Regional knowledge spillovers: A firm-based analysis of nonlinear effects

Carlos Carreira[*] and Luís Lopes[§]

[*] *CeBER and Faculty of Economics, University of Coimbra. Email: ccarreir@fe.uc.pt*

[§] *CeBER and Faculty of Economics, University of Coimbra. Email: perlopes@fe.uc.pt*

Abstract

This article revisits the question of the role of knowledge externalities in firm productivity. It also addresses the overlooked issue of a plausible nonlinear effect and differences among industries. Using a panel of Portuguese manufacturing firms, it finds that regional knowledge spillovers differ substantially across industries and they are nonlinear, which is critical issue to promoting more assertive regional policies.

Keywords: Agglomeration economies; Regional knowledge spillovers; Total factor productivity; Low-tech, medium-tech and high-tech industries.

JEL classification: R12; L25; O12; D24

INTRODUCTION

The literature has greatly emphasized the role played by both knowledge spillovers and agglomeration economies as drivers of firms performance (Audretsch and Feldman, 2004; Döring and Schnellenbach, 2006). However, while it has been widely accepted that knowledge spillovers have a positive effect on firm productivity, the complete picture of how knowledge agglomeration impacts on a firm productivity remains puzzling (Breschi and Lissoni, 2001; Tappeiner et al., 2008; de Groot et al., 2016).

Different streams of literature have provided clear evidence that knowledge transmission is limited by distance and therefore bounded in space (Baptista, 2000; Audretsch and Feldman, 2004; Rosenthal and Strange, 2004). The proximity to knowledge sources encourages the circulation of ideas and the transmission of knowledge thanks to face-to-face contacts and social networks, which allow units operating close to those sources to learn and innovate at a faster rate than rival firms located elsewhere (Baptista, 1999; Storper and Venables, 2004; Fritsch and Kauffeld-Monz, 2010; Lambooy, 2010). The presence of knowledge mediators in the local milieu can accelerate this knowledge-transfer by fostering face-to-face contacts between organizations that are not directly linked (Kauffeld-Monz and Fritsch, 2013). In fact, the local milieu where firms are located clearly matters for transmitting knowledge.

Technologies are path-dependent since they evolve as a historical process along technological trajectories (Nelson and Winter, 1982). Industry-specific technological regimes strongly determine potential learning effects (Marsili, 2002; Carreira and Teixeira, 2011a). This implies that firm learning is likely to vary between industries, but it is relatively invariant across firms within the same industry (Klevorick et al., 1995; Malerba and Orsenigo, 1997; Grimpe and Sofka, 2009). Aside from its academic interest, the question is critical to formulating regional policy because there can be considerable

differences regarding the knowledge base and competences between regions and therefore there can be diverse growth regimes in different regions (Baptista and Swann, 1998; Audretsch and Fritsch, 2002; Fritsch and Falck, 2007).

This article contributes to the debate by studying the role played by regional knowledge spillovers in firm productivity among Portuguese industries. Spillovers occur whenever an individual or organization shares knowledge with its neighbours without having to pay for such knowledge in a market transaction. They have been identified as one of the main explanations of agglomeration externalities, so the productivity of firms should be influenced significantly by the access to these sources of knowledge (Audretsch et al., 2005; Baptista and Mendonça, 2010). Additionally, it looks at the nonlinearities in the effect of knowledge spillovers on firm-level total factor productivity (TFP), exploring differences among high-, medium-high-, medium-low- and low-technology industries, following the OECD classification of manufacturing industries (OECD 2005).

The empirical enquiry is based on a large sample of Portuguese manufacturing firms, covering the period 1996-2004. It is worth noting, first, that the use of firm-level data may help when dealing with endogeneity issues that are often found in analyses that rely on city-industry or region-industry aggregate data (van Oort et al., 2012). Furthermore, as it will be discussed in the next section, it allows us to address the risk of ecological fallacy that can occur when we draw inferences about firms based only on analyses of aggregate data. Second, unlike employment and wages growth, which are observed consequences, TFP is the most direct approach to capture agglomeration advantages at the firm level (Puga, 2010).

Following Cantwell and Piscitello (2005) and Frenken et al. (2007), the article distinguishes three potential sources of knowledge that firms experience in their local milieu: (1) at the intra-industry level, from the spatial concentration of firms belonging to

the same industry, called *specialization* (or localization) economies—in essence, they rely on the spread of specialized knowledge through firms within an industry, the increased variety and quality of shared inputs and the specialized labour market pooling (Marshall, 1890, Arrow, 1962; Romer, 1986); (2) at the inter-industry level, from the wide variety of industrial activities outside the own industry involving some type of cross-fertilization between industries, known as *diversity* economies (Jacobs, 1969);[1] (3) at the regional level, from improved access to public and private science-technology institutions, external to industries but internal to the region, such as universities, research laboratories, trade associations and other knowledge-generating and transmitting organizations, which this study will call *science-technology* economies (Audretsch et al., 2005; Fritsch and Slavtchev, 2007; Cassia et al., 2009; Baptista and Mendonça, 2010; Acs et al., 2013; Fritsch and Aamoucke, 2013).[2] An important question concerns whether each of these agglomeration economies actually contributes to productivity gains. Given that firms can learn from different sources of knowledge, in theory, all three types of agglomeration economies can occur. However, different sources of spillovers are expected to lead to different effects on productivity among industries because firms learn differently. For example, location economies *a priori* should be particularly significant for low-tech industries, while access to science-technology institutions should be more important for high-tech industries. Another key question is whether firm productivity benefits most from a given amount of agglomeration. As the amount of spillovers differs between both industries and regions, the question is which precise density in a region-industry creates most spillovers. One should expect there to be certain nonlinearities, such as marginally decreasing/increasing economies, and consequently a peak agglomeration size that maximizes the productivity gains.

This study extends the literature by adding several significant new pieces to the puzzle of regional knowledge spillovers. First, the findings on the effects of agglomeration economies on TFP are not new in the literature. However, presumably due to data constraints, most previous studies have relied on aggregate productivity data. Only a small fraction draws inferences from firm- or plant-level data in a panel context, but even these studies differ widely in the variables and empirical methodology that they use, as can be seen in the next section.

Second, although Combes (2000) and Henderson et al. (2001), based on analyses of aggregate data, have noted that there are good reasons to believe that the effects of spillovers are not homogeneous across industries, and to the best of our knowledge other comparable articles do not systematically account for cross-industry differences in agglomeration economies (the rare exceptions are Henderson, 2003; Lall et al., 2004; and Ehrl, 2013).

Third, even if there is no reason to assume a linear relationship between knowledge spillovers and firm productivity, since theory in economic geography and urban economics suggests that agglomeration can lead to both economies and diseconomies, little attention has so far been paid to the potential nonlinear features of such a relationship (the few exceptions are de Lucio et al., 2002, and Au and Henderson, 2006, at aggregate-level; and Martin et al., 2011, and Cainelli et al., 2015, at micro-level). Addressing potential nonlinearities, the article not only fills an important gap in the literature, but it is also able to estimate the density of spillovers that maximizes productivity gains. Finally, the impact of science-technology economies on firm productivity, in particular the role of knowledge-intensive business services (KIBS), and the question of specialization and diversity economies were never considered simultaneously. Integrating both streams of research

allows us to provide a broader perspective of the issue by comparing the impact of each of these sources of knowledge.

To anticipate some of the main findings, this study finds considerable heterogeneity in the sources and magnitudes of regional knowledge spillovers among industries. And, moreover, they are nonlinear.

The remainder of the article is organized as follows. The next section presents an overview of the literature on the role of spatial knowledge agglomeration in firm productivity. The third section discusses the methodology and the main empirical results and their discussion are presented in the fourth section. Finally, the last section offers some. The Appendix in the supplemental data online provides a detailed description of dataset, the descriptive statistics of variables, additional econometric issues and some supplementary results.

RELATED LITERATURE AND HYPOTHESES

Most of the previous studies adopt a regional- or city-level perspective. In a seminal work, for example, using a panel of US cities between 1956 and 1987, Glaeser et al. (1992) found that urban variety has a positive impact on employment growth in industries but regional specialization does not, concluding that knowledge spillovers occur between rather than within industries.

The study of Glaeser et al. (1992) has been replicated to other countries, but with ambiguous results, which can be probably explained by (i) the level of aggregation in data and (ii) the dependent variable used in the analyses (see, for example, Cainelli and Leoncini, 1999; and Forni and Paba, 2002, for Italy; Combes, 2000, for France). In fact, as discussed by van Oort et al. (2012), the theories that underlie agglomeration economies are microeconomic in nature, so showing that regional economic performance is higher in

regions where agglomeration economies are also higher does not mean that a similar relationship holds at the level of the individual firm (see also Duranton and Puga, 2004). As a result, the study of the agglomeration issue has greatly shifted towards the understanding of the operation of micro units (see Audretsch and Feldman, 2004, Döring and Schnellenbach, 2006, or Combes and Gobillon, 2015, for a survey; and Melo et al., 2009, or de Groot et al., 2016, for a meta-analysis).

Since data on productivity are generally not available at firm level, most of the studies use employment or wage growth as a proxy for productivity growth. Cingano and Schivardi (2004), using 1991 Italian census data, showed that the specialization impact on local employment growth is negative and variety has a significant positive effect, in line with Glaeser et al.'s results. However, using TFP as the dependent variable, they found that the specialization effect is reversed and becomes positive and industrial variety does not have any effect. Cingano and Schivardi (2004) thus question the conclusions of previous works, arguing that employment growth based regressions suffer from identification problems because the chain of causality from agglomeration economies to employment growth could be reversed.

Given that knowledge spillovers lead to a change in output that is not fully accounted for by a change in inputs, TFP would be a better measure of performance (Puga, 2010; Combes and Gobillon, 2015). Nevertheless, very few studies draw inferences from TFP at firm-level, probably mainly because of data constraints. Anderson and Lööf (2011), for example, found that in the period 1997–2004 Swedish manufacturing firms located in larger regions are more productive after controlling for a firm's participation in international trade, ownership structure and industry affiliation. Moreover, the marginal effect of agglomeration on firm productivity growth can be given a causal interpretation, suggesting a positive learning effect. In the case of Italian manufacturing firms over the

period 1995–2006, di Giacinto et al. (2014) also noted productivity advantages for firms located in both urban areas and industrial districts, with a higher premium for urban areas.

Distinguishing between different sources of knowledge, López and Südekum (2009) found evidence from Chilean manufacturing plants in the period 1990–1999 for intra-industry spillovers, but none for inter-industry spillovers. Similar results were reported by Martin et al. (2011) and Hashiguchi and Tanaka (2015) for French manufacturing plants in the 1996–2004 period and Chinese manufacturing firms in 2004, respectively.

The relevant sources of knowledge may not be limited to industries. Remarkably, however, the previous studies do not address the role of science-technology economies. In fact, while knowledge produced within incumbent firms that spills over to other firms has been extensively analysed, the role of regional science-technology organizations has been mostly neglected. However, favoured by the flourishing endogenous growth theory, empirical evidence seems to support the idea that universities, research laboratories and other local knowledge-generating organizations matter for the innovative activity of firms. For instance, universities and research laboratories have been described in the literature as crucial sources of knowledge, not only for undertaking knowledge-generating research and development (R&D) activities, but also for training qualified human capital capable of absorbing such knowledge (Audretsch et al. 2005; Fritsch and Slavtchev, 2007; Baptista and Mendonça, 2010; Baptista et al., 2011; Fritsch and Aamoucke, 2013). The role of KIBS as a key source of knowledge has also been increasingly recognized. Acting as knowledge mediators that facilitate collaboration between actors in a regional innovation system, KIBS significantly affect the production and diffusion of knowledge within a region (Muller and Zenker, 2001).

Based on these findings the following hypotheses can be formulated:

Hypothesis 1A   Firms benefit from being located near other firms in the same industry.

Hypothesis 1B   Firms benefit from being located near other firms from other industries.

Hypothesis 1C   Firms benefit from being located near science-technology institutions.

Martin et al. (2011) went a step further by testing the existence of nonlinear effects on regional knowledge spillovers. Briefly, the rationale is that based on previous studies it can be concluded that more agglomeration is always better because it increases the productivity of firms, but agglomeration diseconomies such as congestion effects may also exist, besides knowledge spillovers, and these can dominate at a certain level of agglomeration. They concluded that localization economies have the form of a bell-shaped curve with a negative effect for small values of the variable. Cainelli et al. (2015), using a panel of Italian manufacturing firms over the period 1999–2007, also found a positive and statistically significant effect of specialization and diversity economies, but only above a certain threshold, which seems to contradict the congestion theory. Moreover, diversity seems statistically significant only once nonlinear effects are allowed for. These findings give rise to the following, mutually exclusive, hypotheses:

Hypothesis 2A   Agglomeration economies are significant only above a certain critical concentration.

Hypothesis 2B: Congestion effects exist and offset agglomeration economies above a certain threshold.

Very few studies have analysed the effect of local knowledge spillovers across industries, even though we should expect substantial differences in the magnitude of agglomeration economies among industries. Henderson (2003), using the US census data for 1972–1992 period, found that localization economies have a strong positive effect on productivity in high-tech industries, but not in mechanical industries. He also found little evidence of diversity and region size economies. Based on plant-level data from India for

1994–1995 period, Lall et al. (2004) observe considerable variation in the sources and magnitudes of agglomeration economies across industries. Indeed, while they find no statistically significant agglomeration economies in leather products, basic metals and repair industries, agglomeration economies are quite substantial in the paper products and printing, machinery and equipment, and electronics and computers industries. Looking at the different sources of agglomeration economies, the productivity gains from localization economies are somewhat fewer; they are positive and significant in only two industries, paper products and printing, and non-metallic mineral products. The benefits from urban density are not statistically significant either. Ehrl (2013), using German establishment and employment-level data from 2000 to 2007, also found that agglomeration economies differ substantially between industries. The high-tech industry benefits from diversity and job changes of skilled workers, while the medium-tech industry only benefits from specialized labour market pooling. A rather surprising finding, however, is that for low-tech industries, job changes of more qualified workers generate significant productivity gains. One explanation may be that even though R&D investment is relatively low, knowledge is very important for low-tech firms as they compete for the leadership in technology and design. Apparently, local productive specialization has no impact on productivity in any industry. Thus, it will be expected that:

Hypothesis 3   The magnitude and source of agglomeration economies vary between industries.

EMPIRICAL METHODOLOGY

The data

The empirical study is based on an unbalanced panel of Portuguese manufacturing firms covering the period 1996–2004. The survey comprises all firms with more than 100

employees, plus a representative random sample of firms with 20 to 99 employees. It was excluded from the dataset firms located in the island regions (i.e. Madeira and Azores). (Appendix A in the Supplemental data online provides a more detailed description of dataset.)

Figure 1 shows the number of firms per region. There appears to be a substantial agglomeration of manufacturing firms in the coastal regions of central and northern Portugal, mainly in the central region of Grande Lisboa and northern regions of Grande Porto and Ave.

[Figure 1]

The model

The empirical analysis is based on a firm-level Cobb-Douglas production function:[3]

$$Y_{it} = A_{it} K_{it}^{\alpha} L_{it}^{\beta} M_{it}^{\theta}, \tag{1}$$

where subscripts i and t denote firm and year, respectively. Y_{it} is the real gross output and K_{it}, L_{it} and M_{it} are capital, labour and material inputs, respectively; α, β and θ are the output elasticities of each factor; and A_{it} is the TFP total factor productivity, which is assumed to depend on local milieu:

$$A_{it} = \left(LOC_{it}^{jr}\right)^{\gamma} \left(DIV_{it}^{jr}\right)^{\varphi} \left(SCI_{it}^{r}\right)^{\phi}, \tag{2}$$

where LOC_{it}^{jr}, DIV_{it}^{jr} and SCI_{it}^{r} are measures of localization (or specialization), diversity and science-technology economies, respectively, for firm i in year t (which operate in industry j and region r).

Construction of agglomeration variables

The localization economies are measured for each firm as the share of other employees working in the same industry (at the two-digit level) within the region (Cingano and Schivardi, 2004; Saito and Gopinath, 2009; Combes and Gobillon, 2015):[4]

$$LOC_{it}^{jr} = \frac{L_t^{jr} - L_{it}}{L_t^r - L_{it}},$$ (3)

where L_t^{jr} and L_t^r are the number of employees in industry j and in all manufacturing industries, respectively, in region r, in year t.

The diversity economies are proxied by the entropy index of regional employment shares outside the industry j (Frenken et al. 2007; Cainelli et al., 2015):[5]

$$DIV_t^{jr} = \sum_{j' \neq j} \frac{L_t^{j'r}}{L_t^r - L_t^{jr}} \ln\left(\frac{L_t^r - L_t^{jr}}{L_t^{j'r}}\right),$$ (4)

where $L_t^{j'r}$ is the number of employees in industry j' ($j' \neq j$). The index ranges between 0, the minimum value, when all other manufacturing employment in the region is concentrated in a single industry, and $ln(J^r - 1)$, the maximum value, if this employment is uniformly distributed across all (other) industries (J^r denotes the number of industries in region r). The value of diversity index is not directly linked with the specialization index. In fact, for example, if part of regional employment is highly concentrated in a particular industry and if the remaining employment is uniformly distributed over all other industries, the value of both indexes (specialization and diversity) for this industry is high (Combes, 2000).

Finally, the science-technology economies are measured by two variables: the number of R&D workers in region r, RD_t^r, as a proxy of local knowledge spillovers generated by universities, research laboratories, trade associations and other science-

technology institutions;[6] and the number of Knowledge Intensive Business Services workers, $KIBS_t^r$, as a proxy of the role played by knowledge-transmitting organizations.[7]

Estimation strategy

Agglomeration variables may be endogenous. To deal with this issue, Combes and Gobillon (2015) suggest the implementation of a two-step approach where the production function is estimated in a first step without introducing local variables to compute firm-level TFP. Then, the firm TFP is regressed in a second-step on agglomeration economies dealing with the endogeneity of regional characteristics (Martin et al., 2011; di Giacinto et al., 2014; Cainelli et al., 2015).

The log-linear transformation of model (1) is estimated using a two-step system Generalized Method of Moments (GMM) estimator to compute consistent estimates (Arellano and Bover, 1995). Ordinary least squares (OLS) estimates may suffer from endogeneity problems due to simultaneity and selection bias. The equation cannot be estimated using the within estimator (FE) either, since it requires the strict exogeneity of regressors, a non-realistic assumption. In contrast, the system GMM estimator provides consistent estimates in the presence of endogeneity issues (Blundell and Bond, 2000). (More econometric issues and production function results are given in Appendix B.)

After estimating the output elasticity coefficients, the firm-level TFP is computed as the residual:

$$\hat{a}_{it} = y_{it} - \hat{\alpha}\, k_{it} - \hat{\beta}\, l_{it} - \hat{\theta}\, m_{it}, \tag{5}$$

where lower-case letters denote the log of corresponding upper-case variables in model (1).

In the second stage, the TFP is regressed on both proxies aimed at capturing the effect of the regional knowledge spillovers, as proposed in model (2), and firm age to control for firm stock of knowledge.[8] The rationale is that older firms are likely to have accumulated more knowledge (Carreira and Teixeira, 2011b). To account for the existence of nonlinear effects, it is introduced in the log-linear specification quadratic terms as follows:[9]

$$
\begin{aligned}
a_{it} = \lambda &+ \gamma_1 loc_{it}^{jr} + \gamma_2 \left(loc_{it}^{jr}\right)^2 + \varphi_1 div_t^{jr} + \varphi_2 \left(div_t^{jr}\right)^2 + \phi_1 rd_t^r + \phi_2 \left(rd_t^r\right)^2 \\
&+ \psi_1 kibs_t^r + \psi_2 \left(kibs_t^r\right)^2 + \pi_1 age_{it} + \pi_2 \left(age_{it}\right)^2 + \omega_t + v_j + \upsilon_i + \mu_{it}
\end{aligned}
, \tag{6}
$$

where ω_t, v_j and υ_i are year, industry and firm-fixed effects, respectively; and μ_{it} is a standard error term. The model cannot disentangle firm and regional fixed effects. A common strategy used to deal with this issue is to include regional-fixed effects in the estimated specification. However, Combes and Gobillon (2015) highlight several reasons why this strategy may not work. The empirical model includes two regional-level variables that vary little over year, RD_t^r and $KIBS_t^r$, as a result, it is difficult to identify their effect separately form the unobserved, time invariant, characteristics affecting firm productivity across regions. Moreover, if firms do not change region over the year, firm-level local milieu unobserved characteristics can be appropriately dealt with using firm fixed effects. Industry fixed-effects control both for the technological differences across industries and for the well-known problem of comparing productivity levels of different industries.

Estimating empirical model (6), endogeneity can occur due to both unobserved heterogeneity and simultaneity bias. Some regional characteristics (e.g. public infrastructures, local climate, natural resources, etc.) can affect the propensity to agglomerate, while at the same time agglomeration influences these regional characteristics. Additionally, the likelihood of high-productivity firms being attracted to

large markets may not be as a result of agglomeration economies (selection effect). Thus, to deal with this endogeneity problem the model (6) is estimated using the system GMM estimator, as proposed by Combes and Gobillon (2015).

RESULTS

Results for all manufacturing firms

The descriptive statistics and the correlation matrix of variables, and econometric issues are in the Appendix C and D, respectively. Column (1) of Table 1 reports the results for all manufacturing firms. It shows that the linear localization coefficient is positively signed, while the coefficient of the corresponding squared term is negative (both statistically significant at the 0.01 and 0.05 levels, respectively). On average, a 1% increase in the share of employment in neighbouring firms of the same industry implies a 0.237% ($= \hat{\gamma}_1 + 2\hat{\gamma}_2 \, loc$) increase in firm TFP, all else constant—since all variables are expressed in logarithms, the estimated coefficients can be interpreted as elasticity parameters—, a result consistent with hypothesis 1A. As seen in revision literature section, comparing this result with those reported in the literature is a difficult task given the heterogeneity of methodologies used. Using a similar measure of specialization, Cingano and Schivardi (2004) also found an elasticity of 0.230. López and Südekum (2009), Martin et al. (2011) and Hashiguchi and Tanaka (2015) found positive effects of specialization on firm TFP as well, but the magnitude of elasticities cannot be compared because they use different location proxies.

[Table 1]

The relationship between productivity and localization does not seem to be linear— the greater the specialization, the lower the elasticity. In other words, the marginal effect

decreases as the specialization of region increases (Figure 2*a*). This suggests that the measured effect is the average net effect of localization economies and congestion effects. This result may also explain why studies using data from less densely populated regions find positive specialization effects and negative effects on more populated areas (de Groot et al., 2016). Martin et al. (2011) found that the (net) effect of localization economies has an inverted U-shape pattern, which confirms that concentration generates increasing congestion effects that, after a certain threshold (in our case at about the 90th percentile), can dominate localization effects. Therefore, in the case of localization economies, there is broad evidence in favour of the hypothesis 2B.

[Figure 2]

In the case of diversity economies, only the squared coefficient is significantly positive (at the 0.1 level). Finally, in the case of science-technology economies, the linear and squared R&D coefficients are both positively and negatively signed (and significant at the 0.1 and 0.05 levels), respectively, while the KIBS coefficients do not seem to have any statistically significant impact, at conventional levels, on TFP. Moreover, the null hypothesis of existence of an inverted U-shaped relationship between R&D and productivity over the range of the data is not rejected at the 0.05 significance level by the Lind-Mehlum U-test.[10] Hence, all else constant, more R&D employment in a region seems to be beneficial for improving firm productivity, but only up to a certain critical level (approximately at the median point, as can be seen in Figure 2*b*), which is again a clear confirmation of the hypothesis 2B.

Industry-specific results

In order to ascertain whether the impact of regional knowledge spillovers differs among industries, manufacturing industries are grouped in four sectors according to OECD

16

classification: low, medium-low, medium-high and high technology (see Appendix E). Columns (2) to (5) of Table 1 show that there is broad evidence in favour of the hypothesis 3 that the effects of agglomeration economies vary considerably among industries. In particular, the low- and medium-low-tech industries seem to benefit from localization economies, while industry specialization does not impact (at conventional significance levels) on the productivity of medium-high- and high-tech firms. In the low-tech industries, a 1% increase in the specialization index implies an increase of 0.140% in firm TFP, on average—the elasticity ranges from 0.199 (10th percentile) to 0.019 (90th percentile)—, all else constant; in the medium-low-tech industries, the corresponding increase is 0.386%. These results are much in line with those reported by Lall et al. (2004), but do not confirm the significantly positive effect in high-tech industries found by Henderson (2003).

Regarding diversity economies, they are statistically significant only in the case of medium-high-tech industries, where an inverted U-shaped relationship over the range of the data was observed (the Lind-Mehlum U-test is passed at the 0.1 level). More precisely, the positive effect of regional diversity is observed up to the 70th percentile, after which there are agglomeration diseconomies, as can be seen in Figure 2f.

Finally, in the case of science-technology economies, the results are quite interesting. The two variables, R&D and KIBS employment, are statistically significant at the 0.05 level in the case of the medium-low- and high-tech industries (the linear R&D coefficient of the medium-low-tech industries, which is significant at the 0.1 level, is the sole exception). In particular, a 1% increase in the R&D (KIBS) regional employment implies a 0.387% (0.031%) increase in firm productivity, on average, in the case of medium-low-tech industries; the corresponding increasing in the case of high-tech industries is 0.055% (0.090%).

Ehrl (2013) also observed a significant effect of science-technology externalities in lower technology industries. One plausible interpretation for this result is that, even though those firms invest relatively less in R&D and are less innovative than high-tech firms, they do nevertheless create new products and techniques to achieve competitive advantage. A vital part of this process is the external technological knowledge (Grimpe and Sofka, 2009; Robertson et al., 2009).

Surprisingly enough, the data seem to support an inverted U-shaped relationship between regional R&D employment and firm productivity in both industries (the Lind-Mehlum U-test is significant at the 0.05 level). However, the elasticity is always positive over the range of the data in the former case, while it becomes negative at about the 55th percentile in the latter case (Figures 2*d* and 2*g*). There are two potential explanations for these diseconomies of scale. First, as R&D agglomeration increases, congestion or density costs may eventually offset the benefits to firms. Second, high R&D agglomeration can result in a so-called free-rider problem, that is, under high knowledge spillovers firms may find that imitation is cheaper and less-risky than internal R&D (Yang et al. 2013).

In the case of KIBS employment, the negative and positive coefficients on the linear and squared terms, respectively, depicts a U-shaped relationship over the range of the data (the Lind-Mehlum U-test is significant at the 0.05 level). Thus, the medium-low- and high-tech firms do benefit from the presence of KIBS activities, but only when the agglomeration of KIBS employment is higher than the median, the sole confirmation of the hypothesis 2A (Figures 2*e* and 2*h*). This impact is higher in the latter case: in the third quartile, the elasticities are 0.367 and 0.093 in the high- and medium-low-tech industries, respectively.

CONCLUSION

This article contributes to the ongoing debate on whether regional environment is beneficial to firm productivity by adding several significant new pieces to the puzzle. In particular, based on a panel of Portuguese manufacturing firms over the period 1996–2004, it sheds further light on whether regional knowledge spillovers have different, nonlinear, effects among industries, this being a critical issue to understanding the impact of regional disparities on economic growth and therefore to promoting and designing new, more assertive, regional policies.

The results are in line with other past studies and support the idea that agglomeration effects vary and depend on the type of externality and on the type of industry. They confirm the hypothesis that local knowledge spillovers play an important role in explaining firm productivity. Moreover, they occur in different combinations for firms belonging to different industries. Indeed, whereas in the case of estimates for all manufacturing sectors pooled together localization economies have a significantly positive effect, in the case of distinct regressions by sector only high-low- and medium-low-tech firms seem to benefit from this source of economies. Diversity externalities, for their part, have an inverted U-shaped relationship with firms TFP in medium-high-tech industries— agglomeration diseconomies are dominant after the 70th percentile. It is also found an inverted U-shaped relationship between regional R&D employment and firm productivity for all manufacturing firms and firms from medium-low- and high-tech industries—the elasticities become negative for values higher than the median. Finally, medium-low- and high-tech firms benefit from KIBS activities when their employment in the region is higher than the median.

Overall, this study contributes to a better understanding of regional knowledge spillovers at the industry level, which is critical to promoting and designing new, more assertive, regional policies to enhance economic growth. In particular, their findings

suggest that differences in the effects between regions imply that the respective policy recommendations may only hold for certain types of industry. Given the heterogeneity of industries among regions, we should be careful about deriving general policy recommendations. The results suggest that policies should be adjusted to the specific characteristics of each region. Moreover, regional policies should take into account potential critical levels of agglomeration of each source of knowledge. For example, in regions where low-tech economic activities are dominant, policies should target new investors of similar industries. On the contrary, policies to promote the diversification tend to be better in regions where medium-high-tech firms predominating. But there is a limit in this policy. Over-diversification can negatively impact on firm productivity.

It is worth noting that this analysis has two main limitations. Firstly, multi-plant firms may affect results if those plants are located in different regions. Secondly, the study does not control for unobserved spatial correlation.

A number of interesting further extensions to this article could shed further light on regional knowledge spillovers. These include the replication of the study to other economies to identify the optimal regional density of the different sources of knowledge. Furthermore, future research could adopt both efficient instrumental variable quantile regression techniques to control for nonlinear effects with endogeneity and spatial econometric techniques to control for unobserved spatial correlation, two challenging tasks. They could also look more deeply into the role of firm human capital and R&D investment on the absorption of regional knowledge spillovers.

SUPPLEMENTAL MATERIAL

REFERENCES

Acs, Z., Audretsch, J.D.B., & Lehmann, E.E. (2013). The knowledge spillover theory of entrepreneurship. *Small Business Economics*, 41, 757–774. doi:10.1007/s11187-013-9505-9

Anderson, M., & Lööf, H. (2011). Agglomeration and productivity: Evidence from firm-level data. *Annals of Regional Science*, 46, 601–620. doi:10.1007/s00168-009-0352-1

Arellano, M., & Bover, O. (1995). Another look at the instrumental variable estimation of error-components models. *Journal of Econometrics*, 68, 29–51. doi:10.1016/0304-4076(94)01642-D

Arrow, K.J. (1962). The economic implications of learning by doing. *Review of Economic Studies*, 29, 155–173.

Au, C.-C., & Henderson, J.V. (2006). Are Chinese Cities Too Small? *Review of Economic Studies* 73, 549–756. doi:10.1111/j.1467-937X.2006.00387.x

Audretsch, D.B., & Feldman, M.P. (2004). Knowledge spillovers and the geography of innovation. In: Henderson, J.V., Thisse, J.F. (Eds.) *Handbook of Urban and Regional Economics*, Vol. 4, pp. 2713–2739. Amsterdam: Elsevier. doi:10.1016/S1574-0080(04)80018-X

Audretsch, D.B., & Fritsch, M. (2002). Growth Regimes over Time and Space, *Regional Studies* 36, 113–124. doi:10.1080/00343400220121909

Audretsch, D.B., Lehmann, E., & Warning, S. (2005). University spillovers and new firm location. *Research Policy* 34, 1113–1122. doi:10.1016/j.respol.2005.05.009

Baptista, R. (1999) The Diffusion of Process Innovations: A Selective Review, *International Journal of the Economics of Business* 6, 107-129. doi:10.1080/13571519984359

Baptista, R. (2000). Do innovations diffuse faster within geographical clusters? *International Journal of Industrial Organization* 18, 515–535. doi:10.1016/S0167-7187(99)00045-4

Baptista, R., & Mendonça, J. (2010). Proximity to knowledge sources and the location of knowledge-based start-ups. *Annals of Regional Science* 45, 5–29. doi:10.1007/s00168-009-0289-4

Baptista, R., & Swann, P. (1998). Do firms in clusters innovate more? *Research Policy* 27, 525–540. doi:10.1016/S0048-7333(98)00065-1

Baptista, R., Lima, F., & Mendonça, J. (2011). Establishment of higher education institutions and new firm entry. Research Policy 40, 751–760. doi:10.1016/j.respol.2011.02.006

Blundell, R., & Bond, S. (1998). Initial conditions and moment restrictions in dynamic panel data models. *Journal of Econometrics*, 87, 115–143. doi:10.1016/S0304-4076(98)00009-8

Blundell, R., & Bond, S. (2000). GMM Estimation with persistent panel data: an application to production functions. *Econometric Reviews*, 19, 321–340. doi:10.1080/07474930008800475

Breschi, S., & Lissoni, F. (2001). Knowledge Spillovers and Local Innovation Systems: A Critical Survey. *Industrial and Corporate Change*, 10, 975–1005. doi:10.1093/icc/10.4.975

Cainelli, G., & Leoncini, R. (1999). Externalities and Long-Term Local Industrial Development: Some Empirical Evidence from Italy. *Revue d'economie industrielle*, 90, 25–40. doi:10.3406/rei.1999.1762

Cainelli, G., Fracasso, A., & Marzetti, G.V. (2015). Spatial agglomeration and productivity in Italy: A panel smooth transition regression approach. *Papers in Regional Science*, 94(S1), S39–S68. doi:10.1111/pirs.12103

Cantwell J., & Piscitello, L. (2005). Recent Location of Foreign-owned Research and Development Activities by Large Multinational Corporations in the European Regions: The Role of Spillovers and Externalities. *Regional Studies* 39, 1–16. doi:10.1080/0034340052000320824

Carreira, C., & Lopes, L.P. (2015). Are Small Firms More Dependent on the Local Environment than Larger Firms? Evidence from Portuguese Manufacturing Firms In Leitao J.C., & Baptista, R. (Eds.) *Entrepreneurship, Human Capital, and Regional Development*, pp 263–280. Springer. doi:10.1007/978-3-319-12871-9_13

Carreira, C., & Teixeira, P. (2011a). Entry and exit as a source of aggregate productivity growth in two alternative technological regimes. *Structural Change and Economic Dynamics*, 22(2), 135–150. doi:10.1016/j.strueco.2011.01

Carreira, C., & Teixeira, P. (2011a). The shadow of death: Analysing the pre-exit productivity of Portuguese manufacturing firms. Small Business Economics, 36(3), 337–351. doi:10.1007/s11187-009-9221-7

Cassia, L., Colombelli, A., & Paleari, S. (2009). Firms' growth: Does the innovations system matter? *Structural Change and Economic Dynamics*, 20, 211–220. doi:10.1016/j.strueco.2009.01.001

Cingano, F., & Schivardi, F. (2004). Identifying the sources of local productivity growth. *Journal of the European Economic Association*, 2, 720–742. doi: 10.1162/1542476041423322

Combes, P.P. (2000). Economic Structure and Local Growth: France, 1984–1993. *Journal of Urban Economics*, 47, 329–355. doi:10.1006/juec.1999.2143

Combes, P.-P., & Gobillon, L. (2015). The Empirics of Agglomeration Economies. In: Duranton, G., Henderson, J.V., Strange, W.C. (Eds.) *Handbook of Urban and Regional Economics*, Vol. 5, pp. 247-348. Amsterdam: Elsevier.

Döring, T., & Schnellenbach, J. (2006). What do we know about geographical knowledge spillovers and economic growth? A survey of the literature. *Regional Studies*, 40, 375–395.

Duranton, G., & Puga, D. (2004). Micro-Foundations of Urban Agglomeration Economies. In: Henderson, J.V., Thisse, J.F. (Eds.) *Handbook of Urban and Regional Economics*, Vol. 4, pp. 2063–2117. Amsterdam: Elsevier. doi:10.1016/S1574-0080(04)80005-1

Ehrl, P. (2013). Agglomeration economies with consistent productivity estimates. *Regional Science and Urban Economics*, 43, 751–763. doi:10.1016/j.regsciurbeco.2013.06.002

Forni, M., & Paba, S. (2002). Spillover and the growth of local industries. *Journal of Industrial Economics*, 50, 151–171. doi: 10.1111/1467-6451.00172

Frenken, K., van Oort, F., & Verburg T. (2007). Related variety, unrelated variety and regional economic growth. *Regional Studies*, 41, 685–697. doi:10.1080/00343400601120296

Fritsch, M., & Aamoucke, R. (2013). Regional public research, higher education, and innovative start-ups: an empirical investigation. *Small Business Economics* 41, 865–885. doi:10.1007/s11187-013-9510-z

Fritsch, M., & Falck, O. (2007). New Business Formation by Industry over Space and Time: A Multidimensional Analysis, *Regional Studies* 41, 157–172. doi:10.1080/00343400600928301

Fritsch, M., & Kauffeld-Monz, M. (2010). The impact of network structure on knowledge transfer: an application of social network analysis in the context of regional innovation networks. *Annals of Regional Science* 44, 21–38. doi:10.1007/s00168-008-0245-8

Fritsch, M., & Slavtchev, V. (2007). Universities and innovation in space. *Industry Innovation*, 14, 201–218. doi:10.1080/13662710701253466

Fritsch, M., & Slavtchev, V. (2010). How does industry specialization affect the efficiency of regional innovation systems? *Annals of Regional Science* 45, 87–108. doi:10.1007/s00168-009-0292-9

di Giacinto, V., Gomellini, M., Micucci, G., & Pagnini, M. (2014). Mapping local productivity advantages in Italy: Industrial districts, cities or both?. *Journal of Economic Geography*, 14, 365–394. doi:10.1093/jeg/lbt021

Glaeser, E., Kallal, H., Sheikman, J., & Shleifer, A. (1992). Growth in cities. *Journal of Political Economy*, 100, 1126–1152. doi:10.1086/261856

Grimpe, C., & Sofka, W. (2009). Search patterns and absorptive capacity: Low- and high-technology sectors in European countries. *Research Policy*, 8, 495–506. doi:10.1016/j.respol.2008.10.006

de Groot, H.L.F., Poot, J., & Smit, M.J. (2016). Which agglomeration externalities matter most and why? *Journal of Economic Surveys* 30, 756–782. doi:10.1111/joes.12112

Hashiguchi, Y., & Tanaka, K. (2015). Agglomeration and firm-level productivity: A Bayesian spatial approach. *Papers in Regional Science*, 94(S1), S95–S114. doi:10.1111/pirs.12147

Henderson, J.V. (2003). Marshall's scale economies. *Journal of Urban Economics*, 53, 1–28. doi:10.1016/S0094-1190(02)00505-3

Henderson, V., Lee, T., & Lee, Y.J. (2001). Scale Externalities in Korea. *Journal of Urban Economics* 49, 479–504. doi:10.1006/juec.2000.2202

Jacobs, J. (1969). *The economy of cities*, New York: Vintage.

Kauffeld-Monz, M., & Fritsch, M. (2013). Who Are the Knowledge Brokers in Regional Systems of Innovation? A Multi-Actor Network Analysis. *Regional Studies* 47, 669–685. doi: 10.1080/00343401003713365

Klevorick, A., Levin, R., Nelson, R., & Winter, S. (1995). On the Sources and Interindustry Differences in Technological Opportunities. *Research Policy*, 24, 185–205. doi:10.1016/0048-7333(93)00762-I

Lall, S.V., Shalizi, Z., & Deichmann, U. (2004). Agglomeration economies and productivity in Indian industry. *Journal of Development Economics* 73, 643–673. doi:10.1016/j.jdeveco.2003.04.006

Lambooy, J.G. (2010). Knowledge Transfers, Spillovers and Actors: The Role of Context and Social Capital. *European Planning Studies*, 18, 873–891. doi:10.1080/09654311003701407

Lind, J.T., & Mehlum, H. (2010). With or without U? The appropriate test for a U shaped relationship. *Oxford Bulletin of Economics and Statistics*, 72, 109-118. doi:10.1111/j.1468-0084.2009.00569.x

López, R.A., & Südekum, J. (2009). Vertical Industry Relations, Spillovers, and Productivity: Evidence from Chilean Plants. *Journal of Regional Science*, 49, 721–747. doi:10.1111/j.1467-9787.2009.00631.x

de Lucio, J.J., Herce, J.A., & Goicolea, A. (2002). The effects of externalities on productivity growth in Spanish industry. *Regional Science and Urban Economics* 32, 241–258. doi:10.1016/S0166-0462(01)00081-3

Malerba F., & Orsenigo L. (1997). Technological regimes and sectoral patterns of innovative activities. *Industrial and Corporate Change*, 6, 83–117. doi:10.1093/icc/6.1.83

Marshall, A. (1890). *Principles of Economics*, (9.ª ed. 1961). London: Macmillan Press.

Marsili, O. (2002). Technological Regimes and Sources of Entrepreneurship. *Small Business Economics* 19, 217–231. doi:10.1023/A:1019670009693

Martin, P., Mayer, T., & Mayneris, F. (2011). Spatial concentration and plant-level productivity in France. *Journal of Urban Economics*, 69, 182–195. doi:10.1016/j.jue.2010.09.002

Melo, P.C., Graham, D.J., & Noland. R.B. (2009). A meta-analysis of estimates of urban agglomeration economies. *Regional Science and Urban Economics*, 39, 332–342. doi:10.1016/j.regsciurbeco.2008.12.002

Muller, E., & Zenker, A. (2001). Business services as actors of knowledge transformation: The role of KIBS in regional and national innovation systems. *Research Policy*, 30, 1501–1516. doi:10.1016/S0048-7333(01)00164-0

Nelson, R.R., & Winter, S.G. (1982). *An Evolutionary Theory of Economic Change*, Cambridge, MA: Belknap Press of Harvard University Press.

OECD (2005). STAN indicators. http://www.oecd.org/dataoecd/3/33/40230754.pdf

van Oort, F.G., Burger, M.J., Knoben, J., & Raspe, O. (2012). Multilevel approaches and the firm-agglomeration ambiguity in economic growth studies. *Journal of Economic Surveys*, 26, 468–491. doi:10.1111/j.1467-6419.2012.00723.x

Puga, D. (2010). The magnitude and causes of agglomeration economies. *Journal of Regional Science*, 50, 203–219. doi:10.1111/j.1467-9787.2009.00657.x

Robertson, P., Smith, K., & von Tunzelmann, N. (2009). Introduction: Innovation in low- and medium-technology industries. *Research Policy*, 38, 441–446. doi:10.1016/j.respol.2008.10.019

Romer, P.M. (1986). Increasing returns and long-run growth. *Journal of Political Economy*, 94, 1002–1037. doi:10.1086/261420

Roodman, D. (2009). How to do xtabond2: An introduction to difference and system GMM in Stata. *Stata Journal*, 9, 86-136.

Rosenthal, S.S., & Strange, W.C. (2004). Evidence on the nature and sources of agglomeration economies. In Henderson, J.V., Thisse, J.F. (Eds), *Handbook of Urban and Regional Economics*, Vol. 4, pp. 2119–2172. Amsterdam: Elsevier.

Saito, H., & Gopinath, M. (2009). Plants' self-selection, agglomeration economies and regional productivity in Chile. *Journal of Economic Geography*, 9, 539–558. doi:10.1093/jeg/lbp010

Storper, M., & Venables, A.J. (2004). Buzz: Face-to-face contact and the urban economy. *Journal of Economic Geography*, 4, 351–370. doi:10.1093/jnlecg/lbh027

Tappeiner, G., Hauser, C., & Walde, J. (2008). Regional knowledge spillovers: Fact or artifact?. *Research Policy*, 37, 861–874. doi:10.1016/j.respol.2007.07.013

Yang, C.-H, Lin, H.-L., & Li, H.-Y. (2013). Influences of production and R&D agglomeration on productivity: Evidence from Chinese electronics firms. *China Economic Review*, 27, 162–178. doi:10.1016/j.chieco.2013.09.005

Notes

---

[1] As pointed out by Frenken et al. (2007, p. 686), it is important to distinguish diversity economies from urbanization economies. Diversity economies, i.e. variety of industries, are sometimes regarded as part of urbanization economies, i.e. variety of the local actors and infrastructure, as in Cingano and Schivardi (2004) and Martin et al. (2011), for example.

[2] Other authors identify a fourth source, the competition economies. However, in highly integrated markets like those in European countries, the effect of local competition does

not appear to be statistically significant. This is not so surprising since its marginal impact shrinks under high levels of competition (e.g. Martin et al., 2011; Carreira and Lopes, 2015; Cainelli et al., 2015).

[3] It is assumed that each firm operates in a given industry j and is located in a given region r. Subscripts j and r are omitted to simplify the notation except when it causes ambiguity.

[4] Since it is subtracted ith firm employment, LOC variable is firm specific.

[5] Saito and Gopinath (2009), Martin et al. (2011) and Combes and Gobillon (2015) use the log of the inverse Herfindahl-Hirschman index of regional employment shares of the industries different from j. However, the entropy index is a more standard measure of variety (Frenken et al. 2007). We have nonetheless estimated the model with the inverse Herfindahl-Hirschman index. The results do not significantly change (available from authors upon request).

[6] The knowledge produced by universities and other knowledge-generating organizations is also often measured by, *inter al.*, the amount of money spent on R&D, the number of employees engaged in R&D activities, the number of students in higher degree establishments, the number of articles published in scientific and academic journals or the number of patents (Audretsch et al. 2005; Fritsch and Slavtchev, 2007; Cassia et al., 2009; Baptista and Mendonça, 2010; Baptista et al., 2011; Fritsch and Aamoucke, 2013). However, most of these variables are highly correlated due to complementarity. Having a large number of students, for example, means a larger teaching staff and then a greater amount of R&D resources (Fritsch and Slavtchev, 2007). On the other hand, the number of R&D workers (or the R&D expenditures) is a more general and representative measure. We have nonetheless estimated the model with the number of students in higher degree establishments and the number of higher degree establishments, but the results were less significant (these results are available from authors upon request).
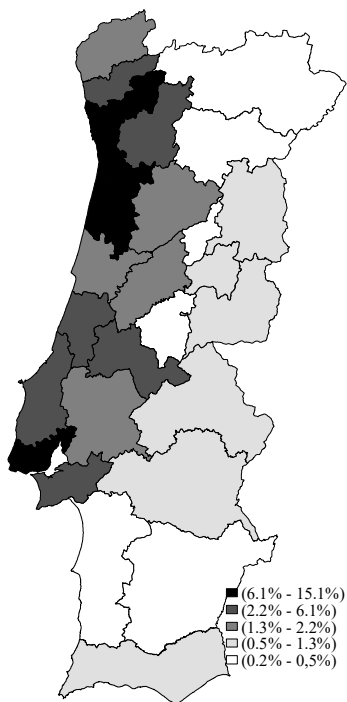
Figure 1   Number of firms by NUTS 3 regions (percentage of total)

Figure 2 Elasticity of (significant) dependent variables by percentile
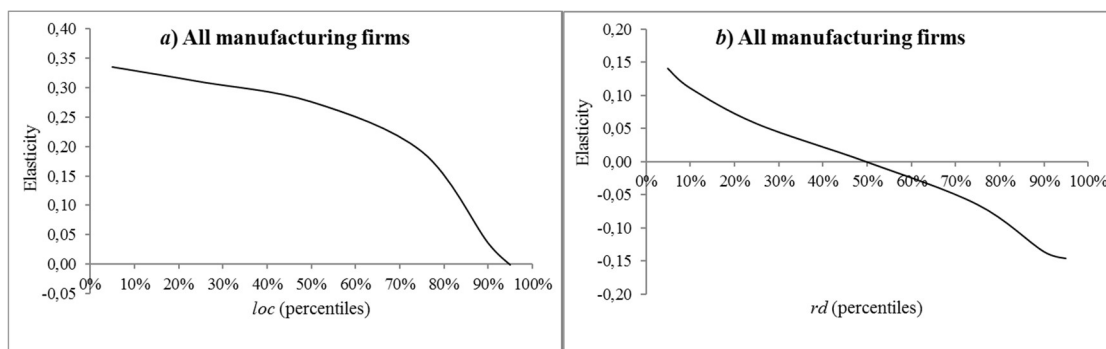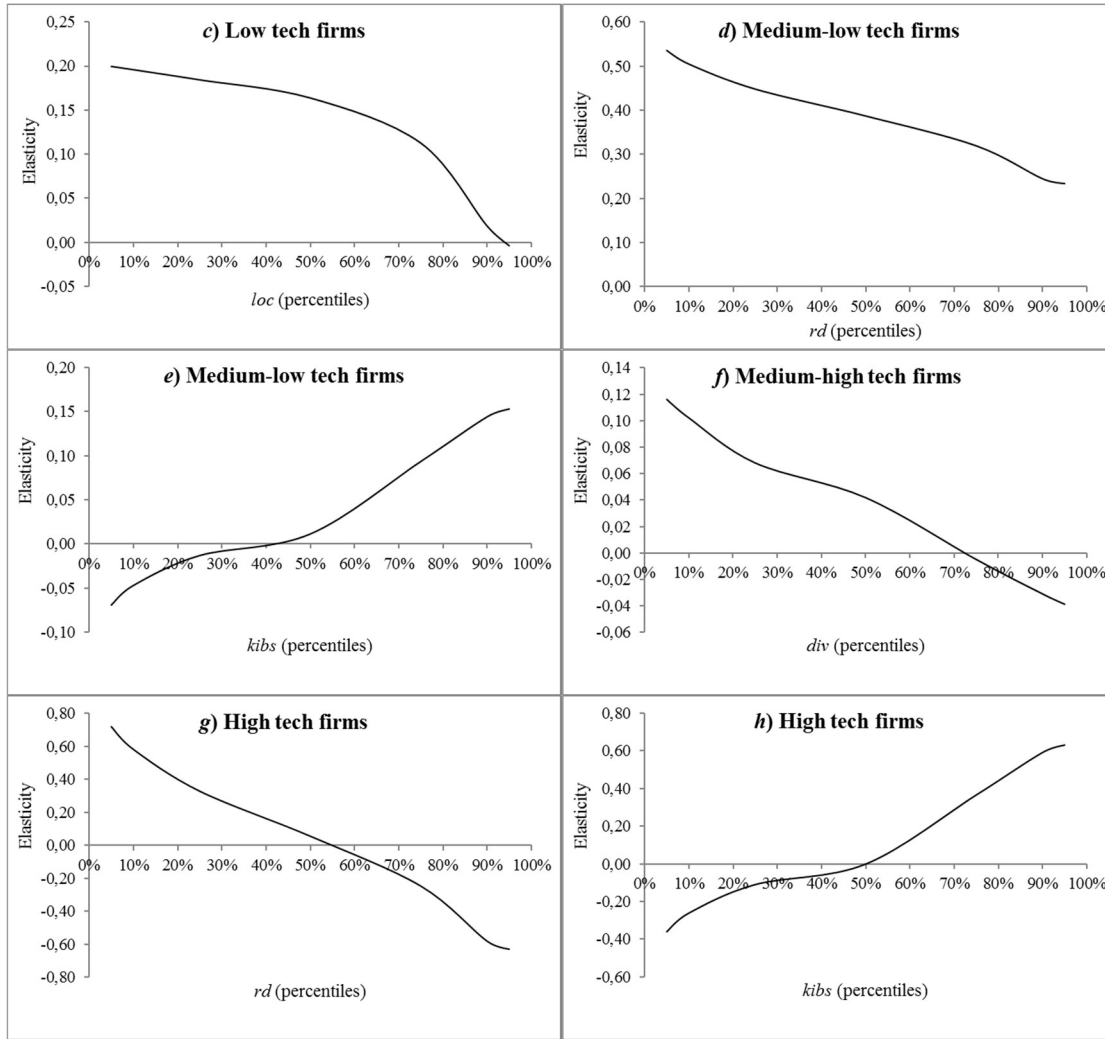
Table 1. Regression results by technology level

Variable	Overall (1)	Low (2)	Medium-low (3)	Medium-high (4)	High (5)
loc_{it}^{jr}	0.348***	0.207***	0.386*	-0.265	0.439
	(0.126)	(0.079)	(0.241)	(0.258)	(0.530)
$\left(loc_{it}^{jr}\right)^2$	-0.332**	-0.200**	-1.315	0.860	-0.265
	(0.167)	(0.093)	(0.834)	(1.251)	(0.267)
div_t^{jr}	-0.228	0.050	-1.120	0.239*	-0.067
	(0.142)	(0.908)	(0.897)	(0.131)	(4.533)
	0.069*	-0.004	0.314	-0.062*	0.112

$\left(div_t^{jr}\right)^2$	(0.039)	(0.227)	(0.248)	(0.035)	(1.297)
rd_t^r	0.324*	0.089	0.285*	-0.185	1.582**
	(0.178)	(0.102)	(0.159)	(0.173)	(0.739)
$\left(rd_t^r\right)^2$	-0.025**	-0.012	-0.026**	0.017	-0.128**
	(0.012)	(0.010)	(0.013)	(0.014)	(0.055)
$kibs_t^r$	-0.106	0.087	-0.186**	0.062	-0.875**
	(0.080)	(0.081)	(0.082)	(0.076)	(0.407)
$\left(kibs_t^r\right)^2$	0.011	-0.001	0.015**	-0.005	0.065**
	(0.007)	(0.006)	(0.006)	(0.005)	(0.028)
age_{it}	-0.050	-0.037	0.002	-0.027	0.037
	(0.045)	(0.044)	(0.035)	(0.041)	(0.190)
$\left(age_{it}\right)^2$	0.007	0.003	-0.002	0.001	-0.010
	(0.008)	(0.007)	(0.006)	(0.006)	(0.034)
$a_{i(t-1)}$	0.499***	0.422***	0.648***	0.561***	0.057
	(0.179)	(0.153)	(0.126)	(0.076)	(0.353)
$a_{i(t-2)}$	0.114	0.018	0.231*	0.130	-0.541
	(0.150)	(0.143)	(0.120)	(0.083)	(0.384)
AR(1) [*p-value*]	-2.52 [*0.012*]	-5.61 [*0.000*]	-3.55 [*0.000*]	-3.49 [*0.000*]	-2.17 [*0.030*]
AR(2) [*p-value*]	0.22 [*0.828*]	1.11 [*0.265*]	-0.07 [*0.946*]	0.66 [*0.511*]	1.29 [*0.198*]
Sargan test [*p-value*]	40.92 [*0.000*]	79.07 [*0.000*]	24.55 [*0.176*]	22.72 [*0.190*]	27.64 [*0.010*]
Hansen test [*p-value*]	9.64 [*0.381*]	9.50 [*0.111*]	24.05 [*0.194*]	18.89 [*0.219*]	9.04 [*0.339*]
N. instruments	42	42	42	39	31
Variance prop.	0.925	0.909	0.931	0.923	0.875
KMO test	0.858	0.848	0.874	0.854	0.844
Observations	14,882	8,335	3,149	2,620	568

Notes: Blundell and Bond (1998) two-step system GMM estimates of model (6). Constant, industry dummies and year dummies included in all models. "Variance Prop." denotes the proportion of the variance of GMM instruments explained by extracted principal components. KMO denotes the Kaiser–Meyer–Olkin test, which is a measure of sampling adequacy of GMM instruments. Sargan test is not robust enough to detect heteroskedasticity and autocorrelation (Roodman, 2009). (1) All manufacturing firms, (2) low tech firms, (3) medium-low tech firms, (4) medium-high tech firms and (5) high tech firms. The number of instruments may vary due to the number of industries considered in each column. Robust standard errors are given in parentheses. ***, **, and * denote statistical significance at the 0.01, 0.05, and 0.10 levels, respectively.
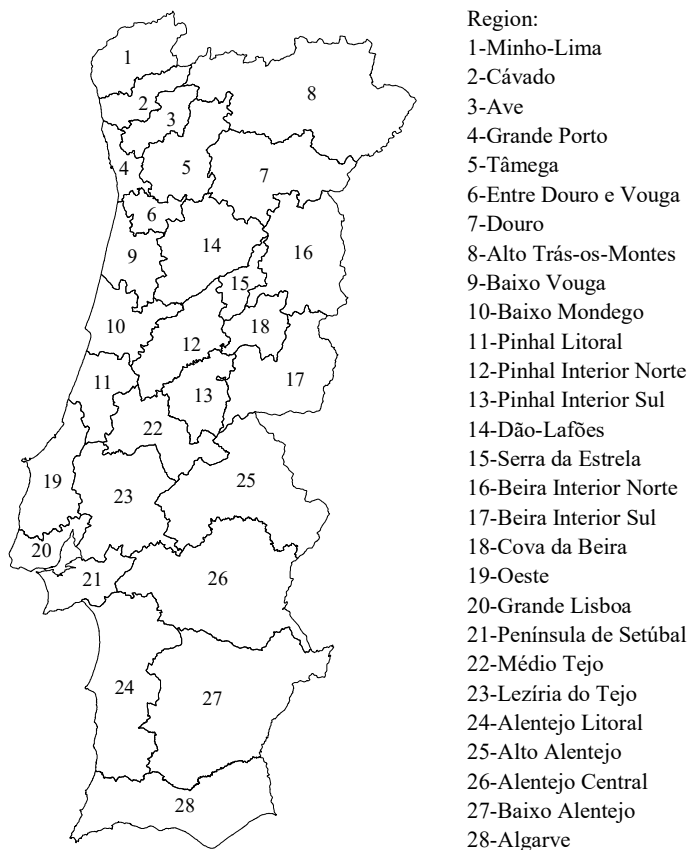
APPENDIX

Appendix A. The data

The empirical study is based on an unbalanced panel of Portuguese manufacturing firms covering the period 1996–2004. The raw data is drawn from the combination of two statistical data sources, both administrated by the Portuguese Statistical Office (INE): *Inquérito às Empresas Harmonizado* (IEH), which is an annual business survey with detailed input and output information required for the computation of firm-level productivity; and *Ficheiro de Unidades Estatísticas* (FUE), a file that contains information on firm location, number of employees and the main economic activity of all Portuguese firms, which is critical to computing spatial agglomeration variables. The longitudinal dimension of the panel, required for our analysis, was constructed using a (unique) firm identification code.

The unit of production considered is therefore the firm. Each firm was assigned to a given NUTS 3-region (Nomenclature of Territorial Units for Statistics at level 3 of the European regional classification, Portuguese definition for 2002) through a spatial identification code (see Figure A1). Thus, the first drawback of our dataset is that multi-plant firms may affect our results if their different plants are located in different regions. However, we note that the different plants of a corporation are often registered as distinct legal entities, so the multi-plant phenomenon impact on results may be negligible.

Figure A1 NUTS 3 regions of mainland Portugal

Region:
1-Minho-Lima
2-Cávado
3-Ave
4-Grande Porto
5-Tâmega
6-Entre Douro e Vouga
7-Douro
8-Alto Trás-os-Montes
9-Baixo Vouga
10-Baixo Mondego
11-Pinhal Litoral
12-Pinhal Interior Norte
13-Pinhal Interior Sul
14-Dão-Lafões
15-Serra da Estrela
16-Beira Interior Norte
17-Beira Interior Sul
18-Cova da Beira
19-Oeste
20-Grande Lisboa
21-Península de Setúbal
22-Médio Tejo
23-Lezíria do Tejo
24-Alentejo Litoral
25-Alto Alentejo
26-Alentejo Central
27-Baixo Alentejo
28-Algarve

The IEH survey comprises all firms operating in Portugal with more than 100 employees, plus a representative random sample of firms with 20 to 99 employees.[10] For reasons of confidentiality restrictions, the raw data from industries with a small number of firms (i.e. *tobacco products*, and *coke and refined petroleum products*) are not available. We also excluded from the dataset firms located in the island regions (i.e. Madeira and Azores). Firms with missing observations or unreasonable values (negative values or outliers) were also ignored.[10] After these procedures, our estimation sample comprises an unbalanced panel of 8,074 firms or 32,003 year-firm observations. Table A1 lists the industries covered by the study and the corresponding number of firms.

Table A1 Number of firms by industry

CAE	Industry	Mean	SD	Min	Max

15	Food products and beverages	459	18.5	429	482
17/18	Textiles and wearing apparel	817	109.8	659	971
19	Leather and leather products	215	51.8	162	339
20	Wood and wood products (except furniture)	192	11.8	168	207
21/22	Pulp, paper, paper products and publishing	207	10.7	191	226
24	Chemical and chemical products	155	3.9	148	161
25	Rubber and plastic products	87	12.3	69	105
26	Other non-metallic mineral products	316	19.7	278	333
27/28	Basic metals and fabricated metal products	359	18.9	327	387
29	Machinery and equipment	279	19.8	244	310
30/31/32/33	Electronic and electrical equipment	159	4.5	154	167
34/35	Motor vehicles, trailers and other transport equipment	133	8.8	120	147
36	Furniture, other manufacturing n.e.c. and recycling	178	16.7	146	203

Notes: The decomposition uses the two-digit level of the Portuguese Classification of Economic Activities (CAE Rev. 2.1). At least at this disaggregation level there is a direct correspondence between this classification and the classifications of both the European Community (NACE-Rev. 1.1) and the United Nations (ISIC Rev. 3.1). Mean values over the period 1996–2004 and standard deviations (SD) of the number of firms.

Appendix B. Production function results

We use the two-step system GMM approach to estimate the (log-linear) Cobb-Douglas production function at the industry (two-digit) level and compute the firm-level TFP.[10] Given that output elasticities may vary significantly across industries due to differences in production technology, we run distinct regressions for each industry. To choose the level of industry disaggregation, we strike a balance between ensuring the homogeneity of the technology within each industry and preserving a significant number of observations by industry.

The *gross output* (Y_{it}) is given by the sum of total revenues from sales, services rendered and production subsidies. It is deflated by the producer price index at the 3-digit level. The *labour* (L_{it}) input is a 12-month employment average. *Materials* (M_{it}) include the cost of materials and services purchased and were deflated by the GDP deflator. *Capital*

stock (K_{it}) is measured as the book value of total net assets excluding financial investments and cash stock (that is, it includes not only tangible and intangible assets but also other elements of the asset side of the balance sheet, such as accounts receivable and inventory, all important to the operation of the firm). Output and input monetary variables are measured in constant 1996 Euro prices.

Table B1 presents the output elasticity parameters for each factor, not imposing constant returns to scale. All regressions include year dummies, exit dummy and firm age variable. Test statistics for auto-correlation of the error term indicate significant negative first-order serial correlation at standard significance levels for all regressions, but no significant second-order serial correlation, as expected. The validity of lagged instruments dated *t*–2 (and earlier) is also clearly not rejected by Hansen tests of overidentifying restrictions in all estimates (the Sargan test is less clear for some regressions, but generally is not robust; Roodman, 2009a).

Table B1. Production function elasticities by industry

Industry	α_K	β_L	θ_M	*COMFAC*
Food products and beverages	0.088**	0.185***	0.722***	0.829
	(0.041)	(0.060)	(0.034)	
Textiles and wearing apparel	0.088**	0.222***	0.570***	0.981
	(0.045)	(0.051)	(0.039)	
Leather and leather products	0.184***	0.197***	0.613***	0.591
	(0.059)	(0.060)	(0.041)	
Wood and wood products (except furniture)	0.083*	0.185***	0.713***	0.527
	(0.045)	(0.049)	(0.034)	
Pulp, paper, paper products and publishing	0.104***	0.224***	0.667***	0.120
	(0.037)	(0.056)	(0.046)	
Chemical and chemical products	0.068***	0.179***	0.774***	0.158
	(0.034)	(0.042)	(0.036)	
Rubber and plastic products	0.023	0.109*	0.708***	0.306
	(0.042)	(0.060)	(0.045)	
Other non-metallic mineral products	0.107*	0.151***	0.760***	0.998
	(0.060)	(0.056)	(0.038)	

	α_K	β_L	θ_M	COMFAC
Basic metals and fabricated metal products	0.146***	0.248***	0.617***	0.081
	(0.048)	(0.046)	(0.043)	
Machinery and equipment	0.262***	0.229***	0.511***	0.248
	(0.062)	(0.077)	(0.052)	
Electrical and optical equipment	0.097**	0.127**	0.743***	0.189
	(0.045)	(0.056)	(0.039)	
Motor vehicles and other transport equipment	0.117***	0.220***	0.646***	0.740
	(0.039)	(0.074)	(0.038)	
Furniture, manufacturing n.e.c. and recycling	0.091**	0.191***	0.669***	0.082
	(0.043)	(0.048)	(0.047)	

Notes: Blundell and Bond (1998) two-step system GMM estimates of model (1). In order to save space, only the three output elasticities are reported (full results available from authors upon request). COMFAC is a minimum distance test of the non-linear common factor restrictions imposed in the restricted model. P-values are reported. Results for the restricted model are reported since the COMFAC test is easily passed in all regressions at standard significance levels. Year dummies, exit dummy and age variable included. α_K, β_L and θ_M denote capital, labour and material elasticities, respectively. Robust standard errors are given in parentheses. ***, **, and * denote statistical significance at the .01, .05, and .10 levels, respectively.

Given that the test of common factor restrictions (COMFAC) is easily accepted in all regressions, we focus on these results henceforth. Estimates of output elasticities range from 0.023 to 0.262 for capital, from 0.109 to 0.248 for labour and from 0.511 to 0.774 for material. They are significant at standard levels, except in the case of capital coefficient for the *rubber and plastic products* industry that it does not seem to be statistically significant. Constant returns to scale seem to be the prevalent regime in our estimates. They are rejected in favour of decreasing returns for only three industries: *textiles and wearing apparel*, *rubber and plastic products* and *manufacturing n.e.c.*.

Appendix C. Descriptive statistics of variables

Tables C1 and C2 show, respectively, the descriptive statistics and the correlation matrix of variables used in our empirical model (6). As can be seen, the standard deviation is about one-third of the respective mean for all variables except that for localization economies, which is two-thirds. Even if the between standard-deviation account for a large part this

variability, within variation is not negligible, accounting for about 13 to 31% of the variability (Table C3). The minimum value for the localization economies variable is 0 ($= e^{0} - 1$), that is, some firms are the sole representative of their industry in their region. The heterogeneity in both R&D and KIBS employment across regions is very high: R&D employment ranges from 0 (no R&D employment) to 11,990 ($= e^{9.392} - 1$); the minimum KIBS employment is 24 ($= e^{3.178}$) and the maximum is 143,322 ($= e^{11.873}$). Finally, it should be further noted that the correlation between proxies of knowledge based externalities (i.e. RD_t^r and $KIBS_t^r$) is significantly high.

Table C1. Descriptive statistics (pooled yearly values)

Variable	Obs.	Mean	Std. Dev.	Skewness	Kurtosis	Min.	Max.
a_{it}	32,003	3.001	0.900	0.684	2.270	1.139	5.472
loc_{it}^{jr}	32,003	0.167	0.1548	1.299	3.608	0	0.622
div_t^{jr}	32,003	2.047	0.294	1.686	5.437	0.490	2.391
rd_t^r	32,003	6.448	1.751	-0.318	3.359	0	9.392
$kibs_t^r$	32,003	7.447	2.302	0.413	2.204	3.178	11.873
age_{it}	31,960	2.923	0.749	-0.541	3.660	0	5.069

Table C2. Correlation across variables (pooled yearly values)

	a_{it}	loc_{it}^{jr}	div_t^{jr}	rd_t^r	$kibs_t^r$
loc_{it}^{jr}	0.592*	1			
div_t^{jr}	0.165*	0.211*	1		
rd_t^r	-0.080*	-0.179*	0.482*	1	
$kibs_t^r$	-0.137*	-0.260*	0.449*	0.851*	1
age_{it}	-0.068*	-0.052*	0.094*	0.154*	0.143*

Note: * denotes statistical significance at the .05 level.

Table C3 Descriptive statistics (between and within variations)

Variable	Mean	Std. Dev.	Min	Max	Observations

a_{it}	overall	3.001	0.900	1.139	5.472	N = 32,003
	between		0.895	1.139	5.323	n = 8074
	within		0.120	0.942	5.083	T-bar = 3.964
loc_{it}^{jr}	overall	0.167	0.155	0	0.622	N = 32,003
	between		0.158	0	0.622	n = 8074
	within		0.017	-0.192	0.625	T-bar = 3.964
div_{t}^{jr}	overall	2.047	0.294	0.490	2.391	N = 32,003
	between		0.298	0.607	2.391	n = 8074
	within		0.048	1.196	2.814	T-bar = 3.964
rd_{t}^{r}	overall	6.448	1.751	0	9.392	N = 32,003
	between		1.738	0	9.392	n = 8074
	within		0.281	1.463	10.970	T-bar = 3.964
$kibs_{t}^{r}$	overall	7.447	2.302	3.178	11.873	N = 32,003
	between		2.213	3.178	11.873	n = 8074
	within		0.518	-0.004	14.415	T-bar = 3.964

Appendix D. Econometric issues

The estimates of model (6) are obtained by applying the two-step system GMM approach. All regressions include both year and industry dummies. We estimated the system GMM by the two-step estimator because this estimator is robust enough to detect heteroskedasticity (Roodman, 2009a). The system GMM easily generates numerous instruments that can be mathematically redundant—i.e. containing no new identifying information (Arellano and Bover, 1995; Roodman, 2009a). To overcome this well-known problem of too many instruments in dynamic panel data GMM, we apply principal components analysis (PCA) to produce a smaller instrument set that is maximally representative of the original (Roodman, 2009b).[10] The proportion of the variance of GMM instruments explained by extracted principal components varies between 87.5% and 93.1%. Moreover, the Kaiser-Meyer-Olkin (KMO) test statistics indicate that the sampling of GMM instruments is adequate.

All of the GMM-estimates for the lagged dependent variable lie within the bounds set by the OLS and FE (upward-bias and downward-bias, respectively; Blundell and Bond, 2000; Roodman, 2009a). Additionally, both the Arellano-Bond test statistics for autocorrelation and the Hansen test statistics of overidentifying restrictions suggest that the models presented in Table 1 yield statistically valid estimates. We used the Hansen test instead of the Sargan test in order to test the exogeneity of the instrumental variables because the Sargan test is not robust enough to detect heteroskedasticity and autocorrelation (Roodman, 2009a).

Appendix E. Classification of manufacturing industries into categories based on R&D intensities

High-technology industries	Medium-high-technology industries
Aircraft and spacecraft	Electrical machinery and apparatus, n.e.c.
Pharmaceuticals	Motor vehicles, trailers and semi-trailers
Office, accounting and computing machinery	Chemicals excluding pharmaceuticals
Radio, TV and communications equipment	Railroad equipment and transport equipment, n.e.c.
Medical, precision and optical instruments	Machinery and equipment, n.e.c.
Medium-low-technology industries	**Low-technology industries**
Building and repairing of ships and boats	Manufacturing, n.e.c.; Recycling
Rubber and plastics products	Wood, pulp, paper, paper products, printing and publishing
Other non-metallic mineral products	Food products, beverages and tobacco
Basic metals and fabricated metal products	Textiles, textile products, leather and footwear

References

Arellano, M., &Bover, O. (1995). Another look at the instrumental variable estimation of error-components models. *Journal of Econometrics*, 68, 29–51.

Blundell, R., & Bond, S. (1998). Initial conditions and moment restrictions in dynamic panel data models. *Journal of Econometrics*, 87, 115–143.

Blundell, R., & Bond, S. (2000). GMM Estimation with persistent panel data: an application to production functions. *Econometric Reviews*, 19, 321–340.

Roodman, D. (2009a). How to do xtabond2: An introduction to difference and system GMM in Stata. *Stata Journal*, 9, 86-136.

Roodman, D. (2009b). A Note on the Theme of Too Many Instruments. *Oxford Bulletin of Economics and Statistics*, 71, 135–158.