufficiently high density of low productivities in the distribution (α not too low) and when these low productivities also get sufficient weight in calculating the sales weighted averages (σ not too high). In our simulations we will assume $\alpha > 1$, implying that the Pareto-distribution of productivities has finite mean.

Propositions 2-4 imply that the composite effect on cluster productivity is unclear. With the share of the more productive innovative firms diminishing, knowledge spillovers may very well imply that cluster productivity decreases. We will get back to this issue in Section 4, where we offer an extensive numerical analysis based on realistic values of the model's parameters. As we will see, knowledge spillovers that benefit non-innovative firms by direct imitation decrease cluster productivity for nearly all parameter configurations that we consider. Only extreme cases such as $\zeta \approx 1$, $\theta \approx 1$ or $\alpha \gg \sigma - 1$ make it likely that cluster productivity increases.

3.2. Knowledge spillovers by indirect imitation

We now consider the situation where firms can only benefit from knowledge spillovers if they invest in R&D. In contrast to the previous case, knowledge spillovers do not imply that firms are able to imitate the *outcome* of market survival R&D, but that firms can imitate the R&D *practices* done by others. We stylize this situation by assuming that knowledge spillovers imply a lower fixed investment cost of market survival R&D, benefitting firms that already innovate. This is somewhat similar to the way Luttmer (2007) models knowledge spillovers. In his model, newly entering firms pay an investment cost in order to attain the same productivity level of an existing firm, essentially copying that firm's innovative output. It is also analogous to Kamien and Zang (2000) and Grünfeld (2003). In their models, firms must invest in absorptive capacity in order to receive (more of) the knowledge spilling over from other firms in the market. We parameterize the reduction in research costs by $0 \leq \theta < 1$.

Knowledge spillovers by indirect imitation imply that the consolidated chance of exit for firms will remain exactly as in the benchmark model: $\delta_{RD} = \zeta$ and $\delta_H = 1$ when $\varepsilon(0) = 1$ and $\varepsilon(\overline{f}_{RD}) = 0$. Note that the extent of knowledge spillovers by indirect imitation θ does not affect firms' exit rates. This is because indirect knowledge spillovers work on the cost side of innovation and not on the benefit side. The benefits of R&D depend on ζ only, implying that when ζ is low, many firms will be inclined to invest in R&D irrespective of the cost. Consequently, when the costs of R&D decline due to knowledge spillovers, this will not affect the decision of those firms that already invest. But with lower R&D cost, some of the most productive non-innovative firms will start investing as well. More firms will do so when ζ is high, implying that a reduction in R&D cost has more impact on the degree of innovation when ζ is high.

In case of knowledge spillovers by indirect imitation, the firm's value function becomes:

$$v(\varphi) = \max\left\{0, \frac{\pi(\varphi)}{\delta_H}, \frac{\pi(\varphi) - (1-\theta)\overline{f}_{RD}}{\delta_{RD}}\right\} \tag{14}$$

showing how knowledge spillovers enhance the value of firms that invest in R&D. Furthermore, the value function implies that while the presence of knowledge spillovers does not have a direct effect on φ^*, it does exert a direct impact on the zero cut-off point for market survival R&D:

$$\frac{\delta_H - \delta_{RD}}{\delta_H} f\left(\left(\frac{\varphi^*_{RD}}{\varphi^*}\right)^{\sigma-1} - 1\right) = (1-\theta)\overline{f}_{RD} \tag{15}$$

Because $0 \leq \theta < 1$, we know that the right-hand side is smaller than in the benchmark model, implying a lower cut-off-point for R&D investment, ceteris paribus the aggregate variables.

To analyze the impact of the aggregate variables on both thresholds, we determine first how the inclusion of θ affects the cut-off point for profitable production φ^*. Once again this is governed by a total differentiation of the free-entry condition, recognizing that the right-hand-side features $(1-\theta)\overline{f}_{RD}$ rather than $\overline{f}_{RD}$. Applying the Pareto-distribution and using that changes in φ^* and φ^*_{RD} stand in a fixed relation:

$$\widehat{\varphi}^*_{RD} = \widehat{\varphi}^* - \frac{1}{(1-\theta)(\sigma-1)}\left(1 - \left(\frac{\varphi^*}{\varphi^*_{RD}}\right)^{\sigma-1}\right)d\theta, \tag{16}$$

total differentiation of the free entry condition yields:

$$\widehat{\varphi}^* = -Z' d\theta > 0$$

where $Z' < 0$ is a shorthand for an expression that is given in Appendix C.1. Also, knowledge spillovers by indirect imitation unambiguously increase the threshold for profitable entry φ^*.

Fig. 2 represents the impact knowledge spillovers by indirect imitation have on the various thresholds and densities. By (16), the proportional change of the threshold for profitable entry as an innovator equals the effect on $\widehat{\varphi}^*$ that is due to the changes in aggregate variables (the first part of the equation) plus a direct effect that knowledge spillovers have on $\widehat{\varphi}^*_{RD}$ (the second part of the equation). Note that in contrast to the direct imitation case the direct effect of knowledge spillovers enters Eq. (16) with a negative sign. Hence, when knowledge spillovers occur by indirect imitation, the (proportional) change of the threshold for entry as an innovator is less than that of the threshold for market entry. Moreover, in Appendix C.1 we show that this implies $\widehat{\varphi}^*_{RD} < 0 < \widehat{\varphi}^*$. Consequently, when knowledge spillovers imply that the costs of doing

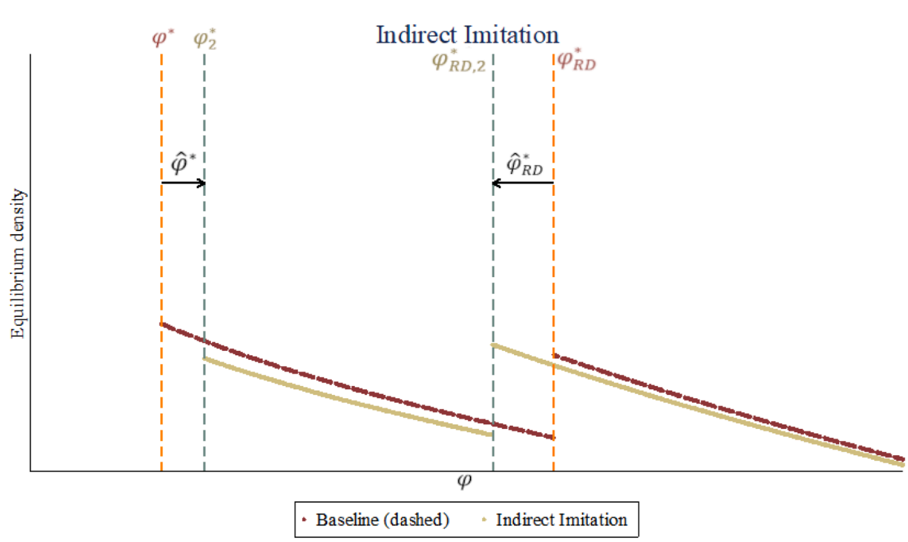


Fig. 2. The effect of spillovers on the composition of the cluster (indirect imitation).

market survival R&D decrease, firms of low productivity are forced out of the cluster ($\widehat{\varphi}^* > 0$), but for those that stay, the required productivity level to become an innovative firm is lower than without knowledge spillovers ($\widehat{\varphi}^*_{RD} < 0$).

Proposition 5. **(Threshold effects):** When firm productivity levels within an industry cluster are Pareto-distributed, knowledge spillovers that benefit innovative firms (indirect imitation) increase the productivity threshold for entering industry and lower the productivity threshold for becoming an innovative firm.

The intuition for these threshold effects is that indirect knowledge spillovers are an externality to firms that engage in R&D themselves. This shifts the balance of required R&D cost and reduced chance of exit for the most productive non-innovators. In terms of Fig. 2, there is an increase in density for the productivity levels $\varphi^*_{RD,\,2} \leq \varphi < \varphi^*_{RD}$. Firms within this range of productivity levels benefit from the selection effect, which they did not before. The increased presence of innovative firms puts pressure on the other firms in the cluster, particularly on those that do not innovate. The least productive firms will even have to exit, as indicated by the truncation at the left-hand side of the productivity distribution.

The overall effect on the density of both types of existing firms is determined by the relative strength of these two effects, see Appendix D. Our simulation results show that in over 99% of the cases, the crowding out by switching firms outweighs the effects of the left truncation. As such, Fig. 2 reflects the most likely results of knowledge spillovers by indirect imitation: there are lower densities for both existing innovative and non-innovative firms, but the density for firms that switch from not innovating to innovating increases. In the remaining less than 1% of the cases, the effect on the densities is positive for *all* productivity levels.

These changes in densities and the threshold effects outlined in Proposition 5 imply an ambiguous effect on cluster productivity, see Appendix C.2 and C.3. With knowledge spillovers by indirect imitation, the composition effect $d(M_{RD}/M)$ is positive, while the innovator effect $d\widetilde{\varphi}_{RD}$ is negative. Furthermore, the sign of the non-innovator effect $d\widetilde{\varphi}_H$ cannot be determined. This renders the overall effect on cluster productivity unclear and again we resort to an extensive numerical analysis to shed light on the matter. We will see that knowledge spillovers benefitting firms via indirect imitation increase cluster productivity for nearly all parameter configurations that we consider. Negative effects become likely in situations with high benefits from innovation, such as $\zeta < 0.5$.

Proposition 6. **(Composition effect):** When firm productivity levels within an industry cluster are Pareto-distributed, knowledge spillovers that benefit innovative firms (indirect imitation) increase the relative number of innovative firms in the cluster. This increases cluster productivity.

Proposition 7. **(Productivity effects):** When firm productivity levels within an industry cluster are Pareto-distributed, knowledge spillovers that benefit innovative firms (indirect imitation) decrease the average productivity of innovative firms. The effect on average productivity of non-innovative firms is unclear.

4. A numerical assessment of the effects of knowledge spillovers

Our previous analysis has shown clear effects on the degree of innovation in the cluster (Propositions 2 and 6). However, no clear conclusions could be drawn with respect to the productivity effects of knowledge spillovers. We therefore use STATA to perform an extensive numerical analysis of the model to shed further light on our analytical results. How exactly will knowledge spillovers affect cluster productivity? Which contextual parameters are particularly influential in this respect? How does the effect of a change of the extent of knowledge spillovers on average productivity $\widetilde{\varphi}$ decompose into the three effects we have identified ($\widetilde{\varphi}_{RD}$, $\widetilde{\varphi}_H$, and $\frac{M_{RD}}{M}$)? Which of these channels dominates the overall effect and on what does it depend? Throughout this section, we answer these questions and summarize them in a number of results that reflect the main conclusions of the numerical analysis. Moreover, we will use country- and sector-specific data from the OECD to match the outcomes of our numerical analysis to realistic clusters (see Appendix E).

We conduct our numerical assessment as follows. In line with Ghironi and Melitz (2005), we first simulate our models for a single standard calibration of contextual parameters that we derive from the literature. This is our baseline simulation and we use it to get a first idea of how cluster productivity is affected by increasing the extent of knowledge spillovers. Then, we check how robust the baseline results are by doing separate Monte Carlo simulations for each type of knowledge spillover.[24] For this, we iterate our model 131,611 times, drawing parameter configurations from random distributions.

As well as running separate simulations for the direct and indirect imitation models, we also include a simulation in which both types of knowledge spillovers are present at the same time. We will refer to this as the ‘hybrid model’. The reason for this extra simulation is twofold. First, in real-world situations both types of imitation can be expected to occur simultaneously (Nooteboom et al., 2007). Second, since the results from the hybrid model simulation appear not to be a straightforward average of the simulation results of both types of imitation in isolation, we believe they provide additional insights. When simulating the hybrid model, we consistently assume that the extent of both types of knowledge spillovers is the same ($\theta_{dir} = \theta_{indir}$).[25]

4.1. Baseline simulation

We use the following values for our baseline simulation:

Parameter choice baseline simulation

b_m	σ	$\alpha+1-\sigma$	ζ	$\varepsilon(0)$	$\varepsilon(\overline{f}_{RD})$	f_{RD}/f	α	θ
1	4	0.5	0.75	1	0	0.25	3.5	0.4

We normalize b_m and $\varepsilon(0)$ to 1, and $\varepsilon(\overline{f}_{RD})$ to 0. We will not vary these variables in our further simulations.[26] The other parameter choices are

[24] There are various methods that are used to simulate economic models (Canova, 1995). In macroeconomic modelling, grid searches are a commonly used method (see, for example, Gregory and Smith, 1993). Calibration within such models, however, is inherently different to the way we do it: in such models, specific numerical outcomes are known and models are calibrated to fit real-world data. As we have no real-world measurements of the size of this cluster-level effect to compare our outcomes to, we choose to use Monte Carlo simulations. Our approach is similar to the one used in simulations of input-output models, as described for example by West (1986).

[25] We restrict our analysis to this specific case of the hybrid model for reasons of clarity. We find that when $\theta_{dir} \gg \theta_{indir}$, the results of the simulation become very close to the results of knowledge spillovers by direct imitation, while in case of $\theta_{dir} \ll \theta_{indir}$ the results of the simulation become very close to the results of knowledge spillovers by indirect imitation. The intermediate cases, where $\theta_{dir} \approx \theta_{indir}$, provide different outcomes. The three cases – direct imitation, indirect imitation, and a hybrid case with $\theta_{dir} = \theta_{indir}$ – are thus representative of a much wider range of spillover combinations.

[26] b_m is the scale parameter of the Pareto distribution, determining the width of the distribution. The choice for not varying the two extreme values for ε is based on the fact that they are essentially the same as ζ. Together, these three variables determine the difference in market exit between innovative and non-innovative firms (δ_{RD}/δ_H), and with that, the benefits of investing in R&D. As such, varying all three parameters at the same time risks muddying the general picture. We ran some simulations to check if changing $\varepsilon(0)$ and $\varepsilon(\overline{f}_{RD})$ has different effects than changing ζ, keeping the ratio δ_{RD}/δ_H fixed. For indirect imitation, we find that changes in these two extreme values are in fact entirely equivalent to changes in ζ; for direct imitation and the hybrid model, the differences are minor.

extensively motivated in Appendix E.[27] For here, it suffices to say that the chosen parameter settings lead to a realistic outcome for the percentage of innovators, which is about 40% of firms in our baseline simulation. Compared to the numbers found by the 2015 UK Innovation Survey (Department for Business Innovation & Skills, 2016), our baseline simulation establishes a very realistic percentage of innovators: in the UK, about 53% of firms were innovative, while 22% performed technological innovation.

Fig. 3 shows the results for the baseline simulation for the direct imitation model (Panel A), the indirect imitation model (Panel B), and the hybrid model (Panel C).[28] The outcomes of the baseline simulation show that the ambiguity we found in the previous sections disappears: knowledge spillovers by direct imitation have a negative effect on cluster productivity, while knowledge spillovers by indirect imitation have a positive effect. Decomposing it into the various channels we have identified in Eq. (13), we see that the direct imitation model entails a negative composition effect, countered by positive productivity effects for innovators and non-innovators. For the indirect imitation model these effects are exactly opposite. Note that for both types of knowledge spillovers it holds that the composition effect determines the sign of the overall effect. This is not true, however, when both types of knowledge spillovers operate at the same time, as in our hybrid model. In that case the signs of the decomposed effects match those of indirect imitation, including a positive composition effect, but the composite effect we end up with is negative. Thus, the outcomes of the hybrid model are not a matter of simply combining the results of the two other models. The intuition behind this is that in both direct and indirect imitation, the change in the cutoff for innovation, φ^*_{RD}, is much larger than the change in the cutoff for profitable entry, φ^*. In the hybrid model, these effects are closer in size. This causes the effect on the composition of firms and on the average productivity of non-innovative firms to be much smaller, leaving the negative effect on the average productivity of innovative firms as the main component of average cluster productivity.

4.2. Monte Carlo analysis

To check how robust these baseline results are we run a Monte Carlo simulation. We want the random draw of parameters to resemble real world industries as closely as possible, so we have different distributions and limits for each parameter. Each simulation consists of 131,611 iterations of randomly drawn parameter configurations.[29] The distributions of the parameters are described in the table below. All variables are generated using normal distributions as is common in sampling theory (West, 1986), except for σ, $\bar{f}_{RD}/f$, and θ. For these variables we use χ^2-distributions to put greater emphasis on the low ends of the distribution, yielding a more realistic representation of the data.

Choice of parameter distributions

Parameter	σ	$\alpha+1-\sigma$	α	ζ	$\bar{f}_{RD}/f$	θ
Type of distribution	χ^2	Normal	n.a.	Normal	χ^2	χ^2
Mean	4	0.5	-	0.75	0.25	0.4
Standard Dev.	1	0.25	-	0.1	0.2	0.2
Minimum	2	0	-	0.3	0.005	0
Maximum	7	3	-	0.95	1	0.9

As the output of these 131,611 random iterations is broad, it can be difficult to interpret what it means for more specific cases. As such, we provide below six realistic example clusters to which our model may be applied. We include the parameter configuration of the baseline simulation, as well as a number of specific country/sector combinations based on real-world clusters. Using a loss function, we determine for each of the six clusters which of the 131,611 iterations is closest to a cluster's particular configuration.[30] In our analysis, we will use these example cluster configurations as additional points of reference to explain the results. We will pay particular attention to the Dutch pharmaceutical and Italian apparel clusters. Both clusters are outliers in the results, which makes it interesting to discuss how this may relate to their specific parameter configurations.

Country	Sector	Parameter choice specific clusters						
		b_m	σ	$\alpha+1-\sigma$	ζ	f_{RD}/f	α	θ
Baseline	*Baseline*	*1*	*4*	*0.5*	*0.75*	*0.25*	*3.5*	*0.4*
USA	Software	1	4†	0.5	0.83**	0.58	3.5	0.4
GER	Cars	1	3.99	0.59	0.49*	0.32	3.58	0.4
NLD	Pharma	1	2.39††	0.19	0.52*	0.33	1.58	0.4
UK	Finan	1	4†	0.5	0.87	0.007	3.5†	0.4
ITA	Appar	1	2.88	0.26	0.63*	0.04	2.14	0.4
JAP	Manufacturing	1	3.06†††	0.49	0.81***	0.22	2.55	0.4

† No data for software or finance was available, so our baseline parameters were used instead.
†† No data for pharmaceuticals was available, so data for the chemical industry was used instead.
††† No data for manufacturing in the broad sense was available, so data for the production of electrical machinery was used instead.
* No data for the specific sector was available, so data for manufacturing was used instead.
** No data for the US was available, so data on the Canadian Information and Communication sector (which includes Software) was used instead.
*** No data for Japan was available, so data for South Korea was used instead.

4.2.1. The effects of the extent of knowledge spillovers

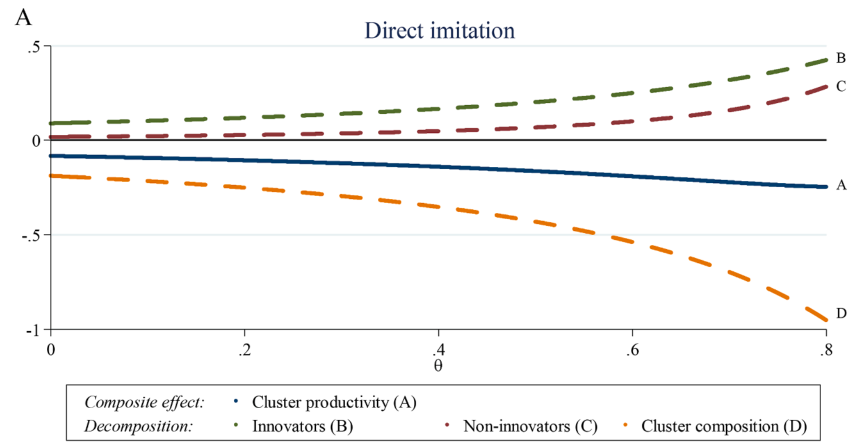
We consider first how the main variable of interest in our analysis – the extent of knowledge spillovers θ – influences the outcomes of the models. The results for θ can be seen in Fig. 4. In the figures, the dots indicated by A represent the top 10% of iterations, while the dots indicated by C represent the bottom 10%. The remaining 80% of all iterations, the dots indicated by B, are in a relatively narrow range, regardless of the type of knowledge spillovers we look at. Most of the variation thus lies in the extremes. The example clusters that we have given are mostly in the middle 80% of iterations, with some exceptions:

[27] A possible limitation is that $\bar{f}_{RD}/f$and ζ can be expected to be negatively correlated. In industries with high ζ the benefits of investing in R&D are low, leading to low R&D investment, just as in case of high R&D cost $\bar{f}_{RD}\ /f$. Hence, $\bar{f}_{RD}/f$ and ζ reflect two sides of the same coin. In our simulations we allow these two parameters to be correlated as a robustness check, finding no qualitative differences in outcomes.

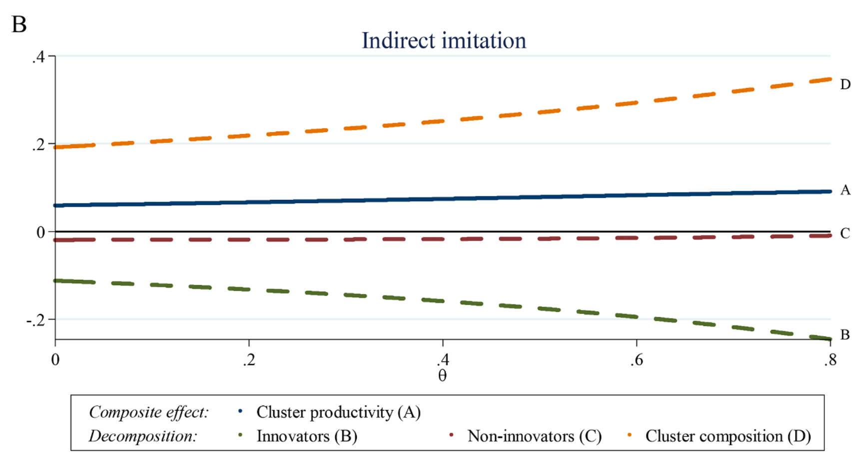
[28] We limit the scale of the graph to $\theta < 0.8$, mainly for expositional reasons: for knowledge spillovers by direct imitation the size of the various effects becomes much larger at $\theta > 0.8$, compromising the legibility of the graph. For the other two models, this limit is non-binding. Regardless, values of $\theta > 0.8$ are unrealistic in our view, as motivated in Appendix E.

[29] We draw 250,000 parameter configurations from full distributions. After we draw them, we cut off the extreme ends of the distribution according to the table, which causes small deviations in means and standard deviations. In this process, we remove 118,389 configurations. Because all data are drawn randomly, removing these outliers does not affect the results. All draws are made with the same seed so that they are comparable, also across models.

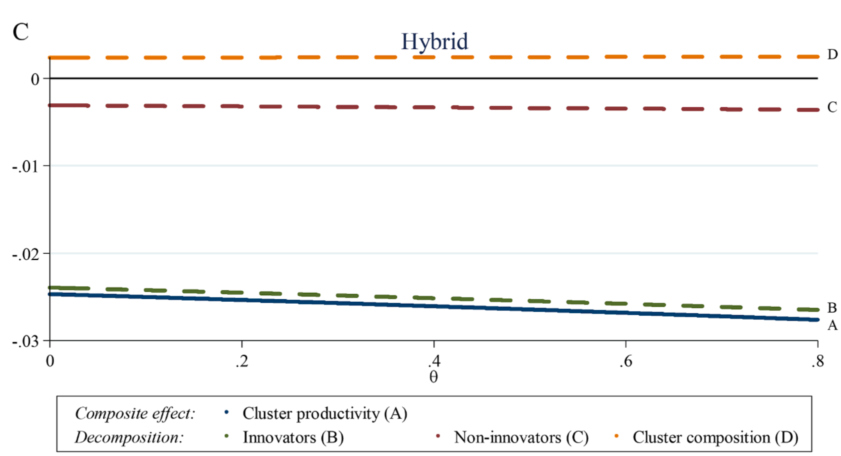
[30] The following procedure is applied. For each iteration, we take the absolute difference between the parameter value of that iterations and the parameter value of the cluster as we established it from the data and square it. We then take the iterations for which the sum of the squared differences is the smallest possible. As such, the exact values of the iterations that we select will deviate slightly from the values in the table. For example, the configuration for the Dutch pharmaceuticals cluster has a value of 2.37 for σ, 0.21 for $\alpha+1-\sigma$, 0.57 for ζ, 0.31 for f_{RD}/f, and 0.36 for θ.



Source: Own calculations with Stata 14.

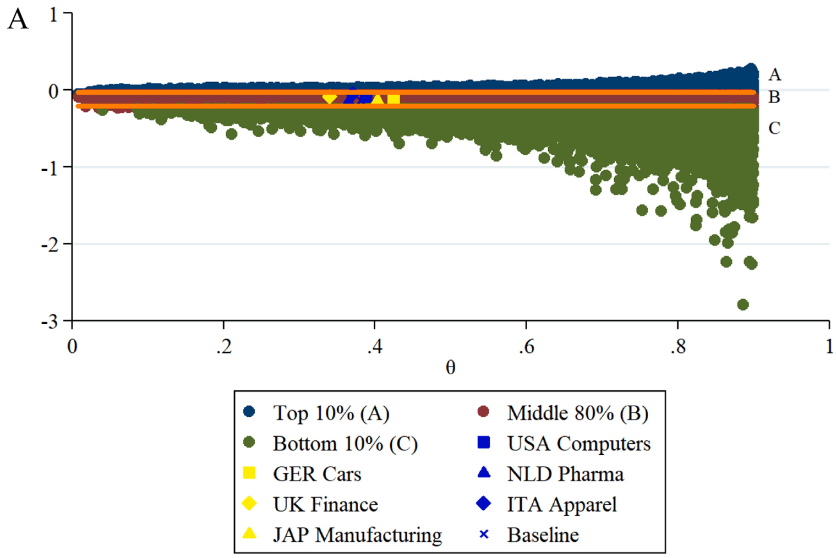


Source: Own calculations with Stata 14.

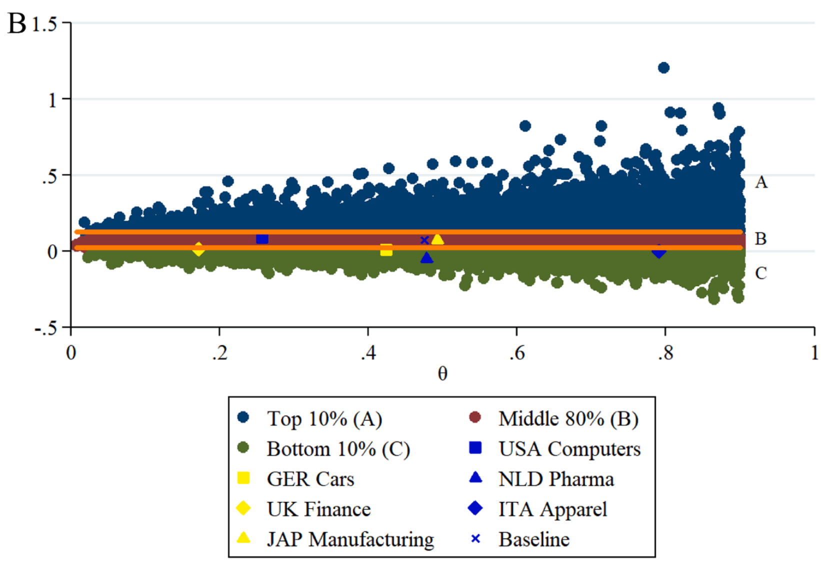


Source: Own calculations with Stata 14.

Fig. 3. A: The effect of the extent of spillovers (baseline, direct imitation), B: The effect of the extent of spillovers (baseline, indirect imitation), C: the effect of the extent of spillovers (baseline, hybrid model).



Source: Own simulations with Stata 14



Source: Own simulations with Stata 14

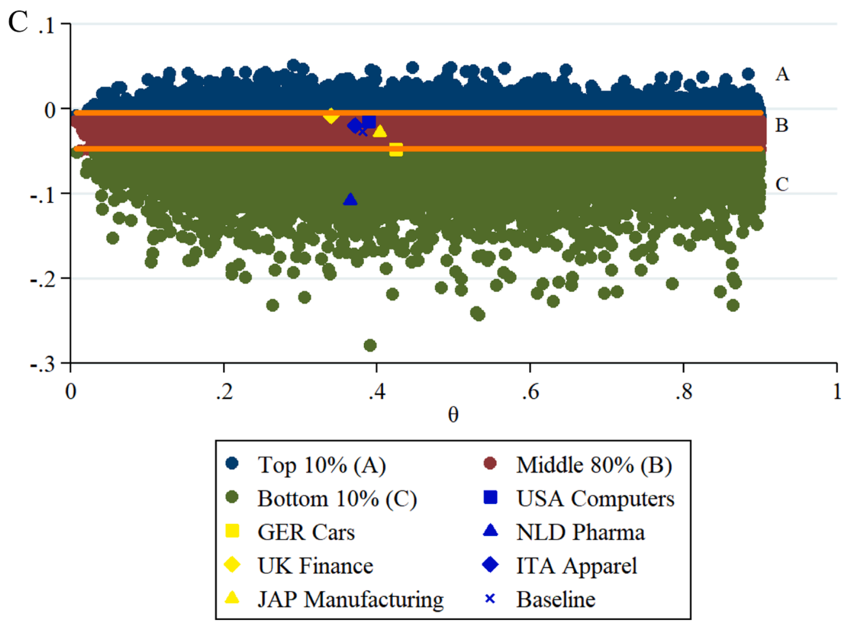
Fig. 4. A: The effect of the extent of spillovers (Monte Carlo, direct imitation).
B: The effect of the extent of spillovers (Monte Carlo, indirect imitation).
C: The effect of the extent of spillovers (Monte Carlo, hybrid model).

in the *direct imitation* model, all clusters are in the middle 80%; in the *indirect imitation* model, the Italian apparel and Dutch pharmaceutical clusters are in the bottom 10% of iterations; and in the *hybrid* model, the Dutch pharmaceutical cluster is in the bottom 10% of iterations.

For all three models it is clear that the unidirectional effects of the baseline specification cannot be upheld: for the direct imitation model we find positive net effects of knowledge spillovers on cluster productivity in about 3,000 (2.2%) iterations and for indirect imitation we find a negative overall effect for 12,000 iterations (9.4%). This includes two of our example clusters (the Dutch pharmaceutical cluster and the Italian apparel cluster. In the hybrid model, only very few iterations result in an overall positive effect on cluster productivity.[31]

Result 1. **(Cluster Productivity Effects):** Knowledge spillovers by direct imitation have a negative effect on cluster productivity, while knowledge spillovers by indirect imitation have a positive effect. This holds for *most* iterations (>90%). When knowledge spillovers occur

[31] In fact, combining the two types of knowledge spillovers leads to negative results in more iterations than in the case of knowledge spillovers by direct imitation. More specifically, 129,392 out of 131,611 of all iterations (98.3%) show a negative sign for $d\tilde{\varphi}$ in the hybrid model versus 128,604 out of 131,611 iterations in the direct imitation model (97.8%). This is in line with the intuition provided before: as the innovator effect tends to be dominant within the hybrid model, its results are more stable than the other two models.



Source: Own simulations with Stata 14

Fig. 4. (*continued*).

through both direct and indirect imitation with the same intensity (hybrid model), the overall effect is negative for *almost all* iterations (>98%).

The decomposition of the overall productivity effects is given in Fig. 5. Panel A gives the innovator, non-innovator and composition effects for the direct imitation model, Panel B for the indirect imitation model and Panel C for the hybrid model. In the direct imitation model (Panel A), all effects are in line with Propositions 2-4: the effects that occur through the average productivity levels of innovative and non-innovative firms are unambiguously positive, while the effect through the ratio of innovative to non-innovative firms is unambiguously negative. Furthermore, in most iterations (about 94%) the innovator effect is greater than the non-innovator effect. The exceptions occur mostly when there are relatively few innovative firms to begin with, which makes sense in view of the M_{RD}/M weights in (13).

The decomposition effects for the indirect imitation model (Panel B) are in line with Propositions 6 and 7. Note that now we also find an unambiguous (negative) effect of knowledge spillovers on the productivity levels of non-innovative firms. Furthermore, an increase in the level of knowledge spillovers only magnifies the innovator effect. A few outliers aside, the non-innovator and composition effects do not become stronger as θ increases. This is in contrast to the direct imitation model, where knowledge spillovers magnify the effects of all three decomposed effects. The Dutch pharmaceutical and the Italian apparel clusters are both in the lowest 10% of iterations, but the figures show that they are in that subsample for very different reasons. The Italian apparel cluster, like in the direct imitation model, appears in the lowest 10% because all three of the decomposed effects are very small. This cluster has the smallest composition effect of all of the example clusters. Conversely, the Dutch pharmaceutical cluster is in the lowest 10% because its innovator effect is very strongly negative, which is not compensated by a similarly strong positive composition effect. The hybrid model seems to average out the decomposed effects of the direct and indirect imitation models. The signs of the decomposed effects are in line with the case of knowledge spillovers by indirect imitation though. If both types of knowledge spillovers are active at the same time, knowledge spillovers will cause more firms to engage in market survival R&D. Overall, however, the average productivity of both innovative and non-innovative firms will decline. In the hybrid model, the Dutch pharmaceutical cluster is the only example cluster that jumps out. It has, by a small margin, the largest positive composition effect of any of the clusters we discuss, but this positive effect is cancelled out by the innovator effect, which is the largest negative effect of any of the clusters we discuss.

Result 2. (Decomposition): The simulation results confirm Propositions 2-4 (direct spillovers) and Propositions 6 and 7 (indirect spillovers): knowledge spillovers by direct imitation increase the average productivity of innovative *and* of non-innovative firms in the cluster. The relative number of innovative firms goes down. The opposite holds for indirect spillovers. This holds for all iterations. In the hybrid model, the relative number of innovative firms always goes up, while the average productivity of innovators always goes down. The effect on the average productivity of non-innovators is negative in most iterations (>90%).

4.2.2. The effects of other parameters

We have established that knowledge spillovers by indirect imitation mostly have a positive effect, while knowledge spillovers by direct imitation, and when both types of spillovers are active, are mostly negative. But how does the impact of increased knowledge spillovers depend on the (other) model parameters?[32] To assess this we divide the 131,611 iterations in subsets from low to high for each parameter, determining the results of the model for each subset accordingly. For example, when applying this procedure to the systemic risk parameter ζ, we compare the mean values of $d\widetilde{\varphi}/d\theta$ across the subsets $0.35 < \zeta < 0.6$, $0.6 < \zeta < 0.7$, $0.7 < \zeta < 0.8$, and $0.8 < \zeta < 1$. Since the other parameter values are drawn independently from ζ, the outcomes of this exercise are comparable to a ceteris paribus variation of ζ.

[32] Note that this means that we examine the effect that these parameters have on the effect of spillovers on cluster productivity, rather than the direct effect they may have on cluster productivity themselves. For example, Grilli et al. (2010) conclude that firm dissolution may be a desired outcome for firms in high-tech sectors, suggesting that a low ζ is beneficial to cluster productivity, at least in these sectors. These effects are interesting, but the goal in this paper is to find which clusters benefit more from knowledge spillovers, rather than to find which parameter values make for the most efficient cluster irrespective of knowledge spillovers.

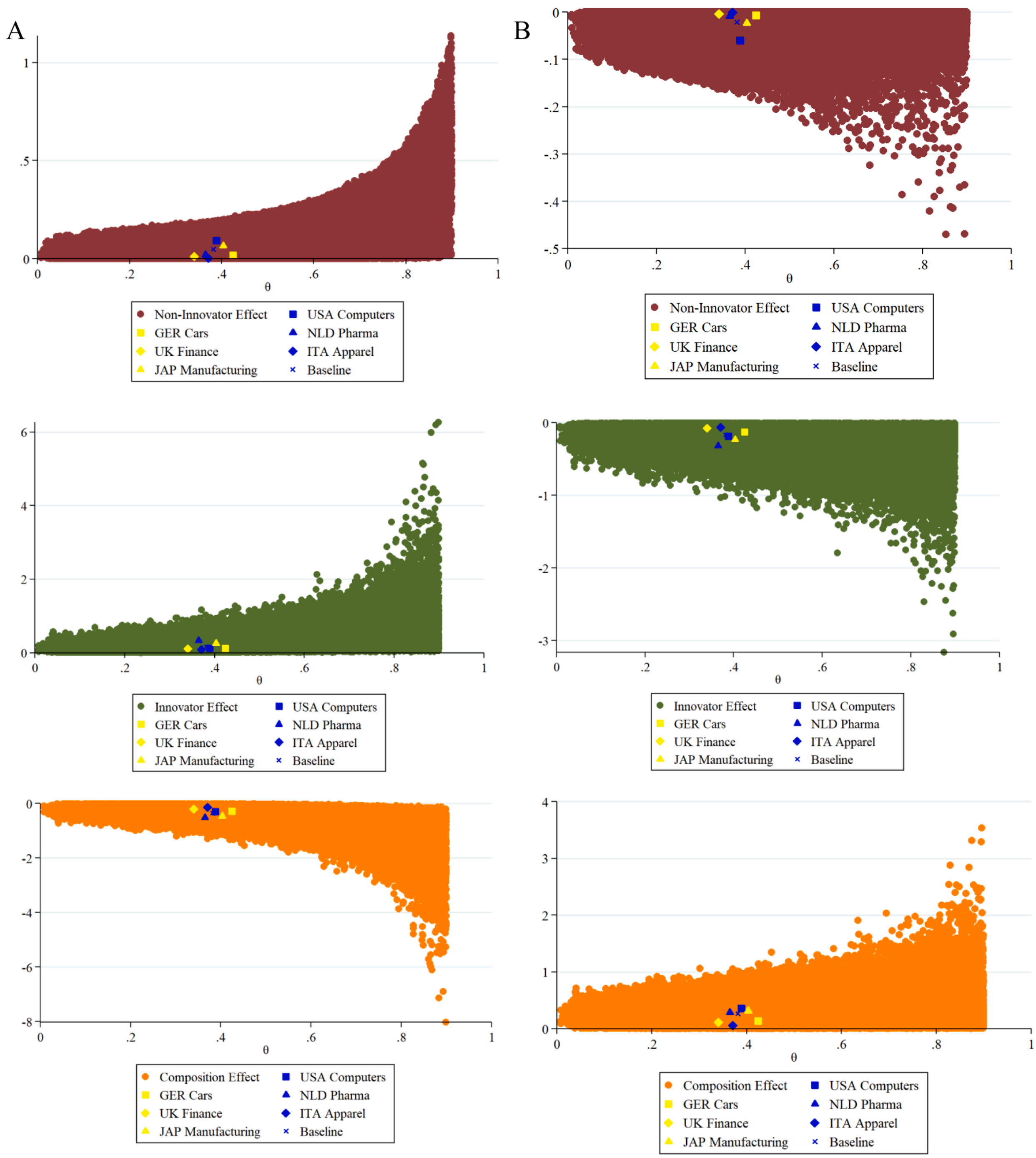


Fig. 5. A: The decomposition of the effects of spillovers (Monte Carlo, direct imitation).
B: The decomposition of the effects of spillovers (Monte Carlo, indirect imitation).
C: The decomposition of the effects of spillovers (Monte Carlo, hybrid model).

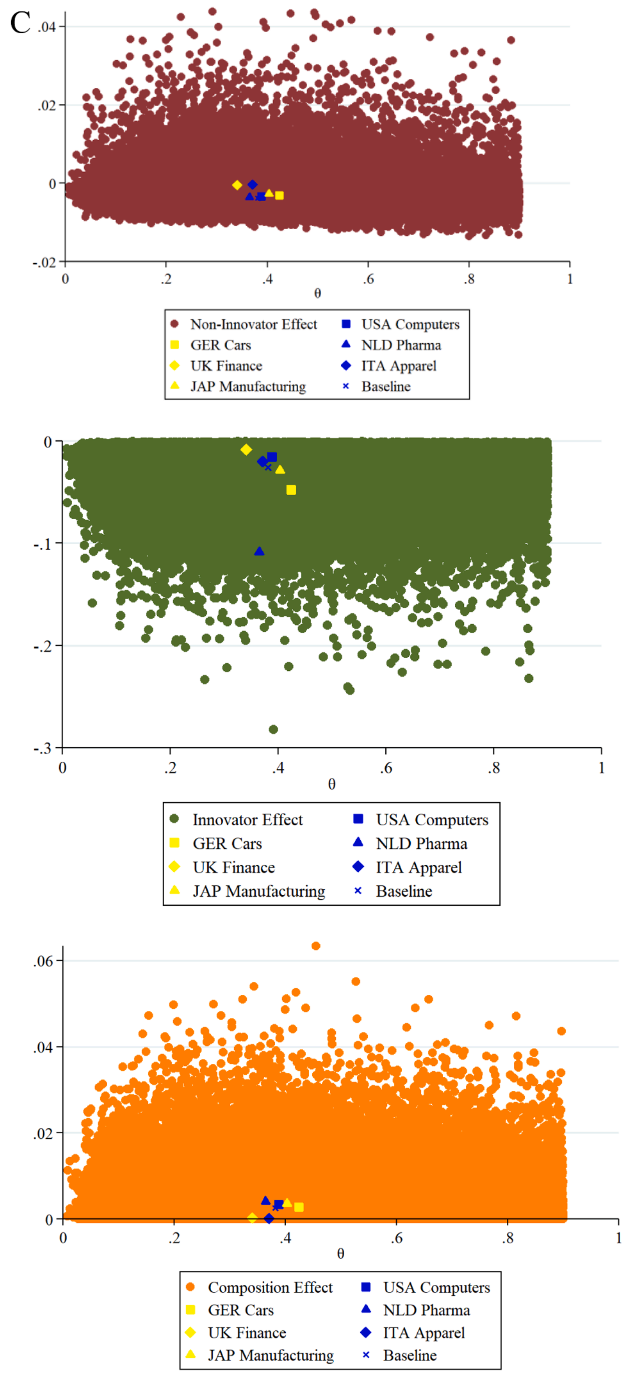


Fig. 5. (*continued*).

Table 1 below provides a summary of the results. The full details of our analysis are given in Tables A.1 and A.2 in Appendix F. For sake of reference, the tables also include the impact of θ.

Result 3. **(Parameter sensitivity):** While most parameters, ceteris paribus, have a fairly consistent influence on the effect of spillovers on cluster productivity, there is a lot of underlying variance in the results.

Table 1 shows that for both direct and indirect knowledge spillovers the effects of increased knowledge spillovers are enhanced when the extent of knowledge spillovers θ is larger: the effect of direct spillovers is more negative for larger values of θ and the effect of indirect spillovers is more positive. For the hybrid model, the particular value of θ has practically no impact on $d\widetilde{\varphi}/d\theta$. By contrast, a higher systemic risk-related exit chance ζ has a positive effect on the effect of knowledge spillovers in all three simulations. Increasing $\bar{f}_{RD}/f$, finally, implies greater benefits of knowledge spillovers in case of indirect spillovers; for the direct imitation and hybrid models we find non-linear effects.

To offer some explanation to these results, we recall that a ζ and $\bar{f}_{RD}/f$ determine the benefits and costs of innovation. Through that, they also determine how much can be gained from knowledge spillovers. As such, the particular values of ζ and $\bar{f}_{RD}/f$ will also affect how firms will change their R&D decisions due to knowledge spillovers. Knowledge spillovers by direct imitation are beneficial for non-innovative firms and the least-productive innovative firms will turn into non-innovative firms. However, when the benefits of innovation are low and/or costs are high (high ζ and/or high $\bar{f}_{RD}/f$), fewer firms will make this switch, lessening the negative effect of direct knowledge spillovers. Similarly, if knowledge

Table 1
Parameter value changes and the effect of spillovers on average cluster productivity.

$d\left(\frac{d\widetilde{\varphi}/d\theta}{dx}\right) =$	Direct spillovers	Indirect spillovers	Hybrid spillovers
$x = \theta$	-	+	0
$x = \zeta$	+	+	+
$x = \bar{f}_{RD}/f$	~	+	~
$x = \sigma$	~	~	+
$x = \alpha + 1 - \sigma$	+	-	+
Sign of $d\widetilde{\varphi}/d\theta$	-	+	-

Explanation of the table: a '+' ('-') indicates that $d\widetilde{\varphi}/d\theta$ monotonically increases (decreases) when moving from low-value subsets to high-value subsets. A '0' indicates that there is no clear difference and a '~' means that the relation is non-monotone. The sign of $d\widetilde{\varphi}/d\theta$ is included as a reminder of our previous discussion.

spillovers occur by indirect imitation, low benefits and/or high costs imply that fewer low-productive, non-innovative firms will decide to become innovative firms, which means that the main beneficiaries are firms that are already highly productive. This amplifies the positive effects of knowledge spillovers by indirect imitation.

The elasticity of substitution σ mostly serves to magnify all the effects of the model: each of the three decomposed effects become smaller when σ increases (in absolute terms). This holds for all types of knowledge spillovers. A higher σ means that competition is fiercer, lowering the number of low-productivity firms. Their weight in the calculation of average productivity levels declines,[33] reducing the impact knowledge spillovers can have. It underlines the importance of firm productivity heterogeneity to the effects we draw attention to. It probably also explains why the Dutch pharmaceutical cluster, which has the lowest value of σ of any of the example clusters that we have, is among the 10% of iterations with the lowest average productivity effect in both the indirect imitation model and the hybrid model. The Italian apparel cluster has the second lowest value of σ of our example clusters, which may explain why it is among the lowest 10% of iterations for the indirect imitation model.

The impact of varying $\alpha + 1 - \sigma$ is explained by understanding the effect of α, as the value of σ is constant across the subsamples.[34] The value of α affects the distribution of drawn productivity levels per se, with a higher value skewing the Pareto-distribution towards less productive firms. When α increases, the chance that a firm invests in market survival after profitable entry diminishes, lowering the number of innovative firms in the industry cluster. This reduces the impact knowledge spillovers can have and therefore reduces the effect of knowledge spillovers in both models. This further explains why the Dutch pharmaceutical cluster and the Italian apparel cluster are in the bottom 10% in the indirect imitation simulation.

Thus far, we have focused on explaining the sensitivity of the model outcomes to changes in each of the separate model parameters while holding all other parameters constant. But which parameters are actually most important in determining the outcomes of the model? To find out, we divide the iterations into four subsamples ranging from low to high values of $d\widetilde{\varphi}/d\theta$. Specifically, we take the lowest and the highest 1% and 10% of iterations for $d\widetilde{\varphi}/d\theta$ as subsamples. We then compare the average parameter values in each of these subsamples with the average parameter values of the entire simulation, using standard deviations to indicate the extent of the difference.

Table 2 shows the outcome of this analysis in summary form; Table A.3 in Appendix F gives the underlying data.

Result 4. (Policy relevant parameters): The parameters that are most important for the effect of knowledge spillovers on cluster productivity depend on the type of knowledge spillovers.

For the *direct imitation model,* the dominant parameters are θ and $\alpha + 1 - \sigma$. Both the high and the low side of the $d\widetilde{\varphi}/d\theta$ distribution feature high levels of θ. Furthermore, the effect of knowledge spillovers on cluster productivity goes up when $\alpha + 1 - \sigma$ increases. The effect of knowledge spillovers on cluster productivity is also very likely to be negative if $\alpha + 1 - \sigma$ is small. The mean for the lowest 1% of iterations is 0.14. As such, this parameter seems to have an extremely strong effect on the outcomes of the model. The example cluster with the lowest value of $\alpha + 1 - \sigma$, the Dutch pharmaceutical cluster, is also the example cluster with the lowest outcome for $d\widetilde{\varphi}$, further supporting this. The other parameters are relatively unimportant for determining whether or not an extreme outcome is reached, though a substantially higher ζ and $\bar{f}_{RD}/f$ influence the impact of knowledge spillovers positively. For the *indirect imitation model*, all parameters have some importance, of which ζ seems to be most important: the extreme outcomes of the model correspond to draws of ζ that are also extreme. For the lower values, σ appears to be a very strong parameter, as the average value of σ for the lowest 1% of iterations is close to the lower bound that we have established for this variable. The results for the *hybrid model* are a combination of the other two models. As in the indirect imitation model, ζ and σ have a strong influence on the results. As in the direct imitation model, $\alpha + 1 - \sigma$ has a very strong effect on the outcomes. Moreover, all parameters except for θ are of some importance for determining extremes.

What do these results imply? The outcomes in Table 2 suggest that when knowledge spillovers take place through *direct imitation*, cluster productivity is best served by intermediate values of θ. High levels of θ are associated with both very low and very high negative cluster productivity effects, but the latter only occurs when the high level of θ goes along with values of ζ and $\bar{f}_{RD}/f$ that limit the proportion of innovative firms in the industry cluster. The importance of $\alpha + 1 - \sigma$ underlines that the more the distribution is skewed towards low-productivity firms, the less negative the impact of knowledge spillovers will be in the direct imitation model. In the *indirect imitation model*, cluster productivity is best served by a relatively high systemic risk of firm exit ζ. But also, then it is a combination of factors that will make the effects of knowledge spillovers highest. While low values of ζ seem to be an important determinant of very low (yet mostly positive) cluster productivity effects, an increase in ζ will only improve knowledge spillover effects when also θ and $\bar{f}_{RD}/f$ are high. The reason is similar to what we have seen before: the combination of high knowledge spillovers and high R&D costs imply great benefits of being innovative, increasing the proportion of innovative firms in the cluster. Likewise, larger knowledge spillover effects are associated with lower values of α.

5. Conclusion and Policy Implications

This paper demonstrates a mechanism by which cluster externalities affecting firms asymmetrically will change the composition of the cluster, thereby affecting cluster-level outcomes. We show this mechanism in a Melitz (2003) type of model that we apply to an industry cluster where knowledge spillovers occur and where firms of different productivity invest in R&D to enhance their market survival. The productivity heterogeneity of firms within the cluster results in a differentiation between high-productivity firms and low-productivity firms. High-productivity firms innovate to increase their chances of survival, and low-productivity firms do not innovate and leave the market after one period. The threshold demarcating innovative and non-innovative firms is endogenous and also depends on the extent of knowledge

[33] As can be seen in equations A.11 and A.12 in the appendix.

[34] Both $\alpha + 1 - \sigma$ and σ are drawn independently, so we can interpret the ordering of the full sample in subsets from low to high values of $\alpha + 1 - \sigma$ as ceteris paribus σ. However, the values are not ceteris paribus α, implying that $\alpha + 1 - \sigma$ and α perfectly correlate.

Table 2
Identifying the most important parameters.

	Direct Lowest	Direct Low	Direct High	Direct Highest	Indirect Lowest	Indirect Low	Indirect High	Indirect Highest	Hybrid Lowest	Hybrid Low	Hybrid High	Hybrid Highest
θ	+	+	0	+	+	0	0^+	+	0	0	0	0
ζ	0	0	0	0	-	0^-	0^+	+	–	0^-	+	+
$\bar{f}_{RD}/f$	0	0	0	+	0	0^-	0^+	+	+	0^+	0	+
σ	0	0	0	0	-	-	0	0	-	0^-	0	-
$\alpha + 1 - \sigma$	-	-	0^+	0^+	0^-	0	0^-	-	-	0^-	0^+	+

Explanation of the table: The columns show how parameter values are different in different subsets of the simulated sample. The lowest subsample is the bottom 1% of values of $d\tilde{\varphi}/d\theta$ in a particular simulation and the low subsample is the bottom 10% of outcomes. The high subsample is the top 10% of outcomes and the highest subsample is the top 1% of outcomes. A $++$ or a - - refers to a deviation of the subsample's mean which is greater than two standard deviations of the parameter's distribution in the full sample and a $+$ or – refers to a deviation that is greater than one standard deviation. A 0^+ or a 0^- refers to a deviation of the subsample mean that is between a half and a full standard deviation and a 0 refers to a deviation which is smaller than half a standard deviation.

spillovers within the cluster. This causes a feedback effect of knowledge spillovers on cluster-level outcomes (here: cluster productivity) through cluster composition. The type of spillovers is of crucial importance for the implications of this feedback effect.

The feedback effect we demonstrate in this paper has been overlooked in the literature thus far. In our model, it arises because knowledge spillovers affect a firm's cost-benefit analysis regarding R&D investment. This changes the composition of innovative and non-innovative firms in the cluster, the average productivity levels of these two groups of firms, and overall cluster productivity. Propositions 1-7 outline these effects, showing a pivotal role for the type of knowledge spillovers. When knowledge spillovers imply that non-innovative firms can improve their chance of survival by imitating the R&D practices of innovative firms, the relative number of innovative firms in the cluster diminishes. The average productivity of both innovator and non-innovators increases. If, by contrast, knowledge spillovers materialize as a reduction of R&D costs, benefitting innovative firms, the relative number of innovative firms will increase. In that case, the effect on average productivity of innovators is negative while for non-innovators it is unclear. The impact on overall cluster productivity is analytically ambiguous for both types of knowledge spillovers.

Extensive numerical simulations, however, show that in the vast majority of cases the effect on cluster productivity is clear. Using STATA, we generated 131,611 randomly drawn parameter configurations for each type of knowledge spillovers. These parameter configurations are drawn from a set of parameter distributions resembling real-life data. Despite some exceptions, the conclusion of these simulations is that knowledge spillovers that improve the survival chances of non-innovative firms have a negative effect on cluster productivity, while knowledge spillovers that reduce R&D costs for innovative firms improve cluster productivity. If both types of knowledge spillovers are active at the same time, the type of spillover that is stronger tends to dominate. If the two types of spillovers are equally strong, the combined effect of these spillovers on cluster productivity is negative. The parameters that are of key importance for these results also depend on the type of spillovers. When knowledge spillovers benefit non-innovative firms, the crucial parameters are the extent of knowledge spillovers and the skewness of the productivity distribution of firms in the cluster. When knowledge spillovers benefit innovative firms, the crucial parameter is the degree of systemic risk of firm exit. These results are supported by six example real-world clusters that we identified based on OECD data.

Our analytical and numerical results contribute to the theoretical and empirical literature on the effects of clusters. Theoretically, we show that firm productivity heterogeneity in the presence of knowledge spillovers within clusters leads to an additional cluster-level effect. By including different forms of knowledge spillovers, our model helps explain the mixed evidence on which firms benefit from knowledge spillovers. We reach these results by letting R&D affect the chance of market exit rather than the productivity level of a firm, a novel approach in the Melitz literature. Furthermore, by modelling firms as being heterogeneous in productivity, we show that effects of knowledge spillovers only affect part of the cluster. This contrasts with most of the clustering literature. The link between the asymmetric effects on firms and cluster productivity sets our analysis further apart from the literature.

For the empirical literature on the effects of clustering, our general point is that there is an ex-post selection effect, as knowledge spillovers cause different exit rates for different firms. We see three implications. First, attempts to measure ex-ante geographical sorting of firms should also take into account this ex-post selection effect to prevent overestimation of the ex-ante sorting. Second, with cluster externalities affecting firm characteristics, empirical analyses incorrectly estimate the cluster-level effect of knowledge spillovers, even when controlling for ex-ante self-selection by firms. Third, measuring the effect on firms is not equivalent to measuring the effect on cluster productivity, as the effect on cluster composition is also important. In order to overcome the ex-post selection effect we highlight here, empirical studies may benefit greatly by including data about firms exiting the market.

Even though our analysis is mainly theoretical, we still believe it offers important insights for policy. The most important policy implication is that cluster externalities will affect the composition of firms within the cluster. This gives rise to a feedback effect at the cluster level that will affect cluster-level outcomes. The ramifications of this effect will depend on the specific variable policy makers wish to target,[35] but the main message for policy is clear: when clusters consist of heterogeneous firms, policy makers would do well to acknowledge the feedback effect on cluster composition when designing cluster policy.

Furthermore, the type of knowledge spillovers is of prime importance for the feedback effect on cluster-level outcomes. This implies that cluster policy design should start with ascertaining the nature of knowledge spillovers in clusters. Our analysis gives some clues regarding the identification of the type of spillover present in the cluster. For example, direct spillovers may be associated with particular spillover mechanisms, such as supply contracts, licensing, and inter-firm labor mobility (McCann and Folta, 2008; Van Looy et al., 2011). Indirect spillovers, by contrast, may be associated with university-firm collaborations, as in Zucker et al. (1998b) for the American biotech industry or Cassiman et al. (2018) for the Belgian micro-electronics

[35] Our paper has shown results for average productivity and innovation. Other objectives are of course also possible. These objectives are likely to be related to each other, however. In similar setups, it has been shown that higher productivity may imply more exports (Melitz, 2003), higher wages (Bernard et al., 2007), and higher-quality goods (Verhoogen, 2008).

industry. But clearly much more research is needed to give specific guidance for the identification of the type of spillover.[36]

A second takeaway for policy is that the effect on cluster-level outcomes depends on the specific industry and country. Both our theoretical and numerical analyses show that cluster effects of knowledge spillovers depend on the specific situation (type of spillovers, extent of firm heterogeneity, industry-specific exit risk, and so on). Consequently, there is no one-size-fits-all approach to spillover policy within clusters. This presents a challenge to policy makers. To illustrate, some policies, such as patent policies, tend to be very broad because they require coordination at a national or even at an international level. Based on the results of our analysis, however, we emphasize the importance for policy makers to develop more narrow and targeted cluster policies.

We conclude by discussing a number of limitations present in our analysis. First, our model does not include location choice, one of the central foci in the literature on clustering and agglomeration. The existence of cluster externalities will likely have an impact on which particular firms are drawn to clusters. Such an impact may then affect cluster composition through an ex-ante adverse selection effect. However, the asymmetric effects of clustering found in the literature also occur in the absence of geographical sorting; they are an intrinsic aspect of clustering. Furthermore, including an ex-ante selection effect will only serve to magnify the effects we highlight: if a cluster entails knowledge spillovers that benefit low-productivity firms, these firms are more likely to locate in such a cluster, strengthening the effects of the asymmetries we discuss. As such, the ex-post selection effect we showcase may be an underlying reason for the existence of an ex-ante selection effect.

Second, the way we have modelled knowledge spillovers – by considering two extreme ways of how they might occur – is highly stylized. In our analysis, either non-innovators or innovators exclusively benefit from knowledge spillovers. But one may argue that firm productivity heterogeneity also matters with regards to knowledge spillovers. Following the literature on knowledge spillovers, it is likely that knowledge spillovers will be most intense when firms are at a sufficiently large distance from the knowledge frontier (they must have something to learn), while they also should not be too far away from it (they must be able to learn). This may lead to a hump-shaped relation between knowledge spillovers and a firm's distance to the knowledge frontier. Presuming that productivity is positively related to knowledge, this implies that firms in the middle of the productivity spectrum may benefit most from knowledge spillovers. If this is the case, the implications of knowledge spillovers on cluster productivity will then depend on whether these firms are non-innovators, innovators, or a combination of both. We leave it to future research to verify if, and how this will change the effects of knowledge spillovers on cluster productivity.

Finally, we chose a very specific way of modelling cluster externalities and heterogeneity to demonstrate the mechanism we focus on. It may be argued that cluster externalities benefit firms in many more ways than just through firm survival chance, for instance through knowledge spillovers that improve firms' productivity levels, and also that firms are heterogeneous in more dimensions than just productivity. Our results, however, are not restricted to these specific modeling choices. The key point we make is that cluster externalities are likely to benefit firms asymmetrically, impacting cluster composition and, through that, the cluster as a whole. Such asymmetries are relevant to the cluster regardless of the exact nature of the knowledge spillovers or heterogeneity involved. The productivity effect at the cluster level they generate should not be further ignored.

CRediT authorship contribution statement

Cornelis W. Haasnoot: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Resources, Data curation, Writing – original draft, Writing – review & editing, Visualization. **Albert de Vaal:** Conceptualization, Methodology, Validation, Formal analysis, Investigation, Writing – original draft, Writing – review & editing, Supervision, Project administration.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

We would like to thank three anonymous reviewers for their profound and constructive comments on earlier versions of this paper. The usual disclaimer applies.

Appendix

Appendix A. Set-up Melitz-model and derivations benchmark model

In Melitz (2003) the labor requirement of a firm producing a quantity q of a distinct variety is:

$$l = f + q/\varphi \tag{A.1}$$

where $f > 0$ is a fixed overhead cost and marginal costs are $1/\varphi$, with productivity $\varphi > 0$. The total supply of labor is perfectly inelastic and fixed. With a 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 (Melitz, 2003):

$$q(\omega) = Q\left[\frac{p(\omega)}{\widetilde{P}}\right]^{-\sigma} \tag{A.2}$$

[36] In general, the nature of knowledge spillovers in clusters may be identified by using enterprise survey data on innovation, a tool often used in the literature on university-industry collaboration (Perkmann et al., 2013). An example of this is the World Bank Enterprise Survey, which includes questions about the nature of knowledge spillovers. When survey data is limited, insights from the spillover literature may be used to find what type of knowledge spillover is present in a cluster. Different types of knowledge spillovers may be associated with different spillover mechanisms or types of clusters. Knowledge can spill over through different mechanisms, such as the ones that we mentioned in the main text. But also spillover-inhibiting strategies such as patents (e.g., Blind et al., 2006), industrial secrets (e.g. Hall, 1992), imitation cost (e.g., Pavitt, 1987), and lead time (e.g., González-Álvarez and Nieto-Antolín, 2007) could be used to identify the specific type of spillovers. Future research is needed to investigate if these mechanisms are related to specific types of spillovers.

$$r(\omega) = R\left[\frac{p(\omega)}{\widetilde{P}}\right]^{1-\sigma} \tag{A.3}$$

where p denotes price and $R = \widetilde{P}Q$ is aggregate expenditure with $\widetilde{P}$ as the aggregate price level

$$\widetilde{P} = \left[\int_{\omega\in\Omega} p(\omega)^{1-\sigma} d\omega\right]^{\frac{1}{1-\sigma}} \tag{A.4}$$

Operational profits for an individual variety (omitting indices) can then be written as:

$$\pi(\varphi) = q(\varphi)p(\varphi) - w\left(\frac{q}{\varphi} + f\right) \tag{A.5}$$

with w denoting the wage rate, which is normalized to 1. Standard profit maximization gives a firm's optimal price and quantity:

$$p(\varphi) = \frac{1}{\rho\varphi} \tag{A.6}$$

$$q(\varphi) = R\widetilde{P}^{\sigma-1}[\rho\varphi]^{\sigma} \tag{A.7}$$

with $\frac{1}{\rho} \equiv \left(\frac{\sigma}{\sigma-1}\right) > 1$ as the mark-up over marginal cost. Operational profits thus become

$$\pi(\varphi) = \frac{r(\varphi)}{\sigma} - f = \frac{R(\widetilde{P}\rho\varphi)^{\sigma-1}}{\sigma} - f. \tag{A.8}$$

Entry and exit of firms in industry are as specified in the main text, as is the way firms optimally determine their investment in R&D to lower their firm-specific risk.

To determine the cut-off points for market entry and innovation, we write:

$$\frac{r(\varphi_1)}{r(\varphi_2)} = \left(\frac{\varphi_1}{\varphi_2}\right)^{\sigma-1} \tag{A.9}$$

$$\pi(\varphi_2) = \left(\frac{\varphi_2}{\varphi_1}\right)^{\sigma-1}\frac{r(\varphi_1)}{\sigma} - f \tag{A.10}$$

Eq. (4) in the main text is achieved by applying (A.10), while using that $\pi(\varphi^*) = 0$ and acknowledging that profitable entry as an innovative firm also implies payment of the fixed R&D costs.

Productivity levels are drawn from an ex-ante probability density function $g(\varphi)$ and associated cumulative distribution function $G(\varphi)$, implying the following ex-post probability distributions of productivity levels: $\mu(\varphi_H) = \frac{g(\varphi)}{G(\varphi_{RD}^*) - G(\varphi^*)}$ and $\mu(\varphi_{RD}) = \frac{g(\varphi)}{1 - G(\varphi_{RD}^*)}$. Average productivity in the market thus becomes:

$$\widetilde{\varphi}_H(\varphi^*, \varphi_{RD}^*) = \left(\frac{1}{G(\varphi_{RD}^*) - G(\varphi^*)}\int_{\varphi^*}^{\varphi_{RD}^*} \varphi^{\sigma-1} g(\varphi) d\varphi\right)^{\frac{1}{\sigma-1}} \tag{A.11}$$

$$\widetilde{\varphi}_{RD}(\varphi_{RD}^*) = \left(\frac{1}{1 - G(\varphi_{RD}^*)}\int_{\varphi_{RD}^*}^{\infty} \varphi^{\sigma-1} g(\varphi) d\varphi\right)^{\frac{1}{\sigma-1}} \tag{A.12}$$

implying average overall profits of

$$\widetilde{\pi}_H = \pi(\widetilde{\varphi}_H) = \left(\left(\frac{\widetilde{\varphi}_H}{\varphi^*}\right)^{\sigma-1} - 1\right)f \tag{A.13}$$

$$\widetilde{\pi}_{RD} = \pi(\widetilde{\varphi}_{RD}) - f_{RD} = \left(\left(\frac{\widetilde{\varphi}_{RD}}{\varphi^*}\right)^{\sigma-1} - 1\right)f - \overline{f}_{RD} \tag{A.14}$$

where we have applied (A.8) and $\pi(\varphi^*) = 0$. These two equations 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.

Using that the expected value of market entry must be zero, Eq. (5) in the main text, we calculate the average profit level in the market:[37]

[37] Derivation follows from rewriting (A.14) to $\widetilde{\pi}_{RD} = \left(\left(\frac{\widetilde{\varphi}_{RD}}{\widetilde{\varphi}_H}\right)^{\sigma-1}\left(\frac{\widetilde{\varphi}_H}{\varphi^*}\right)^{\sigma-1} - 1\right)f - \overline{f}_{RD}$, so that $\widetilde{\pi}_{RD} = \widetilde{\pi}_H + f\left(\left(\frac{\widetilde{\varphi}_{RD}}{\widetilde{\varphi}_H}\right)^{\sigma-1} - 1\right)\left(\frac{\widetilde{\varphi}_H}{\varphi^*}\right)^{\sigma-1} - \overline{f}_{RD}$. Substituting this in the expression for v_e, and rearranging, gives the expression for $\widetilde{\pi}_H$ in (A.15). Equation (A.16) is obtained by substituting $\widetilde{\pi}_H$ in the expression for $\widetilde{\pi}_{RD}$ above. These unwieldy functions can be used to prove existence of equilibrium.

$$\widetilde{\pi}_H = \frac{f_e - \left(1 - G\left(\varphi^*_{RD}\right)\right)\frac{1}{\delta_{RD}}\left(\left(f\left(\left(\frac{\widetilde{\varphi}_{RD}}{\widetilde{\varphi}_H}\right)^{\sigma-1} - 1\right)\left(\frac{\widetilde{\varphi}_H}{\varphi^*}\right)^{\sigma-1}\right) - \overline{f}_{RD}\right)}{\left(G(\varphi^*_{RD}) - G(\varphi^*)\right)\frac{1}{\delta_H} + \left(1 - G(\varphi^*_{RD})\right)\frac{1}{\delta_{RD}}} \tag{A.15}$$

$$\widetilde{\pi}_{RD} = \frac{f_e + \left(G\left(\varphi^*_{RD}\right) - G(\varphi^*)\right)\frac{1}{\delta_H}\left(\left(f\left(\left(\frac{\widetilde{\varphi}_{RD}}{\widetilde{\varphi}_H}\right)^{\sigma-1} - 1\right)\left(\frac{\widetilde{\varphi}_H}{\varphi^*}\right)^{\sigma-1}\right) - \overline{f}_{RD}\right)}{\left(G(\varphi^*_{RD}) - G(\varphi^*)\right)\frac{1}{\delta_H} + \left(1 - G(\varphi^*_{RD})\right)\frac{1}{\delta_{RD}}}. \tag{A.16}$$

To get the equilibrium values for both cut-off points φ^* and φ^*_{RD}, we set the expected value of market entry equal to zero and use (A.13)-(A.14) to obtain $\left(G(\varphi^*_{RD}) - G(\varphi^*)\right)\left(\left(\frac{\widetilde{\varphi}_H}{\varphi^*}\right)^{\sigma-1} - 1\right)\ \frac{f}{\delta_H} + \left(1 - G(\varphi^*_{RD})\right)\frac{1}{\delta_{RD}}\left[\left(\left(\frac{\widetilde{\varphi}_{RD}}{\varphi^*}\right)^{\sigma-1} - 1\right)f - \overline{f}_{RD}\right] = f_e$. Applying the formulas for average productivity (A.11) and (A.12) yields the free entry condition that is mentioned in the main text:

$$\left(\left(\int_{\varphi^*}^{\varphi^*_{RD}}\left(\frac{\varphi}{\varphi^*}\right)^{\sigma-1} g(\varphi)d\varphi\right) - \left(G\left(\varphi^*_{RD}\right) - G\left(\varphi^*\right)\right)\right)\frac{f}{\delta_H} + \left(\left(\int_{\varphi^*_{RD}}^{\infty}\left(\frac{\varphi}{\varphi^*}\right)^{\sigma-1} g(\varphi)d\varphi\right) - \left(1 - G\left(\varphi^*_{RD}\right)\right)\right)\ \frac{f}{\delta_{RD}} = \left(1 - G\left(\varphi^*_{RD}\right)\right)\frac{\overline{f}_{RD}}{\delta_{RD}} + f_e \tag{A.17}$$

Appendix B. Derivations direct imitation

Derivation $\widehat{\varphi}^*$ and $\widehat{\varphi}^*_{RD}$

Rewrite the free-entry condition (A.17) to

$$\frac{1}{\delta_H(\theta)}\left(\left(\int_{\varphi^*}^{\varphi^*_{RD}}\left(\frac{\varphi}{\varphi^*}\right)^{\sigma-1} g(\varphi)d\varphi\right) - \left(G\left(\varphi^*_{RD}\right) - G\left(\varphi^*\right)\right)\right) + \frac{1}{\delta_{RD}(\theta)}\left(\int_{\varphi^*_{RD}}^{\infty}\left(\frac{\varphi}{\varphi^*}\right)^{\sigma-1} g(\varphi)d\varphi\right) = \frac{1}{\delta_{RD}(\theta)}\left(1 - G\left(\varphi^*_{RD}\right)\right)\left(1 + \frac{\overline{f}_{RD}}{f}\right) + \frac{f_e}{f}$$

Totally differentiate and rearrange:

$$-\frac{\delta'_H}{\delta^2_H}\left[\left(\int_{\varphi^*}^{\varphi^*_{RD}}\left(\frac{\varphi}{\varphi^*}\right)^{\sigma-1} g(\varphi)d\varphi\right) - \left(G\left(\varphi^*_{RD}\right) - G\left(\varphi^*\right)\right)\right]\mathrm{d}\theta$$

$$-\frac{\delta'_{RD}}{\delta^2_{RD}}\left[\int_{\varphi^*_{RD}}^{\infty}\left(\frac{\varphi}{\varphi^*}\right)^{\sigma-1} g(\varphi)d\varphi - \left(1 - G\left(\varphi^*_{RD}\right)\right)\left(1 + \frac{\overline{f}_{RD}}{f}\right)\right]\mathrm{d}\theta$$

$$+(1-\sigma)\varphi^{*1-\sigma}\left[\frac{1}{\delta_H(\theta)}\left(\int_{\varphi^*}^{\varphi^*_{RD}}\varphi^{\sigma-1} g(\varphi)d\varphi\right) + \frac{1}{\delta_{RD}(\theta)}\left(\int_{\varphi^*_{RD}}^{\infty}\varphi^{\sigma-1} g(\varphi)d\varphi\right)\right]\widehat{\varphi}^*$$

$$-\frac{1}{\delta_H(\theta)}\left[1 - \left(\varphi^*_{RD}/\varphi^*\right)^{\sigma-1}\right]dG\left(\varphi^*_{RD}\right) - \frac{1}{\delta_{RD}(\theta)}\left(\frac{\varphi^*_{RD}}{\varphi^*}\right)^{\sigma-1} dG\left(\varphi^*_{RD}\right)$$

$$= -\frac{1}{\delta_{RD}(\theta)}\left(1 + \frac{\overline{f}_{RD}}{f}\right)dG\left(\varphi^*_{RD}\right)$$

Applying $dG(\varphi^*_{RD}) = g(\varphi^*_{RD})\varphi^*_{RD}\widehat{\varphi}^*_{RD}$ and $dG(\varphi^*) = g(\varphi^*)\varphi^*\widehat{\varphi}^*$, using (12) we write

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

with:

$$a = -\frac{\delta'_H}{\delta^2_H}\left(\int_{\varphi^*}^{\varphi^*_{RD}}\left(\frac{\varphi}{\varphi^*}\right)^{\sigma-1} g(\varphi)d\varphi - \left(G\left(\varphi^*_{RD}\right) - G\left(\varphi^*\right)\right)\right)$$

$$-\frac{\delta'_{RD}}{\delta^2_{RD}}\left[\int_{\varphi^*_{RD}}^{\infty}\left(\frac{\varphi}{\varphi^*}\right)^{\sigma-1} g(\varphi)d\varphi - \left(1 - G\left(\varphi^*_{RD}\right)\right)\left(1 + \frac{\overline{f}_{RD}}{f}\right)\right]$$

$$b = \varphi^*_{RD} g\left(\varphi^*_{RD}\right)\frac{\zeta}{(1-\theta)(\sigma-1)\delta_H}\left(1 - \left(\frac{\varphi^*}{\varphi^*_{RD}}\right)^{\sigma-1}\right)\left(\frac{1}{\delta_H}\left(1 - \left(\frac{\varphi^*_{RD}}{\varphi^*}\right)^{\sigma-1}\right) + \frac{1}{\delta_{RD}}\left(\frac{\varphi^*_{RD}}{\varphi^*}\right)^{\sigma-1} - \frac{1}{\delta_{RD}}\left(1 + \frac{\overline{f}_{RD}}{f}\right)\right)$$

$$c = (1-\sigma)\varphi^{*1-\sigma}\left[\frac{1}{\delta_H}\left(\int_{\varphi^*}^{\varphi^*_{RD}}\varphi^{\sigma-1}g(\varphi)d\varphi\right) + \frac{1}{\delta_{RD}}\left(\int_{\varphi^*_{RD}}^{\infty}\varphi^{\sigma-1}g(\varphi)d\varphi\right)\right]$$

$$d = \frac{1}{\delta_H}g(\varphi^*_{RD})\varphi^*_{RD}\left(1-\left(\left(\frac{\varphi^*_{RD}}{\varphi^*}\right)^{\sigma-1}\right)\right)$$

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

Define Z ≡ (a-b)/(c-d-e), so that $\widehat{\varphi}^* = -Zd\theta$.

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

$$a = -b_m{}^{\alpha}\varphi^{*-\alpha}_{RD}\left\{\frac{\delta'_H}{\delta_H^2}\underbrace{\left[\left(\frac{\alpha}{\sigma-\alpha-1}\left(\frac{\varphi^*_{RD}}{\varphi^*}\right)^{\sigma-1}+1\right)-\frac{\sigma-1}{\sigma-\alpha-1}\left(\frac{\varphi^*_{RD}}{\varphi^*}\right)^{\alpha}\right]}_{a1}-\frac{\delta'_{RD}}{\delta_{RD}^2}\underbrace{\left[\frac{\alpha}{\sigma-\alpha-1}\left(\frac{\varphi^*_{RD}}{\varphi^*}\right)^{\sigma-1}+\left(1+\frac{\overline{f}_{RD}}{f}\right)\right]}_{a2}\right\}$$

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

$$c = \frac{(1-\sigma)\varphi^{*1-\sigma}\alpha b_m{}^{\alpha}}{\sigma-\alpha-1}\left[\left(\frac{1}{\delta_H}-\frac{1}{\delta_{RD}}\right)\varphi^{*\,\sigma-\alpha-1}_{RD}-\frac{1}{\delta_H}\varphi^{*\sigma-\alpha-1}\right]\Big\langle 0 \text{ (for } \delta_{RD}<\delta_H)$$

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

$$e = \frac{\alpha}{\delta_{RD}}\left(\frac{b_m}{\varphi^*_{RD}}\right)^{\alpha}\left[\left(\left(\frac{\varphi^*_{RD}}{\varphi^*}\right)^{\sigma-1}-1\right)-\frac{\overline{f}_{RD}}{f}\right].$$

Rewriting eqn. (11) to $\frac{\delta_H-\delta_{RD}}{\delta_H}f\left(\left(\frac{\varphi^*_{RD}}{\varphi^*}\right)^{\sigma-1}-1\right) = \overline{f}_{RD}$, it follows that $b=0$ and $d+e = \alpha\left(\frac{b_m}{\varphi^*_{RD}}\right)^{\alpha}\left[\left(\frac{1}{\delta_{RD}}-\frac{1}{\delta_H}\right)\left(\left(\frac{\varphi^*_{RD}}{\varphi^*}\right)^{\sigma-1}-1\right)-\frac{1}{\delta_{RD}}\frac{\overline{f}_{RD}}{f}\right]=0$, so that $Z=a/c$.

The denominator of Z is always negative ($c<0$) and the sign of Z is determined by the sign of a. Noting that $\sigma-1>0$ and $\sigma-\alpha-1<0$, we can show that:

$$a1\ \langle\ (>)0\ \Leftrightarrow \alpha-(\sigma-1)\rangle(<)\ \alpha\mathscr{R}^{\sigma-1}-(\sigma-1)\mathscr{R}^{\alpha}$$

$$\Leftrightarrow\frac{\mathscr{R}^{\sigma-1}-\mathscr{R}^0}{\sigma-1}>(<)\ \frac{\mathscr{R}^{\alpha}-\mathscr{R}^{\sigma-1}}{\alpha-(\sigma-1)}$$

where $\mathscr{R}\equiv\frac{\varphi^*_{RD}}{\varphi^*}>1$. The left-hand-side gives the average value of the function $\mathscr{R}^x$ over range $[0,\sigma-1]$. The right-hand-side gives the average value of $\mathscr{R}^x$ over $[\sigma-1,\alpha]$. With $\mathscr{R}^x$ exponentially increasing for $\mathscr{R}>1$ and $x>0$, $\alpha>\sigma-1>0$ implies that the left-hand-side is smaller than the right-hand-side and $a1>0$. Furthermore, $a2<0$. Using $\frac{\delta_H-\delta_{RD}}{\delta_H}f\left(\left(\frac{\varphi^*_{RD}}{\varphi^*}\right)^{\sigma-1}-1\right) = \overline{f}_{RD}$, $a2$ can be rewritten to $\left[\frac{\sigma-1}{\sigma-\alpha-1}\left(\frac{\varphi^*_{RD}}{\varphi^*}\right)^{\sigma-1}-\frac{\delta_{RD}}{\delta_H}\left(\left(\frac{\varphi^*_{RD}}{\varphi^*}\right)^{\sigma-1}-1\right)\right]<0$ since $\sigma-1>0$, $\sigma-\alpha-1<0$ and $\frac{\varphi^*_{RD}}{\varphi^*}>1$.

With $a2<0$ and $\delta'_H<0$, $\delta'_{RD}<0$ this implies that $a>0$ and $Z<0$. Hence, $\widehat{\varphi}^* = -Zd\theta>0$ and, applying $\widehat{\varphi}^* = -Zd\theta$ to (12) in the main text, $\widehat{\varphi}^*_{RD}/d\theta>0$.

Derivation $d\left(\frac{M_{RD}}{M}\right)/d\theta$

Using $M_{RD}/M = \frac{\delta_H P_{rd}}{\delta_H P_{RD}+(1-P_{RD})\delta_{RD}}$ and applying the Pareto-distribution so that $dP_{rd} = -\alpha XP_{rd}d\theta$, we get:

$$d\left(\frac{M_{RD}}{M}\right) = -\frac{P_{rd}}{[D_{RD}]^2}[\delta_H\delta_{RD}\alpha X-(1-P_{rd})(\delta'_H\delta_{RD}-\delta_H\delta'_{RD})]d\theta.$$

where $X\equiv\left(\frac{\zeta}{(1-\theta)(\sigma-1)\delta_H}\right)\left(1-\left(\frac{\varphi^*}{\varphi^*_{RD}}\right)^{\sigma-1}\right)>0$ and where D_{RD} denotes the denominator of M_{RD}/M. The sign is determined when applying $\delta_H\equiv\zeta+(1-\zeta)(1-\theta)\varepsilon(0)$ and $\delta_{RD}\equiv\zeta+(1-\zeta)(1-\theta)\varepsilon(\overline{f}_{RD})$:

$$d\left(\frac{M_{RD}}{M}\right) = -\frac{P_{rd}}{[D_{RD}]^2}[\delta_H\delta_{RD}\alpha X+(1-P_{rd})(1-\zeta)\{\zeta(\varepsilon(0)-\varepsilon(\overline{f}_{RD}))\}]d\theta<0$$

since $\varepsilon(0)>\varepsilon(\overline{f}_{RD})$.

Derivation of $\frac{d\widetilde{\varphi}_H}{d\theta}$ *and* $\frac{d\widetilde{\varphi}_{RD}}{d\theta}$

Use (A.11) to obtain

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

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,

$$(\sigma-1)\widehat{\widetilde{\varphi}}_H=\frac{(\sigma-1-\alpha)\left({\varphi_{RD}^*}^{\sigma-\alpha-1}\widehat{\varphi}_{RD}^*-{\varphi^*}^{\sigma-\alpha-1}\widehat{\varphi}^*\right)}{{\varphi_{RD}^*}^{\sigma-\alpha-1}-{\varphi^*}^{\sigma-\alpha-1}}-\frac{\alpha\left({\varphi_{RD}^*}^{-\alpha}\widehat{\varphi}_{RD}^*-{\varphi^*}^{-\alpha}\widehat{\varphi}^*\right)}{{\varphi^*}^{-\alpha}-{\varphi_{RD}^*}^{-\alpha}}.$$

and applying (12):

$$\widehat{\widetilde{\varphi}}_H=\widehat{\varphi}^*+\frac{\zeta\left[1-\left(\varphi^*/\varphi_{RD}^*\right)^{\sigma-1}\right]}{(1-\theta)(\sigma-1)^2\delta_H}\left[\frac{\alpha}{1-\left(\varphi_{RD}^*/\varphi^*\right)^{\alpha}}-\frac{\alpha-(\sigma-1)}{1-\left(\varphi_{RD}^*/\varphi^*\right)^{\alpha-(\sigma-1)}}\right]d\theta.$$

In this equation the first term on the RHS is the effect of the proportional change of φ^* and φ_{RD}^*, which is positive since $\widehat{\varphi}^*>0$. The second term on the RHS is positive since

$$\frac{\mathscr{R}^{\alpha}-\mathscr{R}^{0}}{\alpha}>\frac{\mathscr{R}^{\alpha-(\sigma-1)}-\mathscr{R}^{0}}{\alpha-(\sigma-1)}$$

where $\mathscr{R}\equiv\frac{\varphi_{RD}^*}{\varphi^*}>1$ and $\mathscr{R}^x$ is exponentially increasing in x when $x>0$ and $\alpha>\sigma-1>0$.

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

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

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

Applying the Pareto-distribution gives $\widehat{\widetilde{\varphi}}_{RD}=\widehat{\varphi}_{RD}^*>0$.

Appendix C. Derivations indirect imitation

Derivation of $\widehat{\varphi}^*$

Recognizing that with indirect imitation in the right-hand-side of the free entry condition (A.17) features $(1-\theta)\overline{f}_{RD}$ rather than $\overline{f}_{RD}$, we rewrite the free-entry condition to

$$\left(\frac{1}{\delta_H}\left(\int\limits_{\varphi^*}^{\varphi_{RD}^*}\left(\frac{\varphi}{\varphi^*}\right)^{\sigma-1}g(\varphi)d\varphi\right)-\left(G(\varphi_{RD}^*)-G(\varphi^*)\right)\right)+\frac{1}{\delta_{RD}}\left(\left(\int\limits_{\varphi_{RD}^*}^{\infty}\left(\frac{\varphi}{\varphi^*}\right)^{\sigma-1}g(\varphi)d\varphi\right)\right)=\frac{1}{\delta_{RD}}\left(1-G(\varphi_{RD}^*)\right)\left(1+\frac{(1-\theta)\overline{f}_{RD}}{f}\right)+\frac{f_e}{f}.$$

Totally differentiate and rearrange:

$$(1-\sigma){\varphi^*}^{1-\sigma}\left[\frac{1}{\delta_H}\left(\int\limits_{\varphi^*}^{\varphi_{RD}^*}\varphi^{\sigma-1}g(\varphi)d\varphi\right)+\frac{1}{\delta_{RD}}\left(\int\limits_{\varphi_{RD}^*}^{\infty}\varphi^{\sigma-1}g(\varphi)d\varphi\right)\right]\widehat{\varphi}^*$$

$$-\frac{1}{\delta_H}\left[1-\left(\varphi_{RD}^*/\varphi^*\right)^{\sigma-1}\right]dG(\varphi_{RD}^*)-\frac{1}{\delta_{RD}}\left(\frac{\varphi_{RD}^*}{\varphi^*}\right)^{\sigma-1}dG(\varphi_{RD}^*)$$

$$=-\frac{1}{\delta_{RD}}\left(1+\frac{(1-\theta)\overline{f}_{RD}}{f}\right)dG(\varphi_{RD}^*)-\frac{1}{\delta_{RD}}\left(1-G(\varphi_{RD}^*)\right)\frac{\overline{f}_{RD}}{f}d\theta$$

Apply $dG(\varphi_{RD}^*)=g(\varphi_{RD}^*)\varphi_{RD}^*\widehat{\varphi}_{RD}^*$ and $dG(\varphi^*)=g(\varphi^*)\varphi^*\widehat{\varphi}^*$ and use (16) to get,

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

with:

$$a = \frac{\left(1 - G\left(\varphi^*_{RD}\right)\right)}{\delta_{RD}} \frac{\overline{f}_{RD}}{f} > 0$$

$$b = -\frac{\varphi^*_{RD} g\left(\varphi^*_{RD}\right)}{\delta_{RD}} \left(\frac{\overline{f}_{RD}/f}{(1-\zeta)(\sigma-1)} \left(\frac{\varphi^*}{\varphi^*_{RD}}\right)^{\sigma-1} \right) \left(\left(1 - \frac{\delta_{RD}}{\delta_H}\right) \left(\left(\frac{\varphi^*_{RD}}{\varphi^*}\right)^{\sigma-1} - 1 \right) - \frac{(1-\theta)\overline{f}_{RD}}{f} \right)$$

$$c = (1-\sigma)\varphi^{*1-\sigma} \left[\frac{1}{\delta_H} \left(\int_{\varphi^*}^{\varphi^*_{RD}} \varphi^{\sigma-1} g(\varphi) d\varphi \right) + \frac{1}{\delta_{RD}} \left(\int_{\varphi^*_{RD}}^{\infty} \varphi^{\sigma-1} g(\varphi) d\varphi \right) \right]$$

$$d = g\left(\varphi^*_{RD}\right)\varphi^*_{RD} \frac{1}{\delta_H} \left(1 - \left(\left(\frac{\varphi^*_{RD}}{\varphi^*}\right)^{\sigma-1} \right) \right)$$

$$e = \frac{1}{\delta_{RD}} \left(\left(\left(\frac{\varphi^*_{RD}}{\varphi^*}\right)^{\sigma-1} - 1 \right) - \frac{(1-\theta)\overline{f}_{RD}}{f} \right) g\left(\varphi^*_{RD}\right)\varphi^*_{RD} \widehat{\varphi}^*.$$

Define $Z' \equiv$ (a-b)/(c-d-e), so that $\widehat{\varphi}^* = -Z' d\theta$.

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

$$a = \frac{1}{\delta_{RD}} \left(\frac{b_m}{\varphi^*_{RD}}\right)^{\alpha} \frac{\overline{f}_{RD}}{f} > 0$$

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

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

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

$$e = \frac{\alpha}{\delta_{RD}} \left(\frac{b_m}{\varphi^*_{RD}}\right)^{\alpha} \left(\left(\left(\frac{\varphi^*_{RD}}{\varphi^*}\right)^{\sigma-1} - 1 \right) - \frac{(1-\theta)\overline{f}_{RD}}{f} \right) \lesseqgtr 0.$$

Furthermore, when applying eqn. (15) it follows that $b = 0$ and $d + e = \frac{\alpha}{\delta_{RD}} \left(\frac{b_m}{\varphi^*_{RD}}\right)^{\alpha} \left[\left(1 - \frac{\delta_{RD}}{\delta_H}\right) \left(\left(\frac{\varphi^*_{RD}}{\varphi^*}\right)^{\sigma-1} - 1 \right) - \frac{(1-\theta)\overline{f}_{RD}}{f} \right] = 0$. Hence $Z' = a/c < 0$ and $\widehat{\varphi}^* = -Z' d\theta > 0$ for $d\theta > 0$.

Applying $\widehat{\varphi}^* = -Z' d\theta$ to (16) in the main text, using $Z' = a/c$, $\mathscr{R} \equiv \varphi^*_{RD}/\varphi^*$ and (15), gives

$$\widehat{\varphi}^*_{RD} = \frac{\left(1 - \mathscr{R}^{1-\sigma}\right)}{(\sigma-1)(1-\theta)} \left[\frac{\alpha - (\sigma - 1)}{\alpha} \left(\frac{1}{1 + \frac{\delta_{RD}}{\delta_H - \delta_{RD}} \mathscr{R}^{\alpha-(\sigma-1)}} \right) - 1 \right] d\theta.$$

All terms but for the bracketed term in this expression are positive, hence $\widehat{\varphi}^*_{RD} < 0$ when $d\theta > 0$.

(The bracketed term in this expression is negative when $\frac{\alpha}{\alpha} - \frac{(\sigma-1)}{\alpha} < 1 + \frac{\delta_{RD}}{\delta_H - \delta_{RD}} \mathscr{R}^{\alpha-(\sigma-1)}$, which is the case for $\delta_H > \delta_{RD}$).

Derivation $d\left(\frac{M_{RD}}{M}\right)/d\theta$

Using $M_{RD}/M = \frac{\delta_H P_{rd}}{\delta_H P_{RD} + (1-P_{RD})\delta_{RD}}$ and applying the Pareto-distribution so that $dP_{rd} = \alpha X' P_{rd} d\theta$ we get:

$$d\left(\frac{M_{RD}}{M}\right) = \frac{\alpha X' P_{rd} \delta_H \delta_{RD}}{\left[\delta_H P_{RD} + (1 - P_{RD})\delta_{RD}\right]^2} d\theta > 0.$$

where $X' \equiv \left(\frac{1}{(1-\theta)(\sigma-1)}\right)\left(1 - \left(\frac{\varphi^*}{\varphi^*_{RD}}\right)^{\sigma-1}\right) > 0.$

Derivation of $\frac{d\widetilde{\varphi}_H}{d\theta}$ *and* $\frac{d\widetilde{\varphi}_{RD}}{d\theta}$

As before, use (A.11) and the Pareto-distribution to obtain,

$$(\sigma-1)\widehat{\widetilde{\varphi}}_H=\frac{(\sigma-1-\alpha)\left(\varphi_{RD}^{*\ \sigma-\alpha-1}\widehat{\varphi}_{RD}^{*}-\varphi^{*\sigma-\alpha-1}\widehat{\varphi}^{*}\right)}{\varphi_{RD}^{*\ \sigma-\alpha-1}-\varphi^{*\sigma-\alpha-1}}-\frac{\alpha\left(\varphi_{RD}^{*\ -\alpha}\widehat{\varphi}_{RD}^{*}-\varphi^{*-\alpha}\widehat{\varphi}^{*}\right)}{\varphi^{*-\alpha}-\varphi_{RD}^{*\ -\alpha}}.$$

Applying (16) and rearranging gives

$$\widehat{\widetilde{\varphi}}_H=\widehat{\varphi}^*-\frac{\left[1-\left(\varphi^*/\varphi_{RD}^*\right)^{\sigma-1}\right]}{(1-\theta)(\sigma-1)^2}\left[\frac{\alpha}{1-(\varphi_{RD}^*/\varphi^*)^{\alpha}}-\frac{\alpha-(\sigma-1)}{1-(\varphi_{RD}^*/\varphi^*)^{\alpha-(\sigma-1)}}\right]d\theta.$$

The first term on the RHS is positive since $\widehat{\varphi}^*>0$, while the second term on the RHS is negative. The overall effect on $\widetilde{\varphi}_H$ is therefore ambiguous.

Similarly, we can derive that $\frac{\widehat{\widetilde{\varphi}}_{RD}}{d\theta}<0$:

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

Applying the Pareto-distribution and rearranging gives $\widehat{\widetilde{\varphi}}_{RD}=\widehat{\varphi}_{RD}^*<0$.

Appendix D. Mathematical derivation density distribution changes

To determine how knowledge spillovers change the distribution of firms we first introduce a measure showing the overrepresentation of innovative firms – and likewise the underrepresentation of non-innovative firms – in comparison to the original Melitz-model. This is accomplished by dividing (7) in the main text by P_{RD} (which is M_{RD}/M in the Melitz model):

$$D_{RD}=\left(\frac{\delta_H}{\delta_H P_{RD}+(1-P_{RD})\delta_{RD}}\right)>1\text{ and }D_H=\left(\frac{\delta_{RD}}{\delta_H P_{RD}+(1-P_{RD})\delta_{RD}}\right)<1,$$

as measures of overrepresentation of innovative firms and underrepresentation of non-innovative firms, respectively. Consequently, $\frac{dD_H}{dP_{rd}}=-\frac{D_H(\delta_H-\delta_{RD})}{\delta_H P_{RD}+(1-P_{RD})\delta_{RD}}<0$, $\frac{dD_{RD}}{dP_{rd}}=-\frac{D_{RD}(\delta_H-\delta_{RD})}{\delta_H P_{RD}+(1-P_{RD})\delta_{RD}}<0$ and $\frac{\partial\mu(\varphi)}{\partial\varphi^*}=\frac{\alpha\mu(\varphi)}{\varphi^*}>0$ for both cases of knowledge spillovers.

For *direct imitation*, we need to establish

- for $\varphi_2^*\le\varphi<\varphi_{RD}^*$: $\frac{d\mu(\varphi)}{d\theta}=\frac{\partial\mu(\varphi)}{\partial\varphi^*}\frac{d\varphi^*}{d\theta}+\frac{\partial\mu(\varphi)}{\partial D_H}\frac{dD_H}{d\theta}$
- for $\varphi\ge\varphi_{RD,2}^*$: $\frac{d\mu(\varphi)}{d\theta}=\frac{\partial\mu(\varphi)}{\partial\varphi^*}\frac{d\varphi^*}{d\theta}+\frac{\partial\mu(\varphi)}{\partial D_{RD}}\frac{dD_{RD}}{d\theta}$

Knowing that $dP_{rd}=-\alpha XP_{rd}d\theta<0$, $\frac{dD_H}{d\delta_H}=\frac{-\delta_{RD}P_{RD}}{[\delta_H P_{RD}+(1-P_{RD})\delta_{RD}]^2}<0$, $\frac{dD_{RD}}{d\delta_H}=\frac{\delta_{RD}(1-P_{RD})}{[\delta_H P_{RD}+(1-P_{RD})\delta_{RD}]^2}>0$, $d\delta_{RD}/d\theta=0$ and $\frac{d\delta_H}{d\theta}=-(1-\zeta)$, we derive $\frac{dD_H}{d\theta}=\frac{D_H P_{RD}[(1-\zeta)+(\delta_H-\delta_{RD})\alpha X]}{\delta_H P_{RD}+(1-P_{RD})\delta_{RD}}>0$ and $\frac{dD_{RD}}{d\theta}=\frac{D_{RD}(\delta_H-\delta_{RD})\alpha XP_{rd}-D_H(1-P_{RD})(1-\zeta)}{\delta_H P_{RD}+(1-P_{RD})\delta_{RD}}\lesseqgtr 0$.

Consequently, for $\varphi_2^*\le\varphi<\varphi_{RD}^*$: $\frac{d\mu(\varphi)}{d\theta}=\frac{\alpha\mu(\varphi)}{\varphi^*}\cdot\underbrace{\frac{d\varphi^*}{d\theta}}_{>0}+\frac{g(\varphi)}{1-G(\varphi^*)}\cdot\underbrace{\frac{dD_H}{d\theta}}_{>0}>0$ and for $\varphi\ge\varphi_{RD,2}^*$: $\frac{d\mu(\varphi)}{d\theta}=\frac{\alpha\mu(\varphi)}{\varphi^*}\cdot\underbrace{\frac{d\varphi^*}{d\theta}}_{>0}+\frac{g(\varphi)}{1-G(\varphi^*)}\cdot\underbrace{\left(\frac{dD_{RD}}{d\theta}\right)}_{\lesseqgtr 0}\lesseqgtr 0$.

For those firms with productivity levels $\varphi_{RD}^*\le\varphi<\varphi_{RD,2}^*$ (the 'switchers') the effect is that their representation changes from D_{RD} to $D_H+dD_H=D_H'$ (discrete change). Their density will be lower (higher) if $\mathrm{D_H'}(\varphi_{RD}^*)-\mathrm{D_{RD}}(\varphi_{RD}^*)<(>)0$. Using D_{RD} and $D_H'=\frac{\delta_{RD}}{\delta_H}D_{RD}+\left[\frac{\delta_{RD}P_{RD}(1-\zeta)}{[\delta_H P_{RD}+(1-P_{RD})\delta_{RD}]^2}+\frac{D_H(\delta_H-\delta_{RD})\alpha XP_{rd}}{\delta_H P_{RD}+(1-P_{RD})\delta_{RD}}\right]d\theta$, we get $\mathrm{D_H'}(\varphi_2^*)-\mathrm{D_{RD}}(\varphi_2^*)=\underbrace{\left(\frac{\delta_{\mathrm{RD}}}{\delta_{\mathrm{H}}}-1\right)}_{<0}\mathrm{D_{RD}}+\underbrace{\left[\frac{\delta_{\mathrm{RD}}\mathrm{P_{RD}}(1-\zeta)}{[\delta_{\mathrm{H}}\mathrm{P_{RD}}+(1-\mathrm{P_{RD}})\delta_{\mathrm{RD}}]^2}+\frac{\mathrm{D_H}(\delta_{\mathrm{H}}-\delta_{\mathrm{RD}})\alpha\mathrm{XP_{rd}}}{\delta_{\mathrm{H}}\mathrm{P_{RD}}+(1-\mathrm{P_{RD}})\delta_{\mathrm{RD}}}\right]}_{>0}d\theta\lesseqgtr 0$

For *indirect imitation*, we need to establish

- for $\varphi_2^*\le\varphi<\varphi_{RD,2}^*$: $\frac{d\mu(\varphi)}{d\theta}=\frac{\partial\mu(\varphi)}{\partial\varphi^*}\frac{d\varphi^*}{d\theta}+\frac{\partial\mu(\varphi)}{\partial D_H}\frac{dD_H}{d\theta}$
- for $\varphi\ge\varphi_{RD}^*$: $\frac{d\mu(\varphi)}{d\theta}=\frac{\partial\mu(\varphi)}{\partial\varphi^*}\frac{d\varphi^*}{d\theta}+\frac{\partial\mu(\varphi)}{\partial D_{RD}}\frac{dD_{RD}}{d\theta}$

With indirect imitation $dP_{rd}=\alpha X'P_{rd}d\theta>0$, $d\delta_{RD}/d\theta=0$ and $d\delta_H/d\theta=0$, implying $\frac{dD_H}{d\theta}=-\frac{D_H(\delta_H-\delta_{RD})\alpha X'P_{rd}}{\delta_H P_{RD}+(1-P_{RD})\delta_{RD}}<0$ and $\frac{dD_{RD}}{d\theta}=-\frac{D_{RD}(\delta_H-\delta_{RD})\alpha X'P_{rd}}{\delta_H P_{RD}+(1-P_{RD})\delta_{RD}}<0$. This yields, for $\varphi_2^*\le\varphi<\varphi_{RD,2}^*$: $\frac{d\mu(\varphi)}{d\theta}=\frac{\alpha\mu(\varphi)}{\varphi^*}\cdot\underbrace{\frac{d\varphi^*}{d\theta}}_{>0}+\frac{g(\varphi)}{1-G(\varphi^*)}\cdot\underbrace{\frac{dD_H}{d\theta}}_{<0}\lesseqgtr 0$ and for $\varphi\ge\varphi_{RD}^*$: $\frac{d\mu(\varphi)}{d\theta}=\frac{\alpha\mu(\varphi)}{\varphi^*}\cdot\underbrace{\frac{d\varphi^*}{d\theta}}_{>0}+\frac{g(\varphi)}{1-G(\varphi^*)}\cdot\underbrace{\frac{dD_{RD}}{d\theta}}_{<0}\lesseqgtr 0$.

For the firms with productivity levels $\varphi_{RD,\,2}^*\le\varphi<\varphi_{RD}^*$ the effect is that their representation changes from D_H to $D_{RD}+dD_{RD}=D_{RD}'$ (discrete change). Their density will be higher (lower) if $D_{RD}'(\varphi_{RD}^*)-D_H(\varphi_{RD}^*)>(<)0$. Using D_H and $D_{RD}'=\frac{\delta_H}{\delta_{RD}}D_H-\frac{D_{RD}(\delta_H-\delta_{RD})\alpha X'P_{rd}}{\delta_H P_{RD}+(1-P_{RD})\delta_{RD}}d\theta$, we derive $D_{RD}'(\varphi_{RD}^*)-D_H(\varphi_{RD}^*)=\underbrace{\left(\frac{\delta_H}{\delta_{RD}}-1\right)}_{>0}D_H-\underbrace{\left[\frac{D_{RD}(\delta_H-\delta_{RD})\alpha X'P_{rd}}{\delta_H P_{RD}+(1-P_{RD})\delta_{RD}}\right]}_{>0}d\theta\lesseqgtr 0$.

Appendix E. Motivation baseline parameters

In this appendix we provide an extensive motivation for the choice of parameters. For each parameter, we will first outline how we decided on the baseline value to use. After that, we explain how we established the country- and sector-specific values for each parameter.

- *Elasticity of substitution σ and shape parameter α*: σ is typically estimated between 5 and 10 (Bergstrand et al., 2013; Brakman et al., 2005), though there are also papers with estimates of σ between 2 and 5 (e.g., Del Gatto et al. 2008). We follow the calibration used by Ghironi and Melitz (2005), where σ and α are linked through the standard deviation of plant sales. Calibrating their model with U.S. data yields estimates of 3.8 for σ and 3.4 for α. This means that the coefficient for σ is lower than found in previous papers, and that the markup of price over marginal cost is therefore significantly larger in our model. However, most models do not include fixed costs, and therefore have no difference between markup over marginal and markup over average costs. If one looks at markup over average costs instead, the calibration we use is much more similar. For our specification, we set σ at 4 and α at 3.5. It must be noted that our configurations always satisfy $\alpha > 1$ which is required to have a finite mean in productivity levels. Del Gatto et al. (2008) provide sector-specific estimates of σ and α for various industries. We use these for our sector-specific parameter values. However, their estimates of α violate our restriction that $\alpha + 1 - \sigma > 0$. As such, rather than using their numbers directly, we calculate for each industry the number of standard deviations the α of that industry is different from the mean. Having calculated how many standard deviations from the mean each industry is, we then translate that into $\alpha + 1 - \sigma$. As an example, Del Gatto et al. find an α of 1.466 for the apparel sector. This is 0.94 standard deviations below the average α in their dataset. We translate this into an $\alpha + 1 - \sigma$ of 0.26 (which is 0.94 standard deviations below the average of 0.5 that we use in our simulations).
- *Firm exit rate ζ*: Dunne et al. (1988) look at firm exit rates over time in U.S. manufacturing industries. They find that on average, every year about 8% of firms exit the market, numbers confirmed by Agarwal and Gort (1996). We normalize the exit chance of non-innovative firms to 100%, so we are not interested in this aggregate exit rate. Rather we want to look at the boost to survival odds that is given to innovators relative to non-innovators. Cefis and Marsili (2005) look at survival rates of innovative and non-innovative firms and find that, after the first 8 years of their life, 26.3% of non-innovative firms have exited the market, against 22.3% of innovative firms. So, if non-innovators are sure to exit the market, as we assume in our modeling set-up, innovative firms must have an exit rate of 84.7%. This means that the normalized value of ζ must in fact be fairly high, probably between 0.6 and 0.9 in order to generate realistic distribution of innovative firms versus non-innovators. In our baseline specification, we set this normalized ζ equal to 0.75. The OECD does not provide data on the survival rates of innovative versus non-innovative firms. As such, we must use some other kind of proxy for this. Previous research has found that innovative firms are, on average, larger than non-innovative firms (Cefis and Marsili, 2005; Baptista and Swann, 1998). We use this stylized fact to approximate ζ. From the OECD, we gather data on the exit rates of firms of different sizes, then calculate the implied ζ from there. In this case, we compare the exit rate of all employer firms to the exit rate of firms with 1-4 employees. We do not use the exit rate for all firms because these are biased upwards by the inclusion of 0-employee firms, which have much higher exit rates than other firms. Furthermore, 0-employee firms do not exist in the model as we have constructed it, so they must be excluded.
- *Ratio of R&D costs to fixed costs $\bar{f}_{RD}/f$:* for $\bar{f}_{RD}/f$ we use industry-level data on various expenditures of firms. We use data from the OECD, with supplementary data on the Netherlands from the Dutch Central Bureau of Statistics (CBS). The first important choice here is which variables to use for both $\bar{f}_{RD}$ and f. According to the model, this is the ratio of R&D labor costs to total labor costs. These data are available from the CBS, but not the OECD, which does not have data on R&D labor costs. As such, we use the CBS data to establish a baseline value for $\bar{f}_{RD}/f$ and the OECD data for looking at individual clusters. For these OECD data, we use total R&D expenditure rather than R&D labor expenditure. This will overstate the value of $\bar{f}_{RD}/f$ somewhat in comparison to the CBS data, but it is the best that is available.

  The CBS data indicate that our measurement for $\bar{f}_{RD}/f$ varies from about 0.09% for the wholesale and retail trade industry to 21.82% for the industry for electrical machines, suggesting a non-weighted average $\bar{f}_{RD}/f$for Dutch industries of 0.06. There are three reasons, however, why this estimate is on the low side. Firstly, the percentages named before include non-innovative firms, which are excluded from $\bar{f}_{RD}/f$ in our model. Secondly, our model mostly has implications for industries where innovation plays a significant role in, so we must look at industries with a significant amount of innovative activity, such as electrical machines, rather than industries with very low innovative activity, like the wholesale trade industry. Thirdly, the parameter $\bar{f}_{RD}/f$ reflects the *maximum* level of innovative activity, not the *average* level. As such, the statistics we found underestimate what our parameter should be and we therefore use 0.25 as baseline value for $\bar{f}_{RD}/f$.
- *Extent of knowledge spillovers θ:* Regarding the baseline value for θ, many papers have tried to find empirical evidence for knowledge spillovers, usually from FDI. However, few papers provide numbers that correspond to what we mean by θ. For example, Todo and Miyamoto (2006) find that knowledge spillovers from foreign firms to domestic firms cause TFP growth rates of domestic firms to become anywhere from 2 to 5 times as high. While interesting, it is difficult to translate this into a value of the θ parameter, as θ denotes the percentage of a technology that is passed on. For this, Haddad and Harrison (1993) provide a useful indication. They find that if the share of foreign firms in domestic industries increases by 10 per cent, local firms get about 4 per cent closer to the best practice, the most productive they can be. This suggests a value of 0.4 for θ. We note that Haddad and Harrison explicitly look at knowledge spillovers from highly productive foreign firms to less productive local firms. This may cause both an overestimation of knowledge spillovers because there is a large productivity difference to overcome but also an underestimation as the absorptive capacity of the low-productive firms may be lower in their situation than in the industries we try to model here. Regardless, 0.4 seems like a good baseline value to start with. Since Haddad and Harrison do not provide any sector-specific indications, we use 0.4 for all the other sectors as well.

Appendix F. Extensive simulation results

Tables A.1, A.2 and A.3

Table A.1
Comparing outcomes across ranges of parameters.

θ	**Direct**	0-0.2	0.2-0.4	0.4-0.6	0.6-0.9	**Indirect**	0-0.2	0.2-0.4	0.4-0.6	0.6-0.9	**Hybrid**	0-0.2	0.2-0.4	0.4-0.6	0.6-0.9
$d\widetilde{\varphi}_{total}$	-0.124	-0.079	-0.096	-0.128	-0.197	0.061	0.048	0.054	0.063	0.077	-0.025	-0.024	-0.024	-0.025	-0.027
$d\widetilde{\varphi}_{direct}$	-0.124	-0.079	-0.096	-0.128	-0.197	-	-	-	-	-	-0.116	-0.070	-0.084	-0.111	-0.209
$d\widetilde{\varphi}_{indirect}$	-	-	-	-	-	0.061	0.048	0.054	0.063	0.077	0.099	0.054	0.067	0.094	0.191
Non-innovator effect	0.080	0.028	0.041	0.069	0.197	-0.025	-0.024	-0.025	-0.025	-0.024	-0.002	-0.002	-0.002	-0.002	-0.002
NI-effect (unweighted)	0.108	0.046	0.062	0.098	0.244	-0.055	-0.043	-0.049	-0.057	-0.073	-0.004	-0.004	-0.004	-0.004	-0.005
Innovator effect	0.235	0.123	0.157	0.227	0.457	-0.197	-0.142	-0.163	-0.202	-0.284	-0.027	-0.026	-0.026	-0.027	-0.028
I-effect (unweighted)	1.392	0.390	0.578	1.032	4.001	-0.396	-0.361	-0.379	-0.400	-0.444	-0.063	-0.060	-0.062	-0.063	-0.066
Composition effect	-0.440	-0.230	-0.294	-0.424	-0.851	0.282	0.213	0.242	0.290	0.385	0.003	0.003	0.003	0.003	0.004
C-effect (unweighted)	-0.392	-0.246	-0.299	-0.396	-0.638	0.372	0.247	0.295	0.381	0.573	0.003	0.003	0.003	0.003	0.004
Observations	131,611	16,186	48,334	40,242	26,849	131,611	16,186	48,334	40,242	26,849	131,611	16,186	48,334	40,242	26,849
ζ	**Direct**	0.35-0.6	0.6-0.7	0.7-0.8	0.8-1	**Indirect**	0.35-0.6	0.6-0.7	0.7-0.8	0.8-1	**Hybrid**	0.35-0.6	0.6-0.7	0.7-0.8	0.8-1
$d\widetilde{\varphi}_{total}$	-0.124	-0.142	-0.140	-0.130	-0.099	0.061	0.013	0.036	0.062	0.092	-0.025	-0.043	-0.034	-0.025	-0.013
$d\widetilde{\varphi}_{direct}$	-0.124	-0.142	-0.140	-0.130	-0.099	-	-	-	-	-	-0.116	-0.093	-0.109	-0.121	-0.121
$d\widetilde{\varphi}_{indirect}$	-	-	-	-	-	0.061	0.013	0.036	0.062	0.092	0.099	0.062	0.085	0.104	0.112
Non-innovator effect	0.080	0.043	0.061	0.080	0.106	-0.025	-0.008	-0.013	-0.022	-0.043	-0.002	-0.002	-0.002	-0.002	-0.001
NI-effect (unweighted)	0.108	0.073	0.092	0.109	0.130	-0.055	-0.032	-0.041	-0.053	-0.077	-0.004	-0.007	-0.006	-0.004	-0.002
Innovator effect	0.235	0.207	0.231	0.242	0.237	-0.197	-0.146	-0.169	-0.196	-0.234	-0.027	-0.044	-0.036	-0.027	-0.015
I-effect (unweighted)	1.392	0.590	0.885	1.251	2.200	-0.396	-0.199	-0.260	-0.355	-0.615	-0.063	-0.072	-0.068	-0.064	-0.054
Composition effect	-0.440	-0.392	-0.432	-0.451	-0.442	0.282	0.167	0.218	0.280	0.369	0.003	0.004	0.004	0.004	0.003
C-effect (unweighted)	-0.392	-0.436	-0.436	-0.406	-0.324	0.372	0.260	0.321	0.380	0.431	0.003	0.005	0.004	0.004	0.002
Observations	131,611	9,049	32,894	51,132	38,532	131,611	9,049	32,894	51,132	38,532	131,611	9,049	32,894	51,132	38,532
$\bar{f}_{RD}$ /f	**Direct**	0-0.2	0.2-0.4	0.4-0.6	0.6-1	**Indirect**	0-0.2	0.2-0.4	0.4-0.6	0.6-1	**Hybrid**	0-0.2	0.2-0.4	0.4-0.6	0.6-1
$d\widetilde{\varphi}_{total}$	-0.124	-0.119	-0.139	-0.124	-0.102	0.061	0.033	0.070	0.092	0.108	-0.025	-0.019	-0.030	-0.031	-0.028
$d\widetilde{\varphi}_{direct}$	-0.124	-0.119	-0.139	-0.124	-0.102	-	-	-	-	-	-0.116	-0.084	-0.137	-0.148	-0.142
$d\widetilde{\varphi}_{indirect}$	-	-	-	-	-	0.061	0.033	0.070	0.092	0.108	0.099	0.070	0.117	0.128	0.125
Non-innovator effect	0.080	0.048	0.094	0.114	0.123	-0.025	-0.008	-0.027	-0.044	-0.060	-0.002	-0.001	-0.002	-0.003	-0.002
NI-effect (unweighted)	0.108	0.082	0.122	0.135	0.139	-0.055	-0.031	-0.063	-0.083	-0.097	-0.004	-0.003	-0.005	-0.005	-0.004
Innovator effect	0.235	0.223	0.256	0.240	0.212	-0.197	-0.126	-0.228	-0.274	-0.295	-0.027	-0.020	-0.031	-0.034	-0.034
I-effect (unweighted)	1.392	0.676	1.501	2.240	2.924	-0.396	-0.176	-0.424	-0.650	-0.903	-0.063	-0.030	-0.068	-0.101	-0.135
Composition effect	-0.440	-0.391	-0.489	-0.477	-0.436	0.282	0.166	0.324	0.412	0.463	0.003	0.001	0.004	0.006	0.007
C-effect (unweighted)	-0.392	-0.443	-0.403	-0.315	-0.241	0.372	0.263	0.435	0.484	0.478	0.003	0.002	0.004	0.005	0.005
Observations	131,611	57,148	41,751	20,122	12,590	131,611	57,148	41,751	20,122	12,590	131,611	57,148	41,751	20,122	12,590
σ	**Direct**	2-3	3-4	4-5	5-7	**Indirect**	2-3	3-4	4-5	5-7	**Hybrid**	2-3	3-4	4-5	5-7
$d\widetilde{\varphi}_{total}$	-0.124	-0.110	-0.139	-0.134	-0.120	0.061	0.030	0.070	0.077	0.075	-0.025	-0.036	-0.026	-0.020	-0.016
$d\widetilde{\varphi}_{direct}$	-0.124	-0.110	-0.139	-0.134	-0.120	-	-	-	-	-	-0.116	-0.089	-0.131	-0.132	-0.121
$d\widetilde{\varphi}_{indirect}$	-	-	-	-	-	0.061	0.030	0.070	0.077	0.075	0.099	0.071	0.111	0.115	0.108
Non-innovator effect	0.080	0.116	0.082	0.064	0.048	-0.025	-0.035	-0.025	-0.020	-0.016	-0.002	-0.000	-0.002	-0.003	-0.002
NI-effect (unweighted)	0.108	0.151	0.111	0.089	0.069	-0.055	-0.073	-0.058	-0.048	-0.038	-0.004	-0.002	-0.005	-0.005	-0.005
Innovator effect	0.235	0.424	0.218	0.145	0.099	-0.197	-0.332	-0.190	-0.133	-0.093	-0.027	-0.042	-0.027	-0.019	-0.014
I-effect (unweighted)	1.392	3.032	1.057	0.622	0.384	-0.396	-0.721	-0.363	-0.243	-0.165	-0.063	-0.108	-0.060	-0.041	-0.029
Composition effect	-0.440	-0.650	-0.438	-0.343	-0.267	0.282	0.397	0.285	0.231	0.184	0.003	0.007	0.003	0.002	0.001
C-effect (unweighted)	-0.392	-0.386	-0.391	-0.394	-0.397	0.372	0.394	0.371	0.362	0.354	0.003	0.006	0.003	0.002	0.002

(*continued on next page*)

Table A.1 (*continued*)

Observations	131,611	39,299	33,319	25,835	33,158	131,611	39,299	33,319	25,835	33,158	131,611	39,299	33,319	25,835	33,158
$\alpha + 1 - \sigma$	**Direct**	0-0.3	0.3-0.6	0.6-0.9	0.9-3	**Indirect**	0-0.3	0.3-0.6	0.6-0.9	0.9-3	**Hybrid**	0-0.3	0.3-0.6	0.6-0.9	0.9-3
$d\widetilde{\varphi}_{total}$	-0.124	-0.213	-0.122	-0.082	-0.057	0.061	0.091	0.059	0.047	0.040	-0.025	-0.037	-0.026	-0.018	-0.012
$d\widetilde{\varphi}_{direct}$	-0.124	-0.213	-0.122	-0.082	-0.057	-	-	-	-	-	-0.116	-0.186	-0.115	-0.083	-0.061
$d\widetilde{\varphi}_{indirect}$	-	-	-	-	-	0.061	0.091	0.059	0.047	0.040	0.099	0.153	0.097	0.074	0.060
Non-innovator effect	0.080	0.062	0.080	0.089	0.093	-0.025	-0.018	-0.024	-0.028	-0.031	-0.002	-0.003	-0.002	-0.001	0.001
NI-effect (unweighted)	0.108	0.087	0.110	0.118	0.120	-0.055	-0.043	-0.056	-0.061	-0.064	-0.004	-0.006	-0.005	-0.003	0.000
Innovator effect	0.235	0.302	0.238	0.199	0.170	-0.197	-0.225	-0.199	-0.181	-0.167	-0.027	-0.036	-0.027	-0.022	-0.018
I-effect (unweighted)	1.392	1.523	1.367	1.344	1.394	-0.396	-0.419	-0.395	-0.385	-0.391	-0.063	-0.077	-0.063	-0.055	-0.051
Composition effect	-0.440	-0.578	-0.441	-0.370	-0.320	0.282	0.334	0.282	0.257	0.239	0.003	0.002	0.003	0.004	0.006
C-effect (unweighted)	-0.392	-0.396	-0.394	-0.388	-0.378	0.372	0.349	0.369	0.387	0.400	0.003	0.001	0.003	0.005	0.007
Observations	131,611	25,432	59,841	39,049	7,289	131,611	25,432	59,841	39,049	7,289	131,611	25,432	59,841	39,049	7,289
α (c.p. $\sigma = 4$)	**Direct**	2-3	3-4	4-5	5-7	**Indirect**	2-3	3-4	4-5	5-7	**Hybrid**	3-3.8	3.8-4.6	4.6-5.4	5.4-7
$d\widetilde{\varphi}_{total}$	-0.078	-0.165	-0.055	-0.024	-0.007	0.052	0.087	0.042	0.031	0.025	-0.012	-0.024	-0.011	-0.004	0.002
$d\widetilde{\varphi}_{direct}$	-0.078	-0.165	-0.055	-0.024	-0.007	-	-	-	-	-	-0.081	-0.155	-0.063	-0.034	-0.018
$d\widetilde{\varphi}_{indirect}$	-	-	-	-	-	0.052	0.087	0.042	0.031	0.025	0.075	0.134	0.059	0.039	0.028
Non-innovator effect	0.075	0.068	0.079	0.079	0.074	-0.025	-0.021	-0.027	-0.028	-0.029	-0.000	-0.003	-0.000	0.002	0.005
NI-effect (unweighted)	0.099	0.095	0.105	0.100	0.091	-0.054	-0.049	-0.057	-0.055	-0.052	-0.001	-0.006	-0.001	0.003	0.006
Innovator effect	0.133	0.182	0.127	0.099	0.075	-0.131	-0.158	-0.130	-0.112	-0.096	-0.016	-0.024	-0.015	-0.011	-0.008
I-effect (unweighted)	0.719	0.794	0.695	0.679	0.664	-0.270	-0.290	-0.266	-0.257	-0.252	-0.038	-0.050	-0.037	-0.030	-0.027
Composition effect	-0.285	-0.415	-0.261	-0.200	-0.156	0.209	0.266	0.198	0.171	0.150	0.004	0.002	0.004	0.005	0.005
C-effect (unweighted)	-0.378	-0.396	-0.382	-0.365	-0.339	0.399	0.360	0.400	0.429	0.451	0.007	0.002	0.006	0.010	0.013
Observations	188,874	58,805	68,658	43,295	18,116	188,874	58,805	68,658	43,295	18,116	188,874	58,805	68,658	43,295	18,116
σ (c.p. $\alpha = 6$)	**Direct**	2-3	3-4	4-5	5-7	**Indirect**	2-3	3-4	4-5	5-7	**Hybrid**	2-3	3-4	4-5	5-7
$d\widetilde{\varphi}_{total}$	-0.015	0.021	0.005	-0.015	-0.076	0.033	0.028	0.024	0.027	0.054	0.006	0.024	0.008	-0.001	-0.011
$d\widetilde{\varphi}_{direct}$	-0.015	0.021	0.005	-0.015	-0.076	-	-	-	-	-	-0.024	0.011	-0.006	-0.026	-0.082
$d\widetilde{\varphi}_{indirect}$	-	-	-	-	-	0.033	0.028	0.024	0.027	0.054	0.036	0.019	0.022	0.031	0.074
Non-innovator effect	0.067	0.074	0.076	0.068	0.051	-0.031	-0.045	-0.034	-0.026	-0.017	0.008	0.020	0.009	0.003	0.001
NI-effect (unweighted)	0.082	0.082	0.090	0.086	0.070	-0.052	-0.062	-0.055	-0.049	-0.040	0.009	0.022	0.011	0.003	-0.003
Innovator effect	0.065	0.044	0.065	0.075	0.082	-0.093	-0.100	-0.097	-0.091	-0.084	-0.008	-0.006	-0.007	-0.008	-0.011
I-effect (unweighted)	1.265	2.801	0.913	0.543	0.363	-0.378	-0.717	-0.323	-0.216	-0.157	-0.047	-0.093	-0.034	-0.024	-0.024
Composition effect	-0.147	-0.097	-0.136	-0.158	-0.208	0.158	0.173	0.155	0.143	0.154	0.006	0.010	0.007	0.004	0.002
C-effect (unweighted)	-0.299	-0.187	-0.296	-0.356	-0.389	0.443	0.477	0.475	0.435	0.376	0.014	0.021	0.017	0.011	0.004
Observations	134,943	40,247	34,213	26,427	34,056	134,943	40,247	34,213	26,427	34,056	134,943	40,247	34,213	26,427	34,056

Explanation of the table: The columns show how outcomes are different in different subsets of the simulated sample. For each parameter, we show the outcomes across four different subsamples. The standard effects are all weighted as in Eq. (13). This means that the innovator effect, for example, does not reflect the effect of knowledge spillovers on the average productivity of innovative firms. Rather, it reflects the effect on the cluster productivity that stems from a change in the average productivity of innovative firms. The unweighted effects take out this weighting. Table 1 is based on the weighted numbers.

Table A.2
Comparing outcomes between various clusters.

	Direct	Baseline	USA Compute	GER Cars	NLD Pharma	UK Finance	ITA Apparel	JAP Manufac
$d\widetilde{\varphi}_{total}$	-0.124	-0.141	-0.083	-0.130	-0.158	-0.082	-0.044	-0.133
$d\widetilde{\varphi}_{direct}$	-0.124	-0.141	-0.083	-0.130	-0.158	-0.082	-0.044	-0.133
$d\widetilde{\varphi}_{indirect}$	-	-	-	-	-	-	-	-
Non-innovator effect	0.080	0.049	0.092	0.020	0.020	0.012	0.002	0.066
NI-effect (unweighted)	0.108	0.077	0.108	0.048	0.047	0.037	0.017	0.103
Innovator effect	0.235	0.160	0.129	0.129	0.338	0.112	0.092	0.261
I-effect (unweighted)	1.392	0.443	0.869	0.220	1.603	0.165	0.107	0.729
Composition effect	-0.440	-0.350	-0.303	-0.279	-0.516	-0.206	-0.139	-0.460
C-effect (unweighted)	-0.392	-0.394	-0.227	-0.436	-0.400	-0.339	-0.189	-0.417
Observations	131,611	1	1	1	1	1	1	1
	Indirect	Baseline	USA Compute	GER Cars	NLD Pharma	UK Finance	ITA Apparel	JAP Manufac.
$d\widetilde{\varphi}_{total}$	0.061	0.081	0.109	0.007	-0.040	0.037	-0.002	0.063
$d\widetilde{\varphi}_{direct}$	-	-	-	-	-	-	-	-
$d\widetilde{\varphi}_{indirect}$	0.061	0.081	0.109	0.007	-0.040	0.037	-0.002	0.063
Non-innovator effect	-0.025	-0.021	-0.060	-0.006	-0.008	-0.004	-0.001	-0.023
NI-effect (unweighted)	-0.055	-0.054	-0.090	-0.034	-0.036	-0.021	-0.010	-0.063
Innovator effect	-0.197	-0.165	-0.187	-0.126	-0.319	-0.074	-0.062	-0.231
I-effect (unweighted)	-0.396	-0.271	-0.557	-0.155	-0.418	-0.089	-0.065	-0.368
Composition effect	0.282	0.267	0.357	0.140	0.288	0.115	0.060	0.317
C-effect (unweighted)	0.372	0.401	0.393	0.267	0.291	0.215	0.088	0.413
Observations	131,611	1	1	1	1	1	1	1
	Hybrid	Baseline	USA Compute	GER Cars	NLD Pharma	UK Finance	ITA Apparel	JAP Manufac.
$d\widetilde{\varphi}_{total}$	-0.025	-0.027	-0.015	-0.048	-0.108	-0.008	-0.020	-0.027
$d\widetilde{\varphi}_{direct}$	-0.116	-0.128	-0.109	-0.082	-0.094	-0.061	-0.025	-0.112
$d\widetilde{\varphi}_{indirect}$	0.099	0.107	0.098	0.045	0.005	0.055	0.009	0.094
Non-innovator effect	-0.002	-0.004	-0.003	-0.003	-0.004	-0.000	-0.000	-0.003
NI-effect (unweighted)	-0.004	-0.007	-0.004	-0.011	-0.011	-0.002	-0.003	-0.005
Innovator effect	-0.027	-0.026	-0.015	-0.047	-0.108	-0.008	-0.020	-0.028
I-effect (unweighted)	-0.063	-0.053	-0.067	-0.066	-0.162	-0.011	-0.022	-0.058
Composition effect	0.003	0.003	0.003	0.003	0.004	0.000	0.000	0.004
C-effect (unweighted)	0.003	0.003	0.003	0.005	0.004	0.001	0.000	0.004
Observations	131,611	1	1	1	1	1	1	1

Table A.3
Parameter distributions across $d\widetilde{\varphi}$ percentiles.

	Entire Distribution (direct)				Entire Distribution (indirect)				Entire Distribution (hybrid)			
	mean	sd	min	max	mean	sd	min	max	mean	sd	min	max
θ	0.43	0.19	0.01	0.90	0.43	0.19	0.01	0.90	0.43	0.19	0.01	0.90
ζ	0.74	0.09	0.35	0.95	0.74	0.09	0.35	0.95	0.74	0.09	0.35	0.95
$\bar{f}_{RD}/f$	0.28	0.21	0.01	1.00	0.28	0.21	0.01	1.00	0.28	0.21	0.01	1.00
σ	3.99	1.35	2.00	7.00	3.99	1.35	2.00	7.00	3.99	1.35	2.00	7.00
$\alpha+1-\sigma$	0.51	0.24	0.00	1.56	0.51	0.24	0.00	1.56	0.51	0.24	0.00	1.56
	Lowest 10th Percentile (direct)				Lowest 10th Percentile (indirect)				Lowest 10th Percentile (hybrid)			
	mean	sd	min	max	mean	sd	min	max	mean	sd	min	max
θ	0.65	0.16	0.04	0.90	0.48	0.21	0.02	0.90	0.45	0.20	0.01	0.90
ζ	0.72	0.09	0.32	0.95	0.67	0.09	0.30	0.95	0.66	0.08	0.30	0.92
$\bar{f}_{RD}/f$	0.28	0.19	0.01	1.00	0.17	0.15	0.01	1.00	0.40	0.19	0.07	1.00
σ	3.96	1.18	2.00	7.00	2.44	0.49	2.00	6.91	2.72	0.61	2.00	6.89
$\alpha+1-\sigma$	0.27	0.19	0.00	1.30	0.50	0.23	0.00	1.47	0.32	0.17	0.00	1.12
	Highest 10th Percentile (direct)				Highest 10th Percentile (indirect)				Highest 10th Percentile (hybrid)			
	mean	sd	min	max	mean	sd	min	max	mean	sd	min	max
θ	0.43	0.23	0.02	0.90	0.57	0.18	0.02	0.90	0.42	0.19	0.02	0.90
ζ	0.78	0.11	0.36	0.95	0.81	0.07	0.54	0.95	0.83	0.09	0.36	0.95
$\bar{f}_{RD}/f$	0.31	0.29	0.01	1.00	0.46	0.22	0.03	1.00	0.29	0.28	0.01	1.00
σ	3.58	1.39	2.00	7.00	4.41	1.16	2.13	7.00	4.25	1.49	2.00	7.00
$\alpha+1-\sigma$	0.66	0.23	0.01		0.32	0.21	0.00	1.19	0.65	0.24	0.00	1.56

(*continued on next page*)

Table A.3 (*continued*)

	Entire Distribution (direct)				Entire Distribution (indirect)				Entire Distribution (hybrid)			
	mean	sd	min	max	mean	sd	min	max	mean	sd	min	max
	Lowest 1st Percentile (direct)				Lowest 1st Percentile (indirect)				Lowest 1st Percentile (hybrid)			
	mean	sd	min	Max	mean	sd	min	max	mean	sd	min	max
θ	0.79	0.10	0.21	0.90	0.62	0.20	0.06	0.90	0.47	0.21	0.04	0.90
ζ	0.71	0.09	0.39	0.95	0.64	0.09	0.35	0.88	0.64	0.08	0.35	0.86
$\bar{f}_{RD}/f$	0.25	0.18	0.01	0.98	0.32	0.16	0.07	0.99	0.51	0.19	0.14	1.00
σ	3.91	1.13	2.07	6.99	2.13	0.14	2.00	3.32	2.36	0.32	2.00	4.23
$\alpha + 1 - \sigma$	0.16	0.14	0.00	0.73	0.35	0.20	0.00	1.21	0.18	0.12	0.00	0.69
	Highest 1st Percentile (direct)				Highest 1st Percentile (indirect)				Highest 1st Percentile (hybrid)			
	mean	sd	min	max	mean	sd	min	max	mean	sd	min	max
θ	0.78	0.11	0.31	0.90	0.69	0.17	0.12	0.90	0.41	0.20	0.03	0.90
ζ	0.85	0.07	0.59	0.95	0.85	0.07	0.56	0.95	0.85	0.07	0.52	0.95
$\bar{f}_{RD}/f$	0.56	0.22	0.10	1.00	0.50	0.21	0.06	1.00	0.62	0.20	0.12	1.00
σ	3.57	1.21	2.00	6.99	4.38	1.10	2.33	6.99	2.43	0.38	2.00	4.06
$\alpha + 1 - \sigma$	0.73	0.20	0.13	1.42	0.21	0.19	0.00	0.91	0.85	0.18	0.41	1.48

References

Agarwal, R., Gort, M., 1996. The Evolution of Markets and Entry, Exit and Survival of Firms. The Review of Economics and Statistics 78 (3), 489–498.

Alcácer, J., Chung, W., 2007. Location Strategies and Knowledge Spillovers. Management Science 53 (5), 760–776.

Amir, R., Wooders, J., 1999. Effects of one-way spillovers on market shares, industry price, welfare, and R&D cooperation. Journal of Economics & Management Strategy 8 (2), 223–249.

Amir, R., Wooders, J., 2000. One-Way Spillovers, Endogenous Innovator/Imitator Roles, and Research Joint Ventures. Games and Economic Behavior 31, 1–25.

Asheim, B.T., Gertler, M., 2005. The Geography of Innovation: Regional Innovation Systems. In: Fagerberg, J., Mowery, D., Nelson, R. (Eds.), The Oxford Handbook of Innovation. Oxford University Press, Oxford, pp. 291–317.

d'Aspremont, C., Jacquemin, A., 1988. Cooperative and noncooperative R&D in duopoly with spillovers. American Economic Review 78, 1133–1137.

d'Aspremont, C., Bhattacharya, S., Gérard-Varet, L.-A., 1998. Knowledge as a public good: efficient sharing and incentives for development effort. Journal of Mathematical Economics 30, 389–404.

Atkeson, A., Burstein, A.T., 2010. Innovation, Firm Dynamics, and International Trade. Journal of Political Economy 118 (3), 433–484.

Audretsch, D.B., Feldman, M.P., 1996. R&D Spillovers And The Geography of Innovation and Production. American Economic Review 86 (3), 630–640.

Audretsch, D.B., Feldman, M.P., 2004. Knowledge spillovers and the geography of innovation, Chapter 61. In: Henderson, J.V., Thisse, J-F. (Eds.), Handbook of Regional and Urban Economics. Elsevier, Amsterdam, pp. 2713–2739.

Baldwin, R.E., Robert-Nicoud, F., 2008. Trade and growth with heterogeneous firms. Journal of International Economics 74, 21–34.

Baptista, R., Swann, P., 1998. Do firms in clusters innovate more? Research Policy 27, 525–540.

Behrens, K., Duranton, G., Robert-Nicoud, F., 2014. Productive Cities: Sorting, Selection, and Agglomeration. Journal of Political Economy 122 (3), 507–553.

Bergstrand, J.H., Egger, P., Larch, M., 2013. Gravity Redux: Estimation of gravity-equation coefficients, elasticities of substitution, and general equilibrium comparative statics under asymmetric bilateral trade costs. Journal of International Economics 89, 110–121.

Bernard, A.B., Jensen, J.B., Redding, S.J., Schott, P.K., 2007. Firms in international trade. Journal of Economic Perspectives 21 (3), 105–130.

Blind, K., Edler, J., Frietsch, R., Schmoch, U., 2006. Motives to patent: empirical evidence from Germany. Research Policy 35, 655–672.

Bondt, R.De, 1996. Spillovers and innovative activities. International Journal of Industrial Organization 15, 1–28.

Boschma, R.A., 2005. Proximity and Innovation: A Critical Assessment. Regional Studies 39 (1), 61–74.

Brakman, S., Garretsen, H., Gorter, J., van der Horst, A., Schramm, M., 2005. New Economic Geography, Empirics, and Regional Policy. CPB Discussion paper 56, CPB Netherlands Bureau for Economic Policy Analysis. The Hague, the Netherlands.

Braunerhjelm, P., Feldman, M.P., 2006. Cluster Genesis. Technology-Based Industrial Development. Oxford University Press, Oxford.

Callois, J-M., 2008. The two sides of proximity in industrial clusters: The trade-off between process and product innovation. Journal of Urban Economics 63 (1), 146–162.

Canova, F., 1995. Sensitivity Analysis and Model Evaluation in Simulated Dynamic General Equilibrium Models. International Economic Review 36 (2), 477–501.

Cassiman, B., Veugelers, R., 2002. R&D Cooperation and Spillovers: Some Empirical Evidence from Belgium. American Economic Review 92 (4), 1169–1184.

Cassiman, B., Veugelers, R., 2006. In Search of Complementarity in Innovation Strategy: Internal R&D and External Knowledge Acquisition. Management Science 52 (1), 68–82.

Cassiman, B., Veugelers, R., Arts, S., 2018. Mind the gap: Capturing value from basic research through combining mobile inventors and partnerships. Research Policy 47, 1811–1824.

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, E., Marsili, O., 2005. A matter of life and death: innovation and firm survival. Industrial and Corporate Change 14 (6), 1167–1192.

Chung, W., Kalnins, A., 2001. Agglomeration effects and performance: a test of the Texas lodging industry. Strategic Management Journal 22, 969–988.

Cohen, W.M., Levinthal, D.A., 1989. Innovation and Learning: The Two Faces of R&D. The Economic Journal 99, 569–596.

Cohen, W.M., Levinthal, D.A., 1990. Absorptive Capacity: A New Perspective on Learning and Innovation. Administrative Science Quarterly 35 (1), 128–152.

Combes, P.-P., Duranton, G., Gobillon, L., Puga, D., Roux, S., 2012. The Productivity Advantages of Large Cities: Distinguishing Agglomeration from Firm Selection. Econometrica 80 (6), 2543–2594.

Decarolis, D.M., Deeds, D.L., 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.

Del Gatto, M., Ottaviano, G., Pagnini, M., 2008. Openness To Trade And Industry Cost Dispersion: Evidence From A Panel Of Italian Firms. Journal of Regional Science 48, 97–129.

Department for Business Innovation & Skills, 2016. Headlines from the UK Innovation Survey 2015. London.

Drucker, J., Feser, E., 2012. Regional industrial structure and agglomeration economies: An analysis of productivity in three manufacturing industries. Regional Science and Urban Economics 42, 1–14.

Dunne, T., Roberts, M.J., Samuelson, L., 1988. Patters of Firm Entry and Exit in U.S. Manufacturing Industries. The RAND Journal of Economics 19 (4), 495–515.

Ebert, T., Brenner, T., Brixy, U., 2019. New firm survival: the interdependence between regional externalities and innovativeness. Small Business Economics 53, 287–309.

Fenn, J., Raskino, M., 2008. Mastering the Hype Cycle: How to Choose the Right Innovation at the Right Time. Harvard Business Press, Cambridge, Mass.

Figueiredo, O., Guimarães, P., Woodward, D., 2015. Industry localization, distance decay, and knowledge spillovers: following the patent paper trail. Journal of Urban Economics 89, 21–31.

Findlay, R., 1978. Relative Backwardness, Direct Foreign Investment, and the Transfer of Technology: A Simple Dynamic Model. Quarterly Journal of Economics 92 (1), 1–16.

Fischer, W.A., 1978. Follow-Up Strategies for Technological Growth. California Management Review 21 (1), 10–20.

Garcia, R., Calantone, R., 2002. A Critical Look At Technological Innovation Typology And Innovativeness Terminology: A Literature Review. Product Innovation Management 19, 110–132.

Gaubert, C., 2018. Firm Sorting and Agglomeration. American Economic Review 108 (11), 3117–3153.

Ghironi, F., Melitz, M.J., 2005. International trade and macroeconomic dynamics with heterogeneous firms. Quarterly Journal of Economics 120 (3), 865–915.

Gollotto, J.C., Kim, S., 2003. Market Valuation of Dot Com Companies; R&D versus Hype. Managerial Finance 29 (11), 61–72.

González-Álvarez, N., Nieto-Antolín, M., 2007. Appropriability of innovation results: An empirical study in Spanish manufacturing firms. Technovation 27, 280–295.

Grabowski, H.G., Vernon, J.M., 1992. Brand Loyalty, Entry, and Price Competition in Pharmaceuticals after the 1984 Drug Act. The Journal of Law & Economics 35 (2), 331–350.

Grabowski, H.G., Vernon, J.M., 1994. Innovation and Structural Change in Pharmaceuticals and Biotechnology. Industrial and Corporate Change 3 (2), 435–449.

Gregory, A.W., Smith, G.W., 1993. Statistical Aspects of Calibration in Macroeconomics. In: Maddala, G.S., Rao, C.R., Vinod, H.D. (Eds.), Handbook of Statistics. Elsevier, Amsterdam, pp. 703–719.

Grilli, L., Piva, E., Lamastra, C.R., 2010. Firm dissolution in high-tech sectors: An analysis of closure and M&A. Economics Letters 109, 14–16.

Grossman, G.M., Helpman, E., 1991. Quality Ladders in the Theory of Growth. The Review of Economic Studies 58 (1), 43–61.

Grünfeld, L.A., 2003. Meet me halfway but don't rush: absorptive capacity and strategic R&D investment revisited. International Journal of Industrial Organization 21, 1091–1109.

Haddad, M., Harrison, A., 1993. Are there positive spillovers from direct foreign investment? Evidence from panel data for Morocco. Journal of Development Economics 42, 51–74.

Hall, R., 1992. The strategic analysis of intangible resources. Strategic Management Journal 13, 135–144.

Helpman, E., Melitz, M.J., Yeaple, S.R., 2004. Export versus FDI with Heterogeneous Firms. American Economic Review 94 (1), 300–316.

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

Jaffe, A.B., Trajtenberg, M., Henderson, R., 1993. Geographic Localization Of Knowledge Spillovers As Evidenced By Patent Citations. Quarterly Journal of Economics 108 (3), 577–598.

Jaffe, A.B., Trajtenberg, M., 2002. Patents, Citations, and Innovations: A Window on the Knowledge Economy. The M.I.T. Press, Cambridge, Mass.

Järvenpää, H., Mäkinen, S.J., 2008. Empirically Detecting the Hype Cycle with the Life Cycle Indicators: An Exploratory Analysis of Three Technologies. In: Proceeding of the 2008 International Conference on Industrial Engineering and Engineering Management (IEEM). Singapore. Institute of Electrical and Electronics Engineers (IEEE).

Kamien, M.I., Zang, I., 2000. Meet me halfway: research joint ventures and absorptive capacity. International Journal of Industrial Organization 18, 995–1012.

Keller, W., 2002. Geographic localization of international technology diffusion. The American Economic Review 92 (1), 120–142.

Khanna, R., Guler, I., Nerkar, A., 2015. Fail Often, Fail Big, and Fail Fast? Learning from Small Failures and R&D Performance in the Pharmaceutical Industry. Academy of Management Journal 59 (2), 436–459.

Knoben, J., Arikan, A.T., van Oort, F., Raspe, O., 2016. Agglomeration and firm performance: One firm's medicine is another firm's poison. Environment and Planning A 48 (1), 132–153.

Leppälä, S., 2018. Theoretical perspectives on localized knowledge spillovers and agglomeration. Papers in Regional Science 97 (3), 467–484.

Luttmer, E.G.J., 2007. Selection, Growth, and the Size Distribution of Firms. Quarterly Journal of Economics 122 (3), 1103–1144.

Malerba, F., Orsenigo, L., 2001. Towards a History Friendly Model of Innovation, Market Structure and Regulation in the Dynamics of the Pharmaceutical Industry: The Age of Random Screening. KITeS Working Papers 124, KITeS Centre for Knowledge, Internationalization and Technology Studies. Universita' Bocconi, Milano, Italy.

Malerba, F., Orsenigo, L., 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.

Marshall, A., 1920. Principles of economics, 8th ed. Macmillan, London (Original work published 1890).

Martin, P., Mayer, T., Mayneris, F., 2011. Spatial concentration and plant-level productivity in France. Journal of Urban Economics 69, 182–195.

McCann, P., Folta, T.B., 2008. Demand- and Supply-Side Agglomerations: Distinguishing between Fundamentally Different Manifestations of Geographic Concentration. Journal of Management Studies 46, 362–392.

McCann, P., Folta, T.B., 2011. Performance differentials within geographic clusters. Journal of Business Venturing 26, 104–123.

Melitz, M.J., 2003. The impact of trade on intra-industry reallocations and aggregate industry productivity. Econometrica 71 (6), 1695–1725.

Melitz, M.J., Ottaviano, G.I.P., 2008. Market Size, Trade, and Productivity. The Review of Economic Studies 73 (1), 295–316.

Molina-Morales, F.X., Martínez-Fernández, M.T., 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.

Nigai, Sergey, 2017. A Tale of Two Tails: Productivity Distribution and the Gains from Trade. Journal of International Economics 104, 44–62.

Nooteboom, B., 2000. Learning by Interaction: Absorptive Capacity, Cognitive Distance and Governance. Journal of Management and Governance 4, 69–92.

Nooteboom, B., Van Haverbeke, W., Duysters, G., Gilsing, V., Oord, A.van den, 2007. Optimal cognitive distance and absorptive capacity. Research Policy 36, 1016–1034.

Pavitt, K., 1987. The objectives of technology policy. Science and Public Policy 14 (4), 182–188.

Pérez, S.E., Llopis, A.S., Llopis, J.A.S., 2004. The Determinants of Survival of Spanish Manufacturing Firms. Review of Industrial Organization 25, 251–273.

Perkmann, M., Tartari, V., McKelvey, M., Autio, E., Broström, A., D'Este, P., Fini, R., Geuna, A., Grimaldi, R., Hughes, A., Krabel, S., Kitson, M., Llerena, P., Lissoni, F., Salter, A., Sobrero, M., 2013. Academic engagement and commercialisation: A review of the literature on university-industry relations. Research Policy 42, 423–442.

Petruzzelli, A.M., 2011. The impact of technological relatedness, prior ties, and geographical distance on university-industry collaborations: A joint-patent analysis. Technovation 31, 309–319.

Porter, M.E., 1990. The competitive advantage of nations. Free Press, New York.

Porter, M.E., 2000. Location, Competition, and Economic Development: Local Clusters in a Global Economy. Economic Development Quarterly 14 (1), 15–34.

Pouder, R., St. John, C.H., 1996. Hot spots and blind spots: geographical clusters of firms and innovation. Academy of Management Review 21, 1192–1225.

Qiao, Y., Ding, C., Liu, J., 2019. Localization, knowledge spillover, and R&D investment: Evidence of Chinese cities. Journal of Urban Management 8, 303–315.

Rigby, D.L., Brown, W.M., 2015. Who Benefits from Agglomeration? Regional Studies 49 (1), 28–43.

Rocha, H., 2004. Entrepreneurship and Development: The Role of Clusters. Small Business Economics 23, 363–400.

Schewe, G., 1996. Imitation as a strategic option for external acquisition of technology. Journal of Engineering and Technology Management 13, 55–82.

Shaver, J.M., Flyer, F., 2000. Agglomeration Economies, Firm Heterogeneity, and Foreign Direct Investment in the United States. Strategic Management Journal 21, 1175–1193.

Todo, Y., Miyamoto, K., 2006. Knowledge Spillovers from Foreign Direct Investment and the Role of Local R&D Activities: Evidence from Indonesia. Economic Development and Cultural Change 55 (1), 173–200.

Van Looy, B., Landoni, P., Callaert, J., Van Pottelsberghe, B., Sapsalis, E., Debackere, K., 2011. Entrepreneurial effectiveness of European universities: An empirical assessment of antecedents and trade-offs. Research Policy 40, 553–564.

Verhoogen, E.A., 2008. Trade, quality upgrading, and wage inequality in the Mexican manufacturing sector. Quarterly Journal of Economics 123 (2), 489–530.

Vestal, A., Danneels, E., 2018. Knowledge exchange in clusters: The contingent role of regional inventive concentration. Research Policy 47, 1887–1903.

Visser, E.J., 1999. A comparison of clustered and dispersed firms in the small-scale clothing industry of Lima. World Development 27 (9), 1553–1570.

Wennberg, A., Lindqvist, G., 2010. The effect of clusters on the survival and performance of new firms. Small Business Economics 34, 221–241.

West, G.R., 1986. A Stochastic Analysis of an Input-Output Model. Econometrica 54 (2), 363–374.

Wuyts, S., Colombo, M.G., Dutta, S., Nooteboom, B., 2005. Empirical tests of optimal cognitive distance. Journal of Economic Behaviour & Organization 58 (2), 277–302.

Zhou, K.Z., 2006. Innovation, imitation, and new product performance: The case of China. Industrial Marketing Management 35, 394–402.

Zucker, L.G., Darby, M.R., Armstrong, J., 1998a. Geographically Localized Knowledge: Spillovers or Markets? Economic Inquiry 36 (10), 65–86.

Zucker, L.G., Darby, M.R., Brewer, M.B., 1998b. Intellectual Human Capital and the Birth of U.S. Biotechnology Enterprises. American Economic Review 88 (1), 290–306.