University of Reading

The lab and the plant: offshore R&D and co-location with production activities

Article

Accepted Version

Castellani, D. ORCID: https://orcid.org/0000-0002-1823-242X and Lavoratori, K. ORCID: https://orcid.org/0000-0002-0078-4525 (2020) The lab and the plant: offshore R&D and co-location with production activities. Journal of International Business Studies, 51 (1). pp. 121-137. ISSN 1478-6990 doi: 10.1057/s41267-019-00255-3 Available at https://centaur.reading.ac.uk/84907/

It is advisable to refer to the publisher's version if you intend to cite from the work. See Guidance on citing.

To link to this article DOI: http://dx.doi.org/10.1057/s41267-019-00255-3

Publisher: Palgrave Macmillan

All outputs in CentAUR are protected by Intellectual Property Rights law, including copyright law. Copyright and IPR is retained by the creators or other copyright holders. Terms and conditions for use of this material are defined in the End User Agreement.

www.reading.ac.uk/centaur

University of Reading

CentAUR

Central Archive at the University of Reading

Reading's research outputs online

The Lab and the Plant.
Offshore R&D and Co-location with Production Activities

Davide Castellani
Henley Business School, University of Reading,
Whiteknights, Reading, RG6 6UD, United Kingdom
Tel. +44(0)118 378 8846; davide.castellani@henley.ac.uk

Katiuscia Lavoratori
Warwick Business School, University of Warwick
Scarman Road, Coventry, CV4 7AL, United Kingdom
katiuscia.lavoratori@wbs.ac.uk

Forthcoming in *Journal of International Business Studies*

Abstract

The literature has highlighted that the propensity of MNEs to co-locate offshore R&D labs with their production plants can vary substantially according to firm and industry characteristics. In this paper, we apply a novel two-stage estimation procedure that allows us to tease out this heterogenous behaviour and investigate the factors that are associated with a higher propensity to co-locate production and R&D activities abroad. Using data on 1,483 greenfield international investments in R&D activities made by 855 firms in 587 cities worldwide, we uncover that the strength of the co-location effect is indeed highly heterogenous across firms. In particular, it is higher among firms with less international experience and geographical dispersion of international activities, as well as with a lower share of intangible assets. These results are consistent with the idea that co-location is a substitute for firms' ability to coordinate complex and dispersed organizational structures, and that firms relying relatively less on codified knowledge can use co-location of offshore R&D and production to facilitate knowledge transfer across activities.

Acknowledgements
The authors would like to thank the editor Ilan Vertinsky and three anonymous reviewers for their constructive and insightful comments and suggestions. This paper also benefitted from feedback received from reviewers and participants at iBEGIN 2016 (Philadelphia), European International Business Academy 2017 (Vienna), Academy of International Business UK & Ireland 2017 (Reading), R&D Management Conference 2017 (Leuven), Academy of International Business 2018 (Minneapolis), and seminars in Ancona, Lucca, Keele, Trento, Urbino and at NIESR (London). Financial support from a BA/Leverhulme Small Research Grant (SG163244) is gratefully acknowledged.

INTRODUCTION

Co-location between international research and development (R&D) and production activities has a long tradition of International Business (IB) studies, dating back to the earlier work of Cantwell (1989) and Pearce (1989)[i], however open questions still remain. For example, while several studies have shown that R&D activity abroad of German, Japanese and Swedish multinational enterprises (MNEs), tends to co-locate with production activities (Ambos, 2005; Ivarsson, Alvstam, & Vahlne, 2016; Kenney and Florida, 1994), evidence on the U.S. has indicated that the foreign-owned R&D activities are not attracted by the possibility to work with their parent company's manufacturing facilities based in the U.S. (Florida, 1997). Some recent large-scale econometric analyses of location decisions by multinational enterprises (MNEs) have supported the hypothesis that international R&D and production activities of a firm may co-locate in the same country/region (Defever, 2006, 2012) - as well as in global cities (Belderbos, Sleuwaegen, Somers, & De Backer, 2016) - but others have also highlighted that this *propensity to co-locate* production plants and R&D labs can be very heterogeneous across firms and industries (Alcácer and Delgado, 2016; Ivarsson et al. 2016; Mariani, 2002).

In this paper, we investigate whether the nature of knowledge and firms' ability to coordinate geographically dispersed activities moderate the likelihood to locate R&D labs in cities where firms had prior investments in production activities. A key contribution of this study is to introduce, for the first time in an IB context, a novel two-stage empirical strategy - based on the estimation of firm-specific parameters derived from a Random Parameter (Mixed) Logit model - that allows to simultaneously account for several sources of firm and industry heterogeneity. We apply this method to a sample of 1,483 greenfield international investments in R&D activities made by 855 firms over the period 2009-2014, with an unprecedented geographical coverage of 587 cities worldwide. Our results support the idea that the need to co-locate international R&D and production activities increases when firms lack the ability to coordinate complex and dispersed organizational structures. Furthermore, while the tacit and 'synthetic' nature of knowledge in an industry is only imprecisely correlated with the extent of co-location between production and R&D, we find evidence that firms

relying relatively less on codified knowledge may need to co-locate more to allow knowledge transfer across activities.

RELATED LITERATURE

Physical proximity can increase communication and improve coordination between activities (Giroud, 2013). In particular, it can foster flows of sticky knowledge - that does not travel easily over long distances - between different establishments within the firm (Gray, Siemsen, & Vasudeva, 2015; Rawley & Seamans, 2015), and enable face-to-face interactions. This can substantially reduce the costs incurred when firms need to integrate and coordinate knowledge residing in geographically dispersed locations (Ketokivi & Ali-Yrkkö, 2009). Based on these arguments, researchers have hypothesised the benefits from co-locating R&D with production activities. However, previous research has also highlighted a substantial trade-off between benefitting from external agglomeration economies, that would lead to locate R&D and production in different places, and internal agglomeration forces that would lead to intra-firm co-location of activities across the value chain (Alcácer & Delgado, 2016; Blanc and Sierra, 1999). Consistently with this view, research suggests that R&D and manufacturing represent distinct units with inherently different organizational and professional logics that may require structural and spatial separation (Gray et al. 2015). Furthermore, the rapid advancements in information and communication technologies (ICTs) lead to the expectation that any challenges associated with coordination and knowledge transfer across distance would have diminished over the last two decades, thus reducing the need to co-locate (Gray et al., 2015; Hannigan, Cano-Kollmann, & Mudambi, 2015; Mudambi, 2008).

There are a number of moderating circumstances that may affect the need for co-location or allow easier unbundling of these two types of activities, and consequently determine heterogeneous effects of the presence of previous production activities of a firm on the probability to choose the same location for their international R&D facilities. These circumstances have to do with a.) the nature of

knowledge; b.) firms' ability to coordinate and manage complex and dispersed organizational structures.

The nature of knowledge

The need to co-locate R&D and production can be stronger for plants that rely more on tacit knowledge, whereas in the case of codified-intensive knowledge unbundling is easier, because the exchange of knowledge across space is facilitated (Gray et al., 2015).

The economic geography literature has identified 'synthetic' and 'analytical' knowledge bases in manufacturing industries (Asheim & Gertler, 2005). An industrial setting based on 'synthetic' knowledge depends more on applied and engineering-based knowledge, often generated in response to specific problems and by an inductive process, where learning-by-doing, experimentation, testing, technical and engineering solutions are the keywords. Knowledge is more tacit and related to know-how and practical skills. Instead, an 'analytical' knowledge base refers to a context where knowledge creation is based on cognitive and rational processes, codification is more frequent, and scientific results are documented in files, reports and patents. Knowledge inputs are mainly based on reviews of existing (and previously codified) studies or the application of understood scientific methods and principles. When the knowledge base is more 'synthetic' spatial proximity and face-to-face contacts between units are more important, thus knowledge and information are more difficult to transfer among globally dispersed units. In turn, this increases the need to co-locate production and R&D activities (Asheim, 2007; Grillitsch, Martin, & Srholec, 2017; Plum & Hassink, 2011). Consistently with this view, it has been shown that more user-oriented research is more likely to be located closer to production, while geographical separation between production and R&D is possible when R&D activities are science-intensive and give rise to generic knowledge or 'general purpose technologies' (Mariani, 2002).

Firms' ability to coordinate complex and dispersed organizational structures

Notwithstanding differences across industries, the *propensity to co-locate* R&D with production activities can also vary substantially across firms within industries. In particular, firm capabilities in organising, managing and taking advantage from a complex and global architecture can reduce the

coordination problems and costs of a geographically dispersed global network of production and R&D activities. Instead, the lack of structured organizational processes for knowledge exchange across distant units may increase the need to co-locate activities (Alcácer and Delgado, 2016; Gray et al., 2015). Such coordination capabilities can be associated with the *international experience* and *geographical dispersion* of MNEs global operations. On the one hand, the international experience can be a key source of organisational learning (Barkema, Shenkar, Vermeulen, & Bell, 1997; Barkema & Vermeulen, 1999), allowing MNEs to develop organisational and managerial capabilities that enable them to better coordinate and manage knowledge transfer across geographically dispersed units, in turn reducing the need for co-locating different activities across the value chain. On the other hand, firms that operate across geographically dispersed locations face higher coordination, control and transfer costs, and they need to develop coherent coordination capabilities (Gerybadze & Reger, 1999). Consistently with this view, Alcácer & Delgado (2016) find that the co-location across activities is higher for firms operating in a few locations, while geographically diversified firms have better managerial capabilities to disperse value chain activities across distant locations, thus reducing the need for co-location.

The discussion above motivates our empirical analysis on the determinants of MNEs *propensity to co-locate* their international R&D with production activities. In particular, we will investigate the role of the *nature of knowledge* in the industry, as well as the *international experience* and *geographical dispersion* of MNEs global operations in moderating the relation between the location of production plants and R&D labs abroad.

DATA

Our empirical analysis relies on data on international investment projects from *fDi Markets*, a database produced by fDi Intelligence, a division of the Financial Times Ltd., which tracks cross-border greenfield investment projects across different industries and countries worldwide. The database

contains projects in different business activities, including research and development, design, development and testing, manufacturing, business services, logistics, marketing and sales, headquarters, ICT, training and technical support, as well information on the country and city where the investment takes place, the name of the investing company and the location of the headquarters. For the purpose of this analysis, we rely on an initial sample of 6,235 new R&D[ii] cross-border investments over the period 2003-2014[iii]. We focus the analysis on investments made in the manufacturing industries[iv] in the second half of period covered by the dataset (2009-2014). This choice allows us to use the period between 2003 and t-1 to build the cumulated number of investments in production activities, which is our main explanatory variable. Due to missing values, the overall sample of R&D investment decisions consists of 1,483 new R&D projects made by 855 MNEs in 373 cities worldwide. In our estimation, we consider a choice set which includes all the 587 cities receiving at least one investment in R&D during the whole period covered by *fDi Markets* (2003-2014).

Using the city as the unit of analysis allows us to exploit the highest geographical granularity in our data, in order to capture 'true' co-location. In fact, spurious correlations could be picked up if a firm had production plants in the same country (or region within country) where they subsequently locate R&D labs, but these may in fact be fairly distant from one another.

Descriptive statistics[v] reveals that a handful (38) of 'Mega' cities (with more than 5 million habitants) receive 43.3% of total number of investments (642 out of 1,483), but a sizable 30.9% of R&D project takes place in cities with less than 1 million inhabitants. Around 37.9% of projects are located in Europe and North America, where United Kingdom, Germany and USA account for the 20%, while Asia attracts about 51.32% of R&D projects, mainly in China and India, that account for 20.60% and 11.6% respectively. Furthermore, about 80% of projects come from Europe and North America, mainly from Germany, France, United Kingdom and USA. Only 18% of the investments is from Asia, with half of them accounted for by Japanese MNEs. The Electronics and the Pharma, Chemical and Biotech sectors are the two largest investors in R&D activities, with 33.18 and 26.37 percent of all investments, followed by Transport with 11.33% of the projects. Interestingly for the purpose of our study, we find that only in 188 cases (12.62%) firms locate their new R&D investments in a city

where they already had prior activities in production. These cases of co-location are relatively more likely in industries such as Aerospace, Automotive, Plastics & Rubber and Chemicals, as well as in cities located in South-East Asia. This apparent lack of co-location reflects the fact that production and R&D are attracted by different location characteristics (e.g. wage and size of the market for production vs. human capital and local innovation system for R&D), and this will be more rigorously investigated in our econometric analysis, where we will condition on some of the relevant differences across locations.

EMPIRICAL METHODOLOGY

Our estimation strategy needs to allow us a.) to assess the effect of prior investments in production activities of a firm in a certain location on the probability to locate subsequent R&D labs of the same firm in the same location; b.) to account for the possibility that such co-location effect is heterogeneous across firms and industries and c.) to identify the sources of such heterogeneity.

We approach this estimation problem with a two-step procedure. In the first step, we estimate a Random Parameter (Mixed) Logit model (MLM) of the probability that a firm locates an R&D lab in a certain city. Our main independent variable is a measure of the firm prior production plants in the same city, and we allow for this parameter to be randomly distributed across firms. From this first-stage estimation we obtain an average co-location parameter (point a. above), and a standard deviation, which provides an indication of the heterogeneity in the importance of co-location (point b. above).

In the second step, we try and explain the sources of such heterogeneity in two different ways. First, we adjust the model specification by introducing interaction terms of the presence of prior production plants with several moderating factors and investigate whether these factors influence the significance of the second moment (i.e. standard deviation). Second, we exploit the fact that for each firm engaged in an international R&D investment project we can compute the vector of simulated values of the parameter associated with the presence of prior production plants (Revelt & Train, 1999). Following Hornstein & Greene (2012), we use this vector as a dependent variable in a second-stage OLS

estimation. This allows us to assess the relative importance of various firm and industry characteristics in explaining co-location between production and R&D (point c. above).

As highlighted by Alcácer, Chung, Hawk, & Pacheco-de-Almeida (2018), this empirical approach allows researchers to test and build new theories at a more granular level. On the one hand, it may help reconcile contradictory empirical findings that were exclusively based on average effects, while on the other hand it may help revive interest in explanatory variables whose average effects across firms were found to be insignificant. Most notably, this approach seems particularly suited to strategy and IB studies which build on firm heterogeneity and where theories often predict several moderating circumstances.

Location Choice Model

The literature on firm's location choice has mainly estimated Conditional Logit (CLM) and Nested Logit models (Basile, Castellani, & Zanfei, 2009; Li, Meyer, Zhang & Ding, 2018; Nielsen, Asmussen, & Weatherall, 2017). More recent empirical studies on location choices of MNEs estimate Mixed Logit (MLM) models (Train, 2003), where random parameters are introduced in order to allow heterogeneous responses to location characteristics (Basile, Castellani, & Zanfei, 2008; Belderbos et al., 2016; Chung and Alcácer, 2002; Defever, 2012; Rasciute & Downward, 2017). This is particularly important in our case, since we have highlighted a variety of factors that can moderate the effect of previous production plants that a firm has in a certain (foreign) location on the probability to choose that same location for its subsequent R&D labs. These factors can lead to a significant heterogeneity in preferences over location attributes. A random parameter model allows to account for such heterogeneity in a very flexible way.

In formal terms, firm *f* makes a series of (up to) J decisions to locate R&D investment projects in a set of L possible foreign locations. We assume that firm *f* chooses *l** (in location decision *j*), if that yields the highest profits among a set of possible alternatives, that is $\pi_{fjl^*} > \pi_{fjl} \quad \forall l \neq l^*$ (l = 1, …, L-1).[vi]

The profit that firm *f* derives from location *l* (in decision *j*) is not directly observed but it is defined by

$$\pi_{fjl} = \beta_f \; x_{fjl} + \gamma_f \; Z_{fjl} + \delta_f \; M_{jl} + \varepsilon_{fjl}$$

where x_{fjl} is a variable measuring the presence of production activities (in the year prior to location decision *j*) of firm *f* in location *l*, Z_{fjl} and M_{jl} are vectors of characteristics (measured in the year prior to location decision *j*) determining the attractiveness of a certain location and varying over firms-locations and locations, respectively. β_f, γ_f and δ_f are random parameters which represent heterogeneous preferences over the attributes *x*, *Z* and *M*. The CLM specification is a special case of MLM, where these parameters are all fixed and do not vary over firms.

The coefficient β_f is our main object of interest. It varies in the population with a density g(β) and if we assume g(.) is a normal distribution, it can be decomposed in its mean α and standard deviation μ_f, so that we can write $\beta_f x_{fjl} = (\alpha \quad x_{fl} + \mu_f x_{fjl})$ (Train, 2003: 144). A significant μ_f reveals that different firms may in fact have significantly different preferences over co-location between international production and R&D. For some firms, the location of R&D may be very strongly associated with previous production activities, while for others co-location may be unnecessary. Similar considerations apply to γ_f and δ_f, but they are not the main focus of this paper.

Explaining the second moment

Having identified a significant standard deviation of the MLM parameters, the researcher could try and explain this heterogeneous behaviour by augmenting the model with interaction effects between the parameter of interest and firm/industry characteristics[vii]. For example, Chung and Alcácer (2002) use an MLM to study how multinational firms' location decisions are affected by U.S. state R&D activity. After finding a significant second moment for states' R&D activity, they added several industry dummies interacted with state R&D activity, which eventually lead the second moment's estimate to become insignificant. However, as the number of interaction terms increases, this approach can be very taxing on the properties of the estimators (Alcácer et al., 2018), and increase both computational intensity and the risk of non-convergence of the likelihood function[viii].

Using Random Parameters as Dependent Variables

A different strategy would be to simulate the *propensity to co-locate* international production and R&D for each firm ($\hat{\beta}_f$), on the basis of the prior distribution from the MLM model combined with each observed choice pattern (Revelt & Train, 1999). This vector can then be used as a dependent variable in a second-stage regression (Greene, Hornstein, & White, 2009; Hawk & Pacheco-de-Almeida, 2018; Hornstein & Greene, 2012). This allows uncovering systematic patterns in the magnitude of the *propensity to co-locate* international production and R&D across firms and industries. A two-stage estimation like this one - which is inspired by Saxonhouse (1976)[ix] - is appropriate when the variable of interest cannot be measured directly, as it is our case. Furthermore, this approach allows to simultaneously account for a large set of sources of heterogeneity in an efficient way. This is particularly relevant in our case, since we have different sources of heterogeneity that may moderate the importance of co-location between production and R&D and a particularly cumbersome estimation of the first-stage MLM.

Our second stage will then consist of an OLS regression[x] of the simulated $\hat{\beta}_f$ values on a vector of firm (X_{1f}) and industry (X_{2s}) characteristics:

$$\hat{\beta}_f = \theta_1 X_{1f} + \theta_2 X_{2s} + u_f$$

Saxonhouse (1976) and Hornstein & Greene (2012) suggest that the estimated parameters may suffer from heteroscedasticity, which can lead to inefficient second stage estimates and underestimate the variance of θ_1 and θ_2 . In order to address this issue, we compute robust standard errors, using the White sandwich estimator (Duren, Merck, & Yeung, 2004). Results using bootstrapped standard errors are qualitatively similar and are available from the authors upon request.

First-stage model specification

Dependent variable

The dependent variable is the location choice of an international R&D investment project. This variable assumes value 1 if a given R&D project *i* is located in the chosen city *l**. It assumes value zero for all possible alternative cities $l \neq l^*$.

Main Independent variable

Prior production plants

The variable captures whether a firm had established production plants in a certain city, prior to the current location decision *j*. For each investment project, we know the year t in which it is announced. Hence, we count the number of investments in production activities by firm *f* in city *l* from 2003 until time t-1. We then convert this count into a dichotomous variable, which takes value 1 for counts greater than zero.

Second-stage model specification

Dependent Variable

Propensity to co-locate international R&D and production, that is the vector of simulated values of the parameter associated with the effect of prior production plants on the location of R&D labs ($\hat{\beta}_f$) from the first stage MLM estimation.

Independent Variables

In Section 2 we discussed different sources of heterogeneity associated with co-location preferences. These variables are computed at the firm level and need to be time-invariant. Hence, we have chosen to refer to the period 2003-2008, so that they precede the period of observation of our dependent variable.

First, in line with a large IB literature, we construct a measure of the *degree of geographical dispersion* as the number of countries in which a firm has registered at least one foreign investment between 2003 and 2008 (as a share of the total number of countries were all firms had foreign investments) (Castellani & Zanfei, 2004; Caves and Mehra, 1986; Dunning, 1996; Hennart, 2007; Kogut & Singh, 1988). This variable should capture the fact that, by operating across geographically dispersed locations, firms face higher coordination, control and transfer costs and need to develop coherent coordination capabilities. Thus, the need for co-location between international production and R&D is expected to be higher for firms that have a lower international geographical dispersion, since they should be less able to transfer knowledge across distant production and R&D units.

Second, in order to test for the role of *international experience* of the firm, we compute the number of years since the first investment of the firm (in R&D or in any other business activities) recorded by *fDi Markets*. This variable should account for the fact that international experience can be a key source of organisational learning, allowing MNEs to develop organisational and managerial capabilities that enable them to better coordinate dispersed operations across distance and manage knowledge transfer across geographically dispersed units, in turn reducing the need for co-locating international production and R&D activities.

Third, since the literature suggests that the propensity to co-locate production and R&D should be higher if the knowledge base is less codified, we introduce one dummy variable which takes value 1 for projects in industries that are characterised by 'synthetic' knowledge bases (*Synthetic Sectors*)[xi].

We further control for the home country of the firm (US, EU, Rest of the World) and, in further checks on a subsample of firms which could be matched with Bureau Van Dijk's Orbis, we also control for a measure proxying for the importance of codified knowledge for the firm, computed as the share of intangibles to total assets, the share of foreign subsidiaries to the total number of (national and foreign) subsidiaries, the number of patents, age and size of the firm.

Detailed definitions of all the variables are reported in Table 1, while correlation tables and descriptive statistics, both for first- and second- stage variables, are available in the Appendix.

[**Table 1** goes about here]

ECONOMETRIC RESULTS

Location choice model

[**Table 2** goes about here]

Our baseline specification, presented in Column (1) and (2) of Table 2, reveals that prior investments in production activities of a firm in a city are positively associated with new investments in R&D in that city.

However, the significant standard deviation associated with the effect of previous production activities on the location of R&D (Column 2) suggests that this effect is very heterogeneous across firms. Our second stage aims at uncovering the sources of such heterogeneity.

Uncovering the sources of heterogeneous preferences

In Mod. 2 we augment our baseline MLM model by interacting the prior production dummy with three characteristics of the investing firm: geographical dispersion of its international activities, the number of years since its first investment recorded in *fDi Markets* (international experience), and a dummy taking value 1 for firms in industries where the knowledge base is mainly synthetic. Results reveal that both the geographical dispersion and the international experience negatively moderate the correlation between prior production and the location of R&D activities abroad.

As discussed in the "Empirical methodology" section, this approach can be very onerous on the estimator, and in fact we have experienced problems with convergence of the likelihood function and extended length of the estimation[xii]. Moreover, looking at results in Mod. 2, we can appreciate that the standard deviation associated with the prior production dummy is still significant (column 4), suggesting that other sources of heterogeneity can simultaneously moderate the co-location effect and that a more flexible estimation strategy, such as the two-stage approach described in the "Empirical methodology" section may be appropriate.

[**Figure 1** and **Table 3** go about here]

Figure 1 plots the kernel density of the simulated value of the $\hat{\beta}_f$ parameter, which is then used as dependent variable in second-stage OLS regressions[xiii] presented in Table 3. Figure 1 reveals that indeed $\hat{\beta}_f$ is quite heterogeneous across firms, with a ratio of 75th to 25th percentile of 1.36, and a fat right tail of outliers with large values of $\hat{\beta}_f$ (greater than 2).

Results from Table 3 show that firms with lower geographical dispersion have a higher propensity to co-locate production and R&D. As for the correlation between international experience and $\hat{\beta}_f$, we find that in Mod. 2 it is negative, as expected, but rather imprecisely estimated. When estimated jointly with the geographical dispersion, the coefficient of international experience becomes virtually

zero. However, in the sample excluding the outliers identified in Figure 1 above (Mod. 7) it is negatively and significantly associated with the propensity to co-locate international production and R&D. It is worth mentioning that excluding these outliers improves the fit of our regression dramatically, with a nine-fold surge in the Adjusted R-squared (from 2.3% in Mod. 6 to 20.9% in Mod. 7).

These results support the idea that the international experience may contribute to the organisational learning process of the firm, and the higher geographical spread of international operations may be associated with greater ability of the firm to manage geographically dispersed activities across the value chain and to orchestrate more complex coordination and communication mechanisms.

Instead, the synthetic nature of industrial knowledge, is not significantly associated with a higher propensity to co-locate international production and R&D. This result could reveal that the industry characterisation of knowledge bases may be too coarse and hinder heterogeneity. In Mod. 5-7 we replace the 'synthetic' sector dummy with a full set of industry dummies. These regressions highlight that differences across sectors are somewhat limited, but prior production investments have a large and significant correlation with subsequent R&D investments in the Plastics & Rubber and in Transport industries. Both industries are characterised by synthetic knowledge bases, and the Transport industry is also engineering-intensive, where a higher propensity to co-locate has been found in previous studies (Belderbos et at., 2016; Ivarsson et al., 2016). It should be noted that industry dummies contribute to improve significantly the explanatory power of the regression, with an Adjusted R-squared increasing from 0.58% to 2.3%. However, as Pisano and Shih (2009; 2012) already noted, while industry characteristics certainly play a role, the need for interaction between production and R&D is mostly a firm-specific characteristic.

In line with this argument, we move beyond a characterisation of knowledge bases at the industry level and control for the share of intangibles in total assets at the firm level. This variable may capture the extent to which firms (within industries) rely on codified knowledge (e.g. patents trademarks, and copyrights) which can be more easily transferred and shared across locations, thus making co-location between production and R&D less of a necessity. The share of intangibles is obtained from Orbis, but it is available for only 358 out of 855 firms, thus causing a drop in the number of second-stage

observations. Results, presented in Mod. 7-11 confirm the conjecture that firms with a higher share of intangibles have a lower propensity to co-locate production and R&D.

Overall, these results support the idea that the importance of co-locating international production and R&D activities may differ along several dimensions, and a methodological approach such as the one devised in this paper allows to account for this multidimensionality in quite a flexible and efficient way.

These results are robust to a number of firm-level controls, which include a.) two dummies identifying US and EU firms, which appear to be both more likely to co-locate production and R&D than firms from the Rest of the World, and, b.) in the subsample of firms for which we can gather data from Orbis, we control for the age and size of the firm, the share of foreign subsidiaries, as well as the number of patents, but these variables are not significantly associated with the propensity to co-locate production and R&D. Importantly, the coefficients associated with geographical dispersion and international experience are remarkably stable despite the significant reduction in sample size. As discussed above, results are more significantly affected by the exclusion of outliers, which leads to an increase in the magnitude and correlation of international experience with $\hat{\beta}_f$. Instead, removing the outliers weakens the correlation of the share of intangible assets becoming slightly less significant, and the positive industry effect for Plastics & Rubber and Transport industries vanishes. This finding is related to the fact that outliers are overrepresented in these two industries (30.8% of firms in the Plastics & Rubber and 22.3% of firms in the Transport industry are in the top 10% of the distribution of $\hat{\beta}_f$).

Finally, it is worth mentioning that in terms of economic significance, the effect of the presence of prior production plants is quite high. Following Hole (2007)[xiv] and based on the results from our first stage baseline regression (Mod. 1 in Table 2), we run a simulation of the change in probability of locating an R&D lab in a certain city under a scenario where a firm has prior production activities in that city *vs.* the baseline scenario where a firm has no prior production activities, keeping everything else constant. Results from this simulation, reveal that the average probability of locating R&D in a city increases from 0.14% to 0.49%. Similarly, to appreciate the economic significance of the

heterogenous propensity to co-locate international production and R&D activities, we compute the variation in $\hat{\beta}_f$ associated with one-standard deviation change in our main variables of interest. Based on Mod. 11, we find that one-standard deviation drop in geographical dispersion, international experience and the share of intangible assets is associated with a 19%, 10.4% and 4% increase in $\hat{\beta}_f$, respectively.

CONCLUSIONS

This study examines the determinants of firms' decisions for the location of international R&D labs, with a special focus on the role of co-location with their previous production plants. Using data on a sample of 1,483 greenfield cross-border investments in R&D activities by 855 firms in a choice set composed by 587 cities worldwide, we find that MNEs are more likely to locate R&D labs in cities where they had previously set-up production plants. However, our findings underline a substantial heterogeneity in firms' preferences. In particular, the *propensity to co-locate* production and R&D is higher for firms that have less international experience and lower geographical dispersion of international activities. This is consistent with the view that co-location between production and R&D allows to compensate for the lack of ability to coordinate operations and manage knowledge transfers across geographically dispersed units. Instead, firms with higher international experience and geographical dispersion have developed such capabilities and this reduces the need for co-locating different activities along the value chain.

Conversely, industry characteristics, such as the nature of the knowledge base, are not robustly associated with more intra-firm co-location. We submit that this may have to do with the fact that there is substantial heterogeneity across firms and within industries. Consistently with this view, we find that more than the nature of the knowledge base of an industry, the extent to which firms (within industries) rely relatively more on codified knowledge (as measured by patents, trademarks and copyrights) matters. The lower the share of intangibles on total assets, the higher the propensity to co-locate production and R&D.

This study provides a number of important contributions to the literature, as well as implications for managerial practice and policy making. First, we contribute to the IB literature by providing a thorough empirical investigation of the factors that can lead to co-locate international R&D with production activities of the same firm, relying on a very comprehensive sample of greenfield investments in R&D activities worldwide, with a fine-grained and unprecedented geographical coverage. In this respect, we contribute to the growing literature highlighting the city as a relevant unit of analysis for studying the activities of MNEs (Belderbos, Du, & Goerzen, 2017; Cook & Pandit, 2018; Goerzen, Asmussen, & Nielsen, 2013; Iammarino & McCann, 2013; Ma, Delios & Lau, 2013; Mudambi, Ma, Makino, Qian, & Boschma, 2018; Stallkamp, Pinkham, Schotter & Buchel, 2018).
Second, we provide a methodological contribution, by proposing a two-stage regression model where we first allow firm heterogeneity in co-location parameters using a Random Parameter (Mixed) Logit location model, and then we uncover the sources of this heterogeneity. This allows us to simultaneously account for several sources of variation in the importance of co-location between production and R&D across firms and industries. We believe that this methodology can be applied in various contexts where the effect of a variable is moderated by several contextual factors, which is quite frequent in strategy and IB. This could help reconcile contradictory empirical findings that were exclusively based on average effects, or revive interest in explanatory variables whose average effects across firms were found to be insignificant. As highlighted by Alcácer et al. (2018) this could allow researchers to test and build new theories at a more granular level.
Third, our study has implications for managerial practice by providing insights on the conditions that make co-location between international production and R&D activities more attractive, considering both firm and industry specificities. This can help MNEs to properly define their investment decision strategies, and becomes extremely important as the coordination and control of different value chain activities may affect the firm's competitiveness, performance and innovation capabilities.
Fourth, we contribute to the policy debate on the attractiveness of places to MNEs. In particular, our results highlight that places that are able to attract manufacturing activities have the potential to move up the value chain, by attracting R&D activities later on. However, since intra-firm co-location

between production and R&D is highly contextual and depends on firm characteristics, our study indicates that the most successful policies may be those targeted towards specific types of investors.

REFERENCES

Alcácer, J., Chung, W., Hawk, A., & Pacheco-de-Almeida, G. 2018. Applying random coefficient models to strategy research: Identifying and exploring firm heterogeneous effects. *Strategy Science*, *3*(3), 533-553.

Alcácer, J., & Delgado, M. 2016. Spatial organization of firms and location choices through the value chain. *Management Science*, 62(11): 3213-3234.

Ambos, B. 2005. Foreign direct investment in industrial research and development: A study of German MNCs. *Research Policy*, 34(4): 395-410.

Asheim, B. T., & Gertler, M. 2005. The Geography of Innovation: Regional Innovation Systems, in Fagerberg, J., Mowery, D., and Nelson, R. (eds.), The Oxford Handbook of Innovation. Oxford University Press, Oxford, 2005, 291-317.

Asheim, B. T. 2007. Differentiated knowledge bases and varieties of regional innovation systems. Innovation: The European Journal of Social Science Research, 20(3): 223-241.

Barkema H.G., Shenkar, O. Vermeulen, F., & Bell, J. 1997. Working abroad, working with others: how firms learn to operate international joint ventures. *Academy of Management Journal*, 40(2): 426–442.

Barkema, H.G., & Vermeulen, F. 1999. International expansion through start-up or acquisition: a learning perspective. *Academy of Management Journal*, 41(1): 7–26.

Basile, R., Castellani, D., & Zanfei, A. 2008. Location choices of multinational firms in Europe: the role of EU cohesion policy. *Journal of International Economics*, 74 (2): 328–40.

Basile, R., Castellani, D., & Zanfei, A. 2009. National boundaries and the location of multinational firms in Europe. *Papers in Regional Science*, 88 (4): 733–748.

Belderbos, R., Du, H. S., & Goerzen, A. 2017. Global cities, connectivity, and the location choice of MNC regional headquarters. *Journal of Management Studies*, 54(8): 1271–1302.

Belderbos, R., Sleuwaegen, L., Somers, D., & De Backer, K. 2016. Where do locate innovative activities in global value chain. Does co-location matter? OECD Science, Technology and Industry Policy Papers, No. 30, OECD Publishing, Paris. http://dx.doi.org/10.1787/5jlv8zmp86jg-en.

Blanc, H., & Sierra, C. 1999. The internationalisation of R&D by multinationals: a trade-off between external and internal proximity. *Cambridge Journal of Economics*, 23(2): 187-206.

Broekel, T., & Boschma, R. 2011. Aviation, space or aerospace? Exploring the knowledge networks of two industries in the Netherlands. *European Planning Studies*, 19(7):1205– 27.

Cantwell, J.A. 1989. *Technological Innovation and Multinational Corporations*. Blackwell, Oxford.

Castellani, D., & Zanfei A. 2004. Choosing international linkage strategies in the electronics industry: the role of multinational experience. *Journal of Economic Behavior and Organization*,53(4): 447-475.

Caves, R. E., & Mehra, S. K. 1986. Entry of foreign multinationals into the US manufacturing industries. In M. E. Porter (Ed.), *Competition and global industries*: 449-481. Boston, MA: Harvard Business School Press.

Chung, W. & Alcácer, J. 2002. Knowledge Seeking and Location Choice of Foreign Direct Investment in the United States. *Management Science*, 48 (12): 1534-1554. https://doi.org/10.1287/mnsc.48.12.1534.440

Cook, G., & Pandit, N. 2018. Cities and International Business: insights from cross-disciplinary perspectives, in G. Cook, J. Johns, F. McDonald, J. Beaverstock and N. Pandit (eds) *The Routledge Companion to International Business and Economic Geography*, Routledge.

Defever, F. 2006. Functional fragmentation and the location of multinational firms in the enlarged Europe. *Regional Science and Urban Economics*, 36(5): 658-677.

Defever, F. 2012. The spatial organization of multinational firms. *Canadian Journal of Economics/Revue canadienne d'économique*, 45(2): 672–697.

Dow, D. & Karunaratna, A. 2006. Developing a multidimensional instrument to measure psychic distance stimuli. Journal of International Business Studies, 37(5): 578-602.

Dunning, J.H., 1996. The geographical sources of the competitiveness of firms: some results of a new survey. *Transnational Corporations*, 5 (3): 1–29.

Durnev, A., Morck, R., & Yeung, B. 2004. Value enhancing capital budgeting and firm-specific stock returns variation. *Journal of Finance*, 59 (1): 65–105.

Florida, R. 1997. The globalization of R&D: Results of a survey of foreign-affiliated R&D laboratories in the USA. *Research Policy*, 26(1): 85-103.

Gerybadze, A., & Reger, G. 1999. Globalization of R&D: recent changes in the management of innovation in transnational corporations. *Research Policy*, 28(2-3): 251–274.

Giroud, X. 2013. Proximity and investment: evidence from plant-level data, *Quarterly Journal of Economics*, 128(2): 861–915.

Goerzen, A., Asmussen, C.G., & Nielsen, B. 2013. Global cities and multinational enterprise location strategy. *Journal of International Business Studies*, 44 (5): 427–450.

Gray, J.V., Siemsen, E., & Vasudeva, G. 2015. Colocation still matters: conformance quality and the interdependence of R&D and manufacturing in the pharmaceutical industry. *Management Science*, 61(11): 2760-2781.

Greene, W. H., Hornstein, A. S., & White, L. J. 2009. Multinationals do it better: Evidence on the efficiency of corporations' capital budgeting. *Journal of Empirical Finance, 16*(5): 703-720.

Grillitsch M., Martin R., & Srholec, M. 2017. Knowledge Base Combinations and Innovation Performance in Swedish Regions. *Economic Geography*, 93(5): 458-479.

Hannigan, T. J., Cano-Kollmann, M., & Mudambi, R. 2015. Thriving innovation amidst manufacturing decline: the Detroit auto cluster and the resilience of local knowledge production. *Industrial and Corporate Change*, 24(3), 613-634.

Hawk, A., & Pacheco-de-Almeida, G. 2018. Time compression (dis)economies: An empirical analysis, *Strategic Management Journal*, 39(9): 2489–2516.

Hennart, J. M. A. 2007. The theoretical rationale for a multinationality - performance relationship. *Management International Review*, 47(3): 423-452.

Herstad, S., Aslesen, H. W., & Ebersberger, B. 2014. On industrial knowledge bases, technological opportunities and global innovation network linkages. *Research Policy*, Elsevier, vol. 43(3): 495-504.

Hole, A.R. 2007. Estimating mixed logit models using maximum simulated likelihood. *Stata Journal, 7*(3): 388-401.

Hornstein, A.S., & Greene, W.H. 2012. Usage of an estimated coefficient as a dependent variable. *Economics Letters.* 116(3): 316-318.

Iammarino, S., & McCann, P. 2013. *Multinationals and economic geography. Location, technology and innovation*. Edward Elgar Publishing.

Ivarsson, I., Alvstam, G., & Vahlne, J.E. 2016. Global technology development by colocating R&D and manufacturing: the case of Swedish manufacturing MNEs. *Industrial and Corporate Change*, 26 (1): 149-168.

Kenney, M, & Florida R. 2004. *Locating Global Advantage: Industry Dynamics in the International Economy*. Stanford: Stanford University Press.

Ketokivi, M., & Ali-Yrkkö, J. 2009. Unbundling R&D and manufacturing: postindustrial myth or economic reality? *Review of Policy Research*, 26(1–2): 35–54.

Kogut, B., & Singh, H. 1988. The Effect of National Culture on the Choice of Entry Mode, *Journal of International Business Studies*, 19(3): 411–432.

Li, J., Meyer, K. E., Zhang, H., & Ding, Y. 2018. Diplomatic and corporate networks: Bridges to foreign locations. *Journal of International Business Studies*, 49(6): 659 – 683.

Ma, X., Delios, A., & Lau, C. M. (2013). Beijing or Shanghai? The strategic location choice of large MNEs' host-country headquarters in China. *Journal of International Business Studies*, 44(9), 953-961.

Mariani, M. 2002. Next to production or to technological clusters? The economics and management of R&D location. *Journal of Management and Governance*, 6(2): 131–152.

Martin, R., & Moodysson, J. 2013. Comparing knowledge bases: On the geography and organization of knowledge sourcing in the regional innovation system of Scania, Sweden. *European Urban and Regional Studies*, 20(2):170– 87.

Mudambi, R. 2008. Location, control and innovation in knowledge-intensive industries. *Journal of Economic Geography*, *8*(5), 699-725.

Mudambi, R., Li, L., Ma, X., Makino, S., Qian, G. & Boschma, R. 2018. Zoom in, zoom out: Geographic scale and multinational activity'. *Journal of International Business Studies*, 49(8): 929–941

Nielsen, B.B., Asmussen, C.G., & Weatherall, C.D. 2017. The location choice of foreign direct investments: empirical evidence and methodological challenges. *Journal of World Business*, 52(1): 62-82.

Papanastassiou, M., Pearce, R. & Zanfei, A. (2019) Changing Perspectives on the Internationalization of R&D and Innovation by Multinational Enterprises. A Review of the literature, *Journal of International Business Studies, forthcoming*.

Pearce, R. 1989. *The internationalisation of Research and Development by Multinational Enterprises*, Macmillan, Basingstoke.

Pisano, G.P., & Shih, W.C. 2009. Restoring American competitiveness. Harvard Business Review, 87, nos. 7-8 (July–August 2009).

Pisano, G.P., & Shih, W.C. 2012. Does America really need manufacturing? *Harvard Business Review*, 90(3), 94–102.

Plum, O., & Hassink, R. 2011. Comparing knowledge networking in different knowledge bases in Germany. *Papers in Regional Science*, 90(2):355– 71.

Rasciute, S., & Downward, P. 2017. Explaining variability in the investment location choices of MNEs: an exploration of country, industry and firm effects. *International Business Review*, 26(4): 605-613.

Rawley, E., & Seamans, R. 2015. Intra-Firm Spillovers? The Stock and Flow Effects of Collocation. Columbia Business School Research Paper No. 15-2. Available at SSRN: https://ssrn.com/abstract=2544 518 or http://dx.doi.org/10.2139/ssrn.2544518.

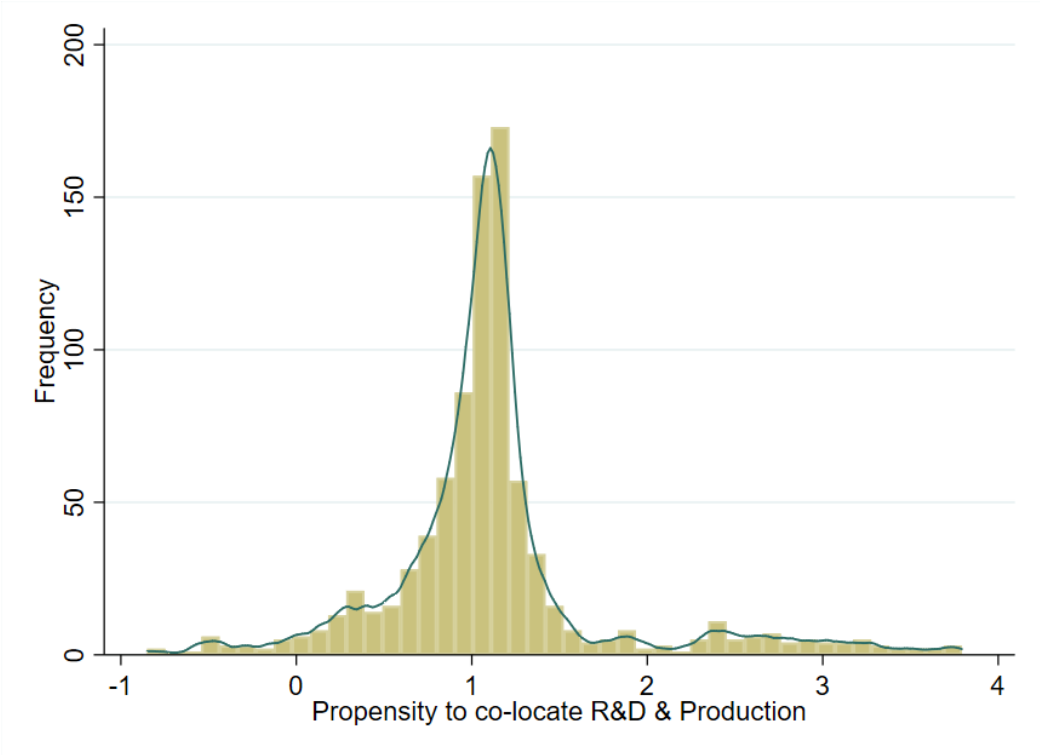
Revelt, D., & Train, K. 1999. Customer-specific taste parameters and mixed logit. *University of California, Berkeley*.

Saxonhouse, G. R. 1976. Estimated parameters as dependent variables. *The American Economic Review*, 66(1): 178-183.

Stallkamp, M., Pinkham, B.C., Schotter, A.P.J., & Buchel, O. 2018. Core or periphery? The effects of country-of origin agglomerations on the within-country expansion of MNEs. *Journal of International Business Studies*, 49(8):942–966.

Train, K.E. .2003. *Discrete Choice Methods with Simulation*. Cambridge University Press, Cambridge

Figure 1: Kernel density distribution of the simulated values of $\hat{\beta}_f$



Source: authors' elaboration

Table 1: Variable List and Description

Variables	Descriptions	Source	Type
First-Stage Model Specification			
DV: Location	R&D location choice among 587 cities over the period 2009-2014; it assumes value 1 if the firm choose the city *l**; 0 otherwise.	fDi Markets	Firm-City
Intra-firm co-location			
Prior Production (dummy)	Dummy =1 if firm cumulated investments in Production, in the city *l* from 2003 until time t-1 is greater than zero	fDi Markets	Firm-City
Prior Other activities (dummy)	Dummy =1 if firm cumulated investments in other activities (e.g. R&D, sales & marketing, HQs, etc.) in the city *l* from 2003 until time t-1 is greater than zero	fDi Markets	Firm-City
City characteristics			
Aggl. economies in R&D	log of the cumulated number of R&D investments made by all MNEs in the city *l* from 2003 until time t-1	fDi Markets	City
Aggl. economies in Prod	log of the cumulated number of Production investments made by all MNEs in the city *l* from 2003 until time t-1	fDi Markets	City
Aggl. economies in Other activities	log of the cumulated number of investments in other activities made by all MNEs in the city *l* from 2003 until time t-1	fDi Markets	City
Geographical Distance	log of Geographical distance between firm home city and the potential host cities	fDi Markets	Firm-City
Population	log of population in the city *l*	UN	City
Home-country specific agglomerations	log of the cumulated number of investments made by MNEs from the same country of company, in the city l in the period 2003 until time t-1	fDi Markets	City
Num. patents	log of the number of patents, by the city of inventors	Harvard Dataverse	City
Country characteristics			
Firm host-country specific experience	Firm cumulated investments in the destination country from 2003 until time t-1, in all the activities	fDi Markets	Country
GDP per capita	GDP per capita (in log)	World Bank	Country
GDP	GDP (in log)	World Bank	Country
Taxation	Taxation rate (in log)	World Bank	Country
Rule of Law	Rule of Law indicator, capturing the rules of society, and in particular the quality of contract enforcement, property rights, and the courts, likelihood of crime and violence.	World Bank	Country
Psychic Distance			
Industrial development	Differences in the degree of industrial development between home and host countries	Dow and Karunaratna	Country
degree of democracy	Differences in the degree of democracy between home and host countries	Dow and Karunaratna	Country
Education Level	Differences in the education levels between home and host countries	Dow and Karunaratna	Country
Religion	Differences in religion between home and host countries	Dow and Karunaratna	Country
Language	Differences in language between home and host countries	Dow and Karunaratna	Country
Second-Stage Model Specification			
Firm Characteristics			
Geographical dispersion	Number of foreign countries in which the company has at least one investment in the period 2003-2008 (as a share of the total number of countries where all firms have at least one investment)	fDi Markets	Firm
International experience	Number of years since the first investment of the firm, until 2008	fDi Markets	Firm
Firm age	Log of the number of years since the year of incorporation of the company, until 2008	Orbis	Firm
Share of intangible assets	Share of intangible assets in total assets	Orbis	Firm
Share of foreign subsidiaries	Share of foreign subsidiaries in the total number of (national and foreign) subsidiaries	Orbis	Firm
Num. patents	Log of number of patents owned by the firm	Orbis	Firm
Large firm	Dummy=1 if the company is classified as 'large' or 'vary large' by Bureau Van Dijk's Orbis	Orbis	Firm
Sector Characteristics			
Synthetic Sectors (dummy)	Dummy =1 if the sector is classified as 'synthetic', 0 as 'analytical', by the empirical EG literature	EG Literature	Firm

Table 2: The determinants of the location of offshore R&D labs, first-stage Mixed Logit regression

	R&D Location Decision							
	Mod. 1				Mod. 2			
	Mean		*Std. Dev.*		*Mean*		*Std. Dev.*	
	(1)		(2)		(3)		(4)	
Intra-firm co-location city level								
Prior firm Production	1.111	[0.0000]	1.8733	[0.0000]	3.2978	[0.0000]	1.58	[0.0000]
	(0.1774)		(0.2348)		(0.5592)		(0.3342)	
Prior firm Production*Geographical Dispersion					-5.7977	[0.0139]		
					(2.3581)			
Prior firm Production*International experience					-0.2909	[0.0077]		
					(0.1092)			
Prior firm Production*Synthetic Sector					-0.2989	[0.3010]		
					(0.2890)			
Prior firm Other activities	0.7097	[0.0000]	-0.6116	[0.0020]	0.7966	[0.0000]	-0.6429	[0.0043]
	(0.1106)		(0.1979)		(0.1078)		(0.2253)	
City characteristics								
Agglomeration economies in R&D	0.3435	[0.0000]	0.3273	[0.0000]	0.3039	[0.0000]	0.1591	[0.0014]
	(0.0455)		(0.0380)		(0.0449)		(0.0497)	
Agglomeration economies in Prod	0.1171	[0.0017]	0.0732	[0.5287]	0.1669	[0.0000]	0.2275	[0.0000]
	(0.0373)		(0.1162)		(0.0405)		(0.0472)	
Agglomeration economies in Other activities	0.0373	[0.3715]	0.0069	[0.8388]	0.0309	[0.4642]	-0.0694	[0.0151]
	(0.0417)		(0.0338)		(0.0423)		(0.0285)	
Geographical Distance between home and host cities (log)	0.0586	[0.2321]	0.3376	[0.0000]	0.0253	[0.5690]	0.2036	[0.0458]
	(0.0491)		(0.0761)		(0.0445)		(0.1019)	
Population	0.0985	[0.0158]	0.2455	[0.0000]	0.1069	[0.0093]	0.196	[0.0009]
	(0.0408)		(0.0404)		(0.0411)		(0.0591)	
Home-country specific agglomerations	0.4162	[0.0000]	0.0156	[0.8445]	0.4145	[0.0000]	0.0762	[0.0925]
	(0.0473)		(0.0798)		(0.0477)		(0.0453)	
Num. patents	0.0626	[0.0020]	0.0271	[0.6743]	0.0658	[0.0012]	-0.0744	[0.0271]
	(0.0203)		(0.0646)		(0.0202)		(0.0337)	
Country characteristics								
Firm host-country specific experience	0.0043	[0.6336]	0.0497	[0.0015]	0.0046	[0.6219]	0.0377	[0.0010]
	(0.0089)		(0.0157)		(0.0092)		(0.0115)	
GDP per capita	0.1578	[0.0643]	0.0276	[0.6407]	0.1322	[0.1271]	0.0385	[0.4473]
	(0.0853)		(0.0592)		(0.0867)		(0.0507)	
GDP	0.169	[0.0000]	-0.1766	[0.0001]	0.1819	[0.0000]	-0.2313	
	(0.0347)		(0.0451)		(0.0364)		(0.0401)	
Taxation	-0.342	[0.0116]	0.088	[0.6359]	-0.3621	[0.0090]	-0.3917	[0.0320]
	(0.1355)		(0.1859)		(0.1386)		(0.1827)	
Rule of Law	0.1082	[0.1841]	-0.1414	[0.0569]	0.1262	[0.1217]	0.0179	[0.7920]
	(0.0815)		(0.0743)		(0.0815)		(0.0680)	
Psychic Distance: Industrial development	0.2342	[0.1353]	-0.0585	[0.4325]	0.3056	[0.0526]	0.0144	[0.9035]
	(0.1568)		(0.0745)		(0.1577)		(0.1189)	
Psychic Distance: degree of democracy	-0.0803	[0.1827]	-0.0391	[0.6102]	-0.1252	[0.0406]	-0.0795	[0.2024]
	(0.0603)		(0.0768)		(0.0611)		(0.0623)	
Psychic Distance: Education Level	0.2704	[0.0006]	0.2658	[0.0050]	0.2627	[0.0008]	-0.1305	[0.2175]
	(0.0783)		(0.0947)		(0.0782)		(0.1058)	
Psychic Distance: Religion	-0.1565	[0.0034]	0.3174	[0.0000]	-0.1669	[0.0020]	0.3326	[0.0000]
	(0.0534)		(0.0746)		(0.0540)		(0.0668)	
Psychic Distance: Language	-0.0318	[0.4528]	0.2777	[0.0002]	-0.0456	[0.2763]	0.226	[0.0006]
	(0.0423)		(0.0749)		(0.0419)		(0.0658)	
No. observations	795,326				779,501			
No. projects	1,483				1,483			
No. cities	587				587			
Time period	2009-2014				2009-2014			
Simulated Log-L MLM	-7428.955				-7277.171			
χ^2	168.9685				125.3505			

Note: The dependent variable is equal to 1 if firm f sets up an R&D lab in city l* and zero for all cities l different from l*. The choice set consists of a maximum of 587 cities. Cities in the home country of the investing firm are removed from the location choice set.
Standard errors in parenthesis below point estimates, p-values in square brackets next to the point estimates.

Table 3: Uncovering the sources of heterogeneity in the propensity to co-locate production and R&D, second-stage OLS regression

	Propensity to co-locate R&D and Production										
	Total Sample							Orbis subsample			
OLS	**Mod. 1**	**Mod. 2**	**Mod. 3**	**Mod. 4**	**Mod. 5**	**Mod. 6**	**Mod. 7[§]**	**Mod. 8**	**Mod. 9**	**Mod. 10[§]**	**Mod. 11[§]**
Firm's Characteristics											
Geographical dispersion	-1.6479			-1.9405	-2.0697	-2.0829	-3.3036	-2.9916	-3.4966	-3.7025	-4.0443
	(0.8794)			(1.1789)	(1.1742)	(1.1715)	(1.0086)	(1.6513)	(1.9200)	(1.4069)	(1.6067)
	[0.0613]			[0.1001]	[0.0783]	[0.0758]	[0.0011]	[0.0709]	[0.0695]	[0.0089]	[0.0124]
International experience		-0.0119		0.0052	0.0031	0.0022	-0.0356	0.0098	-0.0106	-0.0352	-0.0363
		(0.0094)		(0.0129)	(0.0125)	(0.0124)	(0.0081)	(0.0201)	(0.0225)	(0.0123)	(0.0135)
		[0.2045]		[0.6856]	[0.8067]	[0.8572]	[0.0000]	[0.6271]	[0.6391]	[0.0045]	[0.0075]
Share of intangible assets								-0.352	-0.3421	-0.1768	-0.1763
								(0.2072)	(0.2143)	(0.1153)	(0.1216)
								[0.0902]	[0.1115]	[0.1262]	[0.1481]
Firm age									0.0415		0.0481
									(0.0475)		(0.0283)
									[0.3829]		[0.0902]
Number of patents									0.0205		-0.0107
									(0.0166)		(0.0092)
									[0.2169]		[0.2460]
Large firm									0.1112		0.0245
									(0.0945)		(0.0525)
									[0.2401]		[0.6408]
Share of foreign subsidiaries									-0.1678		0.051
									(0.1619)		(0.0764)
									[0.3008]		[0.5052]
Industry Characteristics											
Synthetic Sectors (dummy)			0.0317	0.0526							
			(0.0581)	(0.0581)							
			[0.5853]	[0.3656]							
Food & Beverages					-0.0036	-0.0044	0.0541	-0.0219	0.0337	0.0397	0.0272
					(0.0902)	(0.0917)	(0.0515)	(0.1588)	(0.1691)	(0.0737)	(0.0736)
					[0.9684]	[0.9621]	[0.2937]	[0.8905]	[0.8420]	[0.5907]	[0.7116]
Machinery					-0.08	-0.0702	-0.0636	-0.1152	-0.0966	-0.1587	-0.1767
					(0.0710)	(0.0703)	(0.0420)	(0.1353)	(0.1372)	(0.0779)	(0.0769)
					[0.2604]	[0.3186]	[0.1297]	[0.3954]	[0.4821]	[0.0424]	[0.0223]
Metals					0.0209	0.052	0.0046	0.5963	0.6087	0.2882	0.2403
					(0.2053)	(0.2053)	(0.0913)	(0.4857)	(0.4857)	(0.1806)	(0.1951)
					[0.9188]	[0.8002]	[0.9601]	[0.2203]	[0.2110]	[0.1116]	[0.2191]
Mining					-0.0735	-0.0603	0.0068	-0.3582	-0.4145	-0.0775	-0.1109
					(0.1738)	(0.1725)	(0.0745)	(0.1295)	(0.1422)	(0.1001)	(0.1036)
					[0.6724]	[0.7269]	[0.9274]	[0.0060]	[0.0038]	[0.4394]	[0.2852]
Other					0.0472	0.0537	0.0495	-0.2528	-0.2323	-0.0772	-0.0878
					(0.0903)	(0.0910)	(0.0468)	(0.1521)	(0.1545)	(0.1026)	(0.1050)
					[0.6011]	[0.5551]	[0.2901]	[0.0973]	[0.1335]	[0.4527]	[0.4036]
Pharma, Chemicals & Biotech					0.0497	0.0446	-0.0327	0.0105	0.0463	-0.0861	-0.0985
					(0.0637)	(0.0636)	(0.0374)	(0.1141)	(0.1204)	(0.0632)	(0.0648)
					[0.4352]	[0.4831]	[0.3827]	[0.9264]	[0.7005]	[0.1737]	[0.1298]
Plastics & Rubber					0.3916	0.4115	-0.1078	0.5942	0.6904	-0.1833	-0.1943
					(0.2225)	(0.2221)	(0.0770)	(0.3489)	(0.3681)	(0.1000)	(0.1067)
					[0.0788]	[0.0643]	[0.1618]	[0.0895]	[0.0616]	[0.0678]	[0.0698]
Transport					0.2681	0.2778	0.048	0.2972	0.2901	0.0737	0.0431
					(0.1119)	(0.1116)	(0.0598)	(0.1636)	(0.1713)	(0.0839)	(0.0815)
					[0.0168]	[0.0130]	[0.4219]	[0.0701]	[0.0913]	[0.3804]	[0.5975]
USA (dummy)						0.1429	0.084	0.4032	0.4853	0.1668	0.1689
						(0.0639)	(0.0351)	(0.1282)	(0.1307)	(0.0731)	(0.0713)
						[0.0256]	[0.0169]	[0.0018]	[0.0002]	[0.0231]	[0.0186]
EU (dummy)						0.099	0.0706	0.2813	0.4192	0.18	0.1451
						(0.0607)	(0.0325)	(0.1020)	(0.1190)	(0.0541)	(0.0659)
						[0.1035]	[0.0300]	[0.0061]	[0.0005]	[0.0010]	[0.0285]
Constant	1.1739	1.1687	1.1087	1.1231	1.1359	1.0416	1.0581	0.9907	0.7332	1.0415	0.9177
	(0.0244)	(0.0215)	(0.0511)	(0.0489)	(0.0374)	(0.0554)	(0.0322)	(0.1022)	(0.1958)	(0.0610)	(0.1192)
	[0.0000]	[0.0000]	[0.0000]	[0.0000]	[0.0000]	[0.0000]	[0.0000]	[0.0000]	[0.0002]	[0.0000]	[0.0000]
R-squared	0.0082	0.0020	0.0003	0.0093	0.0325	0.0379	0.2229	0.0962	0.1117	0.2725	0.2869
Adjusted R-squared	0.0070	0.0008	-0.0009	0.0058	0.0199	0.0231	0.2095	0.0593	0.0631	0.2376	0.2409
No. of observations	855	855	855	855	855	855	770	358	348	307	298

Note: The dependent variable is the propensity to co-locate R&D and production, measured as the firm-level parameters associated with prior production of the same firm city l simulated from the first-stage MLM ($\hat{\beta}_f$)

Robust Standard errors in parenthesis, p-values in square brackets below the point estimates.

[§]Model estimated eliminating observations (firms) with a propensity to co-locate ($\hat{\beta}_f$) greater than the 90th percentile.

[i] See Papanastassiou et al., (2019) for a recent survey of studies on the internationalization of R&D.

[ii] In particular, 1,770 of 6,235 (28.39%) identify projects in Research & Development (R&D) activity**,** while 4,465 (71.61%) identify projects in Design, Development and Testing (DDT). We could expect that the former involves more basic (or science-intensive) research activities, whereas the latter reflects the more applied/adaptive type of research. However, an inspection of the description of the projects does not allow to clearly gauge differences between the two. Therefore, we decided to consider them jointly.

[iii] *fDi Markets* provides information on expansion of existing projects, which account for about 19.47% of all projects in R&D and DDT. We have decided not to consider these projects because these are by definition determined by prior R&D investments, therefore the effect of any prior production plants would 'absorbed' in the effect of prior R&D labs.

[iv] The choice of focussing on R&D projects in manufacturing industries is dictated by the fact that service firms are not expected to engage in manufacturing regularly, thus they not constitute a proper setting to test our hypotheses.

[v] Descriptive statistics are available in the Appendix.

[vi] In our empirical analysis, we allow J to vary across firms from a minimum of 1 to a maximum of 25. In our data, 621 firms (72.6%) engage in only one R&D investment decision, and are responsible for 42% of all R&D investment projects. Instead, 58% of the projects are made by 234 firms (27.4%) that make more than one location decision throughout the period. Out of these firms, 90% make up to 6 decisions and only 25 firms make more than 6 decisions.

[vii] A possible route would also be to run regressions on sub-samples, which has the advantage of not imposing common coefficients on the non-interacted variables, but it does not allow to account for several simultaneous sources of heterogeneity.

[viii] Indeed, the interacted model has also advantages, since it allows more directly to estimate the range of the moderating variables for which the effect is statistically different from zero.

[ix] Saxonhouse (1976) proposed to recover a vector of individual-specific estimates from separate regressions for each individual and then use this vector as a dependent variable in a second-stage OLS regression. Clearly, this approach requires long time series for each firm, which is unlikely in most circumstances, and it certainly does not fit the structure of our data.

[x] In principle, the second stage need not be an OLS regression. For example, it could be modelled using a quantile regression model or transform the dependent variable into a binary indicator and estimate a probit model. This makes this approach particularly flexible and interesting, as opposed to, for example, the interacted model.

[xi] This dummy is based on an industry taxonomy in line with the innovation and economic geography literature. The following industries are classified as 'synthetic' knowledge bases: aerospace, automotive, business machines and equipment, coal, oil and natural gas, consumer electronics, electronic components, engines and turbines, food and tobacco, metals, minerals, plastics, rubber, semiconductors and textiles. The following industries are classified as analytical knowledge base: biotechnology, chemicals, healthcare, medical devices, pharmaceuticals, space and defence. For more details, Asheim (2007); Broekel & Boschma (2011); Herstad, Aslesen & Ebersberger (2014); Martin & Moodysson (2013).

[xii] Our baseline regression requires quite some time to converge, and with three interactions it requires a careful choice of the starting values. In our case, with more than three interactions convergence is quite difficult to achieve.

[xiii] To obtain the simulated firm-level parameters, we use the Stata command `mixlbeta`, which is a post-estimation command of `mixlogit` (Hole, 2007). More information at http://fmwww.bc.edu/RePEc/bocode/m/mixlogit.html..

[xiv] To obtain the predicted probabilities, we use the Stata command `mixlpred`, which is a post-estimation command of `mixlogit` (Hole, 2007).