On regional specialization of high- and low-tech industries

Janne Huovari* and Eero Lehto**

Abstract

Industries have varying abilities to benefit from externalities associated with geographical concentration, and are also likely to suffer in different degrees from crowding costs. This makes industries differ in their concentration process. We hypothesize that firms with low education levels (low tech) tend to be concentrated in areas with low urban costs and small populations, and firms with high education levels (high tech) to be concentrated in areas with high urban costs and large populations. This is shown to be true with Finnish regional-industry data.

Key words: regional data, specialization, location quotient, high-tech and low-tech industries

JEL Classification: R100, R120

* Pellervo Economic Research Institute (PTT)

** Labour Institute for Economic Research

1. Introduction

In this study we focus on regional specialization and, in particular, whether high- and low-tech industries differ in their specialization. Industries are likely to have varying abilities to benefit from pecuniary and technological externalities associated with geographical concentration, and are also likely to suffer in different degrees from crowding costs (the costs from intense competition, urban costs and congestion). This makes industries differ in their concentration process and some areas specialize in high-tech industries and others in low-tech industries. In this study we analyse whether variables that indicate regional agglomeration and crowding have different implications for the specialization of high- and low-tech industries identified by the education level of their workforce. The local forward and backward linkages are also taken into account, as is local productivity.

In the economic literature the set-up considered in this study arises in a natural way. The geographical concentration of economic activity is based on externalities. The theories of new economic geography (NEG) are grounded on pecuniary externalities and state that increasing returns, low transport costs, the strong preference for differentiated goods - which are weak substitutes - cause economic activity to cluster (see Ottaviano and Thisse, 2001). In addition to the pecuniary externalities, agglomeration gives rise to technological externalities, so-called spillovers, the most common of which are knowledge spillovers, which are generated by non-commercial transfers of knowledge between firms and individuals. The sensitivity to circumstances and the tacitness of knowledge both imply that technological externalities are geographically restricted (see Breschi and Lissoni, 2001a, 2001b; Morgan, 2004). The geographical proximity of spillovers ï which we also regard as being relevant when constructing the variables of our empirical analysis ï is extensively considered in the empirical research (see, for example, Jaffe et al., 1993; Keller, 2002; Orlando, 2004; Lehto, 2007).

In the framework of the NEG literature, agglomeration is customarily brought about by the demand effect. But in the models of Krugman and Venables (1996) and Venables (1996) interactions between a firm and its local customer or a firm and its local supplier is an agglomeration mechanism that creates local clusters. In this set-up, which assumes low trade costs between regions, deliveries to local customers (forward linkages) or purchases from local suppliers (backward linkages) also lead to local specialization. Agglomeration and specialization, in a way, co-exist, and specialized clusters in this literature need not be different as regards the clustering firmsô knowledge intensity and the rate of agglomeration outside the cluster concerned.

As regards technological externalities, it is important to notice that external knowledge in the form of spillovers is useful only for the firms who can utilize it and for the firm to which knowledge is also an important input. In line with this, Cohen and Levinthal (1989) have suggested that the cost of utilising knowledge in the public domain fruitfully is minimal only for firms that have accumulated sufficient technological capability to absorb external knowledge. On the industry level, it is evident that the utilization and processing of information play a bigger part in knowledge-intensive industries than in other industries. The materialization of technological externalities benefits these industries more than the others and makes these industries agglomerate.

But, as a counterforce to agglomeration, crowding costs in the form of intense competition, as suggested by Baldwin and Okubo (2006), may generate dispersion forces. Even more importantly, high urban costs or congestion externalities associated with agglomeration may force some firms to move to the periphery. It is evident that this concerns, above all, less knowledge-intensive firms ï who do not benefit from local knowledge spillovers.

If, however, one also takes into account the above-discussed trade-off between the concentration and dispersion tendencies, one gets a prophecy in which the specialized clusters differ from each other regarding the knowledge intensity of clustering firms and the amount of local agglomeration outside the local specialized clusters in question. According to this, the specialized clusters that consist of ñlow-techò firms with lower educational requirements for staff would be located on the periphery. On the other hand, the specialized clusters of knowledge users and producers with higher educational requirements could be located in the core. These clusters can take advantage of intra- and inter-industry spillovers and are not so easily damaged by congestion costs.

The approach of Brezis and Krugman (1997), which distinguishes between old and new users of technology, shows that the old users rather stick to the old technology and newcomers are more apt to adapt emerging technology on the periphery, where commuting and transport costs are low. In the dynamic equilibrium model of Duranton and Puga (2001) the economic landscape is characterized by the coexistence of diversified and specialised cities. These scenarios then propose that the high-tech activity does not necessarily tend to agglomerate and become located in the biggest cities. But after all, abstracting from technological externalities and from the heterogeneity of firms in their capabilities and needs to use and produce knowledge, these models cannot give an answer to the question that has arisen in our study: Do specialized clusters differ with regard to their knowledge intensity?

The empirical research on regional specialization typically evaluates whether industries tend to specialize (or concentrate) within a given time-period in a given geographical entity. Middelfart-Knarvig et al. (2000) discovered an overall increase in the specialization and concentration of industries since the 1980s in European manufacturing sectors. On the other hand, Krieger-Boden

(2000) found that specialization tended to decrease in French regions in the years 1973 - 1996. Some studies distinguish between concentrated and dispersed industries. According to Brülhart (1998) industries with increasing returns tend to be highly localized, i.e. they form clusters predicted by the NEG theory. By contrast, some labour-intensive industries are still much more evenly dispersed across European countries. Studies on urban areas show that mature industries benefit from localization economies, unlike new industries (see Henderson et el. 1995). Duranton and Puga (2001) discovered that in diversified environments cities specialize in churning new ideas, whereas in a more specialized environment cities produce standardised products.

Some studies examine whether the linkages between the users and suppliers of intermediate inputs have an impact on regional specialization. According to Paluzie at al. (2001) inter-industry linkages in Spain have a rather negative effect on specialization that is measured by gini indices of the geographical concentration of industries. Tohmo et al. (2006) obtained a similar result for the Finnish data. Audretsch and Feldman (1996), who also analysed the gini coefficient over U.S. regions, found that an industryôs R&D-intensity and the staffôs educational level increased the geographic concentration of production. On the same lines, Alonso-Villar et al. (2004) discovered that higher agglomeration in Spain is typical of industries with a higher technological level.

Rather few empirical studies place an emphasis on the economic impacts of urban and congestion costs. Broersma and Dijk (2008) examine the productivity impacts of traffic costs. Graham (2007) investigates the links between returns to urban density, productivity and traffic congestion.

The novelty of our study is that it examines the impacts of urban costs on regional specialization ï the implications suggested by theoretical studies. In our study the region is defined as being specialized in a certain industry, if its employment share in that industry exceeds its aggregate share of overall employment by a remarkable amount. The determination of this often used employment location quotient is, however, rarely analysed empirically. Owing to this definition for specialization, the concentration of aggregate employment in certain locations does not imply specialization as it does in more popular regional gini indices. In addition, by focusing on the region- and industry-specific specialization index, we are also able to control not only the industry level but also regional level variables, which can also be regarded as a contribution to the previous research. In this respect the specialization measure of this study also differs from the spatial concentration measure - given by Ellison Glaeser (1997) - which is calculated over all regions and industries in a larger geographic area.

The paper proceeds as follows. Section 2 discusses the theoretical arguments and introduces the hypotheses, which are empirically tested. Section 3 contains a description of the data and defines the variables that are used in the empirical analysis. Section 4 reports our results and the last section concludes.

2. Hypotheses

2.1. Specialization indices

Let emp_r^k be the amount of employment in region r and in industry k. For the total employment in region r ï which is $\sum_{k=1}^{K} emp_r^k$ - we use the notation emp_r and, respectively, emp^k describes the total employment in industry k ï which is $\sum_{r=1}^{R} emp_r^k$. The total employment in the country is then

$emp \, (= \sum_{k=1}^{K} \sum_{r=1}^{R} emp_r^k)$. We consider regional specialization by using the Location Quotient (LQ) ï also known as the Hoover-Balassa coefficient[1].

$$(1) \qquad Q_r^k = \frac{emp_r^k / emp_r}{emp^k / emp}.$$

The higher Q_r^k ôs value is, the more specialized region r is in industry k. Analysing this index, we are able to control both industry- and region-specific factors.

It must also be noticed that the Location Quotient for employment $\log(Q_r^k)$ can be decomposed into two additively separate factors, of which the first describes the impact of the average plant size and the second the impact of the number of plants on the concentration measure concerned. Let n_r^k denote the number of plants in region r and in industry k and n^k denote the number of plants in industry k. As Holmes and Stevens (2002) shows, the index (1) can be expressed in the form

$$(2) \qquad Q_r^k = \left(\frac{emp_r^k / n_r^k}{emp^k / n^k}\right)\left(\frac{n_r^k / emp_r}{n^k / emp}\right).$$

Taking logs we obtain

$$(3) \qquad \log(Q_r^k) = \log\left(\frac{emp_r^k / n_r^k}{emp^k / n^k}\right) + \log\left(\frac{n_r^k / emp_r}{n^k / emp}\right).$$

---

[1] Lafourcade and Mion (2003) analysed Italian regions by this index.

In this decomposition $\log(Q_r^k)$ depends on the relative size of the industry plants - in terms of their employment ï and, on the other hand, on the relative number of industry plants per capita. It is then obvious that in the econometric analysis the specialization index is also positively related to the difference in the local industry plant size from its national average. This was discovered by Holmes and Stevens (2002), who showed that in (3) $\log(Q_r^k)$ correlates positively with $\log(\frac{emp_r^k / n_r^k}{emp_r / n^k})$.

The analysis of regional specialization is complicated by the fact that the value of the above index may be systematically higher for the smaller regions because of the small number of plants. Why does the value of $\log(Q_r^k)$ decrease in the regionĝs size? Assume that the equal amount of plants belong to either an industry A or to an industry B and that the plantĝs value is zero if it belongs to industry A, and that it is one if it belongs to industry B. Suppose also that these plants are randomly picked and located in regions of different sizes. The expected value for the plants in each region is then 0.5, regardless of the regionĝs size. According to our definition, the region is the more specialized, the more the average value of its plants departs from 0.5. Let the number of plants that is picked for a region be n. Then the standard deviation of a regionĝs average value is $\frac{0.5}{\sqrt{n}}$ which decreases in n. In fact, when the region increases in the number of its plants, the mass of the distribution function for the average value of its plants becomes concentrated in the neighbourhood of 0.5. Smaller regions are, on the average, more specialized than larger regions when the exceptionally large share of either type of plants is the measure of specialization. According to this, the values of the $\log(Q_r^k)$ index can be either exceptionally high or low for small regions. But because some industries are totally missing in small areas, the low values of $\log(Q_r^k)$ are also omitted and so the average $\log(Q_r^k)$ becomes upward-biased in small areas. In this study we are interested in differences of regional specialization between low-skill and high-skill activities. Therefore we would like to analyse them with a specialization indicator that is not biased according to the regionô s size. To eliminate the possible bias, in the $\log(Q_r^k)$ index we regress it with the number of employees in the region.

We estimated

(4) $\log(Q_r^k) = \alpha + \beta \log(emp_r)$

with OLS and obtained -0.2178 for ß and 0.0067 for its standard deviation. Using this result we transform indicator $\log(Q_r^k)$ into form

(5) $lq_r^k = \log(Q_r^k) + 0.2178 * \log(emp_r)$

We then regard the corrected index lq_r^k as the specialization index that is analysed in this study.

The size correction according to the procedure (4) can already be expected to take into account some of the variation in the average plant size, although we do not correct the specialization index according to the average plant size in the given region and industry. But we assume that the average plant size still has a positive impact on lq_r^k. In this study we shall analyse the behaviour of lq_r^k without controlling $\log(\frac{emp_r^k / n_r^k}{emp^k / n^k})$ and with controlling this factor. In this way we shall see to what extent the regional specialization is explained by the plant size.

It is possible that the industries in which lq_r^k varies more than in the others have a dominant impact on the results obtained in the regression analysis of lq_r^k. To control the impact of this phenomenon we formulate an alternative index that is normalized to vary as much in each industry. We then define the categorized index $lq_r^k(c)$ so that

$lq_r^k(c)$ = 1, when lq_r^k < 30. percentile

$lq_r^k(c)$ = 2, when lq_r^k Ó30. percentile and lq_r^k < 70. percentile

$lq_r^k(c)$ = 3, when lq_r^k Ó70. percentile.

In the categorized index $lq_r^k(c)$ the smaller variation in industry i gets as big a weight as the larger variation in industry j, because the lower, middle and upper segments of lq_r^k within each three-digit NACE industry are defined as belonging to three different categories. The determination of $lq_r^k(c)$ is analysed by means of ordered probit analysis. The results obtained from this analysis are only commented on, not reported.

2.2. Hypotheses derived

In deriving hypotheses we do not foresee that the central variables of interest will have a different impact on $lq_r^k(c)$ than on lq_r^k. So the hypotheses presented concern lq_r^k as well as $lq_r^k(c)$.

We expect that agglomeration favours knowledge-intensive industries more than other industries. Owing to this, the urban costs, as a centrifugal force, push other industries more effectively toward the periphery than knowledge-intensive industries. This makes us believe that the industries with lowly educated staff (low-tech industries) tend to be located on the periphery, where the population (and its density) is small and where urban costs are also low. It is evident that the link of these plants to their local customers and suppliers also makes these firms cluster in the same regions. These regions then tend to specialize in low-tech industries so that the specialization indices considered obtain relatively high values. According to one of the main hypotheses of this study we believe that

(i) firms with a low educational level of staff tend to be concentrated in areas with low urban costs and small populations.

We also consider the specialization of knowledge-intensive industries. It is then thought that these industries characterize the staffö́s high educational level. Because knowledge-intensive industries can be thought of as being able to take advantage of the technological and pecuniary externalities of agglomeration, they are less damaged by high urban cost. Therefore we expect high-tech firms to be located in the core with a large population and high urban costs. Can these industries then be specialized according to our definition given above? Yes, they can. First, if, in particular, low-tech industries tend to avoid dense urban areas, this alone makes the core area specialize, more or less, in activities with a staff with a rather high educational level. Secondly, the specialization index is corrected according to equation (5) so that in heavily populated areas the specialization index tends to obtain larger values. Local forward and backward linkages can also promote the regional clustering of high-tech firms. According to another major hypothesis we then expect that

(ii) firms with a high educational level of staff tend to be concentrated in areas with high urban costs and a large population.

Figures 1 ï 3 give preliminary support for the hypotheses above. The heavily populated areas seem to be specialized in the activities with high educational level for staff and it looks as if low-tech concentrations are located mostly in remote areas.

[Figures 1 ï 3]

We then consider some other hypotheses that are not central to this study. These hypotheses concern backward and forward linkages and the interaction between regional productivity and specialization. We expect, as Krugman and Venables (1996) and Venables (1996), that the local linkages to the purchasers and suppliers of intermediated goods can lead to local concentration.

Accordingly, we hypothesize that

(iii) the presence of local intermediate good suppliers or purchasers increases local specialization.

In this study we also consider the variation of productivity over regions within each industry. According to Baldwin and Okubo (2006), the most productive firms tend to move and be located in the core regions owing to the demand effect associated with agglomeration and the potentiality to take advantage of scale economies. The most productive firms in all industries could move from the periphery to the core area, where specialization indices are typically at the average level. This tendency would not necessarily have any unambiguous effect on the local specialization indices considered. The situation will not differ, even if one considers knowledge-intensive and other industries separately. So we think that

(iv) the regional productivity would not have an unambiguous - positive or negative - impact on specialization.

In this study we also examine how local specialization reacts to some other factors, related to the regional incomes and to the local presence of other high- or low-tech clusters. These implications are discussed in more detail as we report the results.

3. The data and variables

The original data is the plant-level data set for Finnish industries. The data is constructed using the plant-level information of the Business Register, Employment Statistics, Financial Statements Statistics, Input-output Statistics, and Prices and Wages Statistics maintained by Statistics Finland (SF). The data covers all the firms in Finland whose annual turnover is above 0.5 million euros at least in one year during the period 1989 ï 2006. Owing to this restriction only a minor amount of total activity is removed. Therefore the region- and industry-level aggregates obtained from the data can be regarded as good estimates of actual aggregates. The data was aggregated to the NUTS4 regional and NACE three-digit industry level and it covers the years 1997 ï 2005.

The explanatory variable of this data is the specialization index lq_r^k (and $lq_r^k(c)$) in two different categories for the staffôs average educational level in the industry considered. For this kind of classification, we first calculated average post-basic-education education years in plants (edl) so that edl = 1*share1+2*share2, where share1 = the share of those who have had secondary-level education (ISCED levels 3 and 4) and share2 = the share of those who have had tertiary-level education (ISCED level 5 or higher). Using the edl variable we then calculated an aggregate educational level in each NACE three-digit industry over all years using the number of plantsô employees as weights. Using these industry-specific measures for the educational levels, we then pooled the data over all industries and calculated the 30th and 70th percentiles of this index for each year. Let edl_t^k denote the educational index in industry k in year t and edl_t the respective aggregate over all industries. According to our definition, the educational level is high in industry k in year t, if

(a) edl_t^k > 70th percentile of edl_t,

and the educational level is low in industry k in year t, if

(b) edl_t^k < 30th percentile of edl_t.

When considering specialization in activities with high educational requirements, we analyse lq_r^k or $lq_r^k(c)$ subject to restriction (a). Respectively, considering specialization in activities with low educational requirements, restriction (b) is presumed to be valid. The given formulation of explanatory variables collapses our data to be the region- and industry-level panel for the years 1997 ï 2005.

We use the notation lqh_r^k and $lqh_r^k(c)$ for the specialization index in the activities with the high level of education and lql_r^k and $lql_r^k(c)$ in the activities with the low level of education. The determination of specialization is analysed with controlling (and without controlling) the relative size of an average plant in the region (av_r^k), which is determined from

$$av_r^k = (\frac{emp_r^k / n_r^k}{emp^k / n^k}).$$

The key explanatory variables in our analysis are the urban cost and agglomeration variables. The price level for old dwellings with two rooms (dwprice) in each region is a proxy for urban costs or congestion costs. The size of population (popu) in each region is the agglomeration variable. Each of Finlandós 82 regions can be regarded as a commuting area whose population density is closely correlated with the size of its population. The size of the population variable is also a good measure of the demand for products in the area.

The share employment in the specialized clusters with the high level of education in other industries than in industry k is defined by the formula

$$shareh_r^k = \frac{\sum_{\substack{j=1 \\ j \neq k}}^{K} emp_r^j * lqh_r^k}{emp_r}.$$

The respective measure for the specialized cluster with the low level of education is obtained from

$$sharel_r^k = \frac{\sum_{\substack{j=1 \\ j \neq k}}^{K} emp_r^j * lqh_r^k}{emp_r}.$$

The forward and backward linkages reflecting the use of intermediate products are obtained from the input-output statistics for the year 2005. We use the value of 2005 for all years, because the information for the years 1996 ï 2005 provided by SF is not consistent. We then assume that the use of intermediate products in 2005 is a good approximate of the conduct in other years, too.

The data for the year 2005 gives us two NACE three-digit level measures for the use of intermediate goods:

vh^{kj} = the share of industry j of all intermediate goods purchased by industry k.

vt^{kj} = the share of industry j of all intermediate goods sold by industry k.

Using vh^{kj} we then obtain

$$(6) \qquad veh_r^k = \sum_{j=1}^{K} vh^{kj} * \frac{emp_r^j}{emp_r}$$

for the supply of intermediate products for industry k in region r. Respectively,

$$(7) \qquad vet_r^k = \sum_{j=1}^{K} vt^{kj} * \frac{emp_r^j}{emp_r}$$

describes the demand for intermediate products produced by industry k in region r.

Deriving the labour productivity variable we first calculate each plantôs labour productivity by dividing the plantôs real turnover by its size (the number of employees). This measure is then divided by the average (weighted) productivity for the whole NACE three-digit industry. In this way we obtained a plant-level productivity index that is comparable over industries. Let $prod_r^{i,k}$

denote this index for plant i in industry k and in region r. The region- and industry-level productivity ($prod_r^k$) is then obtained from

$$(8) \qquad prod_r^k = \sum_{i=1}^{n} prod_r^{i,k} * \frac{emp_r^{i,k}}{emp_r^k},$$

where n is the number of plants in the given industry and region.

Analysing specialization we also control the unemployment rate in the region (un_r), the average income in the region (y_r), the share of a NACE three-digit industryĝs fixed assets in relation to its total assets (cap^k) in industry k and year-dummies (year). Descriptive statistics for all variables are presented in Table 1.

Estimating the models, we take logs of all the variables (except those of the dummies). This makes the specialization measure considered additive as seen above in equation (3). Without controlling variable av_r^k , specialization can accrue from the location of one or more large plants in the area. If we control this variable , specialization describes solely the clustering of plants in the same region.

Interpreting the results, we address the fact that, in the real world, firms in the same industry do not pop up in a certain region where, for example, intermediate firms and population already exist. In the setting considered, the formation of a specialized cluster is rather an interactive process in which a plant and other plants in the same industry and even their intermediate clients and providers simultaneously enter or move to a certain region. They may also push other activities out of the region. The results tell us about this interaction rather than about one-way causality.

The main results are obtained by estimating OLS models. Estimation then takes into account both between and within variations. Because we are explaining flocking rather than the behaviour of a representative unit, we also find it difficult to solve the endogeneity problem that has arisen, by instrumenting the explanatory variables.

The relative productivity in the region and the specialization considered can typically be regarded as being closely related to each other. Suppose that the estimated coefficient for the productivity variable is positive. The sign of this coefficient does not, however, necessarily tell us that the majority of productivity plants tend to move into specialized clusters, which is propounded by Baldwin and Okubo (2006). It is also possible that plants in specialized clusters enjoy from each otherþs presence. We regard this interaction as being particularly interesting and therefore we have also estimated a model in which the productivity variable is instrumented. For this purpose we have specified the data as a panel form and so we can also use the productivity variableþs lagged value as an instrument. The results obtained from this analysis are reported in the appendix (Table A2).

To test the robustness of the results we have also estimated the ordered probit model in which the explanatory variable is $lq_r^k(c)$. These results are only commented upon.

4. Results

The results obtained by estimating the OLS model are reported in Tables 1 and 2.[2] The average plant size (av_r^k) increases the specialization rate in a remarkable way, as expected. But, on the other hand, the results reported in Tables 1 and 2 show that the inclusion of the average plant size variable does not essentially change the impacts of other factors. This indicates that clustering in

[2] Descriptive statistics for the variables are documented in table A1 in the Appendix.

terms of plantsô numbers alone is basically governed by the same factors as clustering in terms of plantsô employment.

The results reported in Table 1 tell us that the dwelling price has a negative impact on specialization in the low-education segment, whereas the respective impact is zero in the high-education segment. The coefficient of the population variable is negative in the low-education segment and positive in the high-education segment. These results confirm both hypothesis (i), in which we expected that firms with staff of a low educational level tend to be specialized in the areas with low urban costs and small populations, and hypothesis (ii), in which we expected that specialization in activities with high requirements for a staffôs education tends to be located in dense areas with rather high urban costs.

Table 1 around here

The above results reflect more or less the behaviour in the service industries. This is seen from the results that are reported in Table 2. In manufacturing the behavioural pattern is less clear. The coefficient of the population variable fulfils expectations, but its negative sign for the dwelling price variable in both low- and high-education segments is against expectations.

Unemployment is an indicator of under-utilization. Therefore one could expect that areas with high unemployment would also attract activities with a low educational level. The results in Table 1, however, show that in the clusters with lowly educated staff the unemployment rate tends to be low. The result seems to indicate that specialization is a good solution for the unemployment problem in remote areas. Specialization in industries with well-educated staff does not seem to be associated with low unemployment.

Table 2 around here

The proximity of the suppliers of intermediate products (backward linkages) contributes positively to specialization, at least in the low-education segment. The proximity of intermediate product customers (forward linkages) has a negative impact on specialization. Hypothesis (ii) is thus only partly realized. The results are not much different when manufacturing and services are analysed separately (Table 2).

The regionꞌs average income is positively related to specialization in both low- and high-education segments. This may also reflect the economic success of specialization as such. Capital intensity is negatively related to specialization, indicating that the internalisation of the externalities - which specialization creates - is based on the interaction between persons. The capital-intensive plants, which are also regularly large-sized, do not enjoy these externalities to such an extent as the labour-intensive plants. In each area there is no room for many specialized clusters. More specifically, the results indicate that the existence of other specialized clusters is not typical in areas in which well-educated clusters already exist.

It is also remarkable that high productivity seems to be associated with a high rate of specialization in both lowly and well-educated industries. It is difficult to say whether this result tells us that those plants that are originally efficient tend to cluster, as Baldwin and Okubo (2006) hypothesize or whether specialization, as a positive externality, increases the plantsꞌ productivity.

In the appendix in Table A2 we have reported the results obtained by estimating a random effect model. The productivity variable is also instrumented. The central hypotheses (i) and (ii) are still verified and the impact of the productivity variable is still positive. The ordered probit analysis, in which the explaining variable is an indicator $lq_r^k(c)$, confirms the results obtained in OLS (reported in Table 2) regarding the sign of the coefficient for the population and the dwelling price variable.

5. Conclusions

The geographic concentration of economic activity does not gather all firms into one region or make all agglomerations equal. Quite the opposite. Agglomerations greatly vary in size and in their structure of production. This follows from the fact that industries have different abilities to internalise the externalities associated with agglomeration and that they suffer in different degrees from the congestion related to agglomeration. This makes the knowledge-using industries that benefit most from wide-based agglomeration locate themselves in areas with large populations; the other industries locate themselves in remote or semi-remote areas with low urban costs. Some firms and plants belonging to both categories of industries enjoy co-location and they tend to form regionally specialized clusters.

According to our hypothesis ï for which we obtain empirical support in this study - the plants with high requirements for their staffsô education tend to form specialized clusters in areas with large populations and high urban costs, and the plants with rather low educational requirements cluster in sparsely inhabited areas with low urban costs.

Literature:

Alonso-Villar, O., Chamorro-Rivas, J-M., Gonzales-Cerdeira, X., 2004. Agglomeration economies in manufacturing industries: the case of Spain. Applied Economics 36, 2103-2116.

Audretsch, D.B., Feldman, M.P., 1996. R&D spillovers and the geography of innovation and production. The American Economic Review 86, 630-640.

Baldwin, R.E., Okubo, T., 2006. Heterogeneous firms, agglomeration and economic geography: spatial selection and sorting. Journal of Economic Geography 6, 323-346.

Breschi, S., Lissoni, F., 2001a. Localised knowledge spillovers versus innovative milieux: knowledge ñtacitnessò reconsidered. Papers in Regional Science 80, 255-273.

Breschi, S., Lissoni, F., 2001b. Knowledge spillovers and local innovative systems: a critical survey. Industrial and Corporate Change 10, 975-1005.

Breziz, E.S., Krugman, P., 1997. Technology and the life cycle of cities. Journal of Economic Growth 2, 369-383.

Broersma, L., van Dijk, J., 2008. The effect of congestion and agglomeration on multifactor productivity growth in Dutch regions. Journal of Economic Geography 8, 181-209.

Brülhart, M., 1998. Trading places: Industrial specialization in the European Union. Journal of Common Market Studies 36, 319-346.

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

Duranton, G., Puga, D., 2001. Nursery cities: urban diversity, process innovation, and the life cycle of products. American Economic Review 91, 1454-1478.

Ellison, G., Glaeser, E., 1997. Geographic concentration in US manufacturing industries: A dartboard approach. Journal of Political Economy 105, 889-927.

Graham; D.J., 2007. Variable returns to agglomeration and the effect of road traffic congestion. Journal of Urban Economics 62, 103-120.

Henderson, J.V., Kuncoro, A., Turner, M., 1995. Industrial development in cities. Journal of Political Economy 103, 387-410.

Holmes, T., Stevens, J., 2002. Geographic concentration and establishment scale. Review of Economics and Statistics 84, 682-690.

Jaffe, A., Trajtenberg, M., Henderson R., 1993. Geographic localisation of knowledge spillovers as evidence from patent citations. Quarterly Journal of Economics 108, 577-598.

Keller, W., 2002. Geographic localisation of international technology diffusion. American Economic Review 92, 120-142.

Krieger-Boden, C., 2000. Globalization, integration and regional specialization. Kiel Working Paper No. 1009, Kiel Institute of World Economies.

Krugman, P., Venables, A.J., 1996. Integration, specialization, and adjustment. European Economic Review 40, 959-967.

Lafourcade, M., Mion, G., 2007. Concentration, spatial clustering and the size of plants. Regional Science and Urban Economics 37, 46-68.

Lehto, E., 2007. Regional impact of research and development on productivity. Regional Studies 41, 623-638.

Middelfart-Knarvig, K. H., Overman H. G., Redding, S. J., Venables A. J., 2000. The location of European industry. Economic Papers No. 142, European Commission, D-G for Economic and Financial affairs, Brussels.

Morgan, K., 2004. The exaggerated death of geography: learning, proximity and territorial innovation systems. Journal of Economic Geography 4, 3-21.

Orlando, M., 2004. Measuring spillovers from industrial R&D: on the importance of geographical and technological proximity. RAND Journal of Economics 35, 777-786.

Ottaviano, G.I.P., Thisse, J-F., 2001. On economic geography in economic theory: increasing returns and pecuniary externalities. Journal of Economic Geography 1, 153-179.

Paluzie, E., Pons, J., Tirado, D.A., 2001. Regional integration and specialization patterns in Spain. Regional Studies 35, 285-296.

Tohmo. T., Littunen, H., Tanninen, H., 2006. Backward and forward linkages and concentration in Finnish manufacturing in the period 1995-1999. European Journal of Spatial Development (ESBD) (Refereed articles April 2006 - no 19).

Venables, A.J., 1996. Equilibrium locations of vertically linked industries. International Economic Review 37, 341-359.

Table 1. Specialization in the private sector for years 1997 ï 2005, OLS model

Variables in log form

explanatory variable	lql_r^k (with low education	lql_r^k (with low education)	lqh_r^k (with high education)	lqh_r^k (with high education)
av_r^k (average plant size)	0.9085*** (0.0121)		0.9604*** (0.0114)	
$popu_r$ (regionõs population)	-0.144*** (0.0168)	-0.1919*** (0.0219)	0.0860*** (0.0180)	0.1688*** (0.0251)
un_r (regionõs unemployment rate)	-0.2183*** (0.0417)	-0.0427 (0.0541)	0.1400*** (0.0430)	-0.0082 (0.0601)
$dwprice_r$ (dwelling price in the region)	-0.5959*** (0.0621)	-0.3162*** (0.0807)	0.0376 (0.0635)	-0.0298 (0.0887)
veh_r^k (supply of intermediate products)	0.3043*** (0.0180)	0.3366*** (0.0235)	-0.1633*** (0.0200)	0.1098*** (0.0276)
vet_r^k (demand for intermediate products)	-0.1427*** (0.0120)	-0.0853*** (0.0156)	-0.0381*** (0.0115)	0.0193 (0.0161)
$prod_r^k$ (regionõs productivity)	0.2650*** (0.0201)	0.3363*** (0.0261)	0.1598*** (0.0149)	0.173*** (0.0208)
y_r (average incomes in the region)	0.2284 (0.1578)	0.5503*** (0.2053)	0.6652*** (0.1571)	0.5202** (0.2195)
$sharel_r^k$ (other clusters with low education)	-0.0120 (0.0089)	0.028** (0.0116)	-0.0335*** (0.0101)	-0.0765*** (0.0141)
$shareh_r^k$ (other clusters with high education)	-0.0193** (0.0094)	-0.0452*** (0.0123)	-0.0328*** (0.0056)	-0.0484*** (0.0078)
cap^k (capital intensity in an industry)	-1.5880*** (0.0383)	-1.7969*** (0.0497)	-0.2483*** (0.0204)	-0.1746*** (0.0284)
constant	7.3048*** (0.3658)	5.5376*** (0.4749)	-1.4604*** (0.3734)	-0.9972* (0.5217)
Year dummies	yes	yes	yes	yes
N	8179	8179	7507	7507
R^2	0.5329	0.2092	0.5217	0.0662

Notes: Standard errors in parentheses. * Significant at 10 %, ** Significant at 5 %, *** Significant at 1 %.

Table 2. Specialization in manufacturing and services for years 1997 ï 2005, OLS model

Variables in log form

	Manufacturing	Services	Manufacturing	Services
explanatory variable	lql_r^k (with low education	lql_r^k (with low education	lqh_r^k (with high education)	lqh_r^k (with high education)
av_r^k (average plant size)	0.9355*** (0.0148)	0.6928*** (0.019)	1.0024*** (0.0143)	0.8769*** (0.0166)
$popu_r$ (regionôs population)	-0.2247*** (0.0218)	-0.0054 (0.0207)	0.0022 (0.0261)	0.1831*** (0.0214)
un_r (regionôs unemployment rate)	-0.1286** (0.0536)	0.0749 (0.0515)	0.089 (0.0619)	0.1625*** (0.0506)
$dwprice_r$ (dwelling price in the region)	-0.5485*** (0.0807)	0.0351 (0.0741)	-0.234*** (0.0889)	0.5262*** (0.0756)
veh_r^k (supply of intermediate products)	0.2157*** (0.0213)	0.3606*** (0.0431)	-0.1472*** (0.0288)	-0.2975*** (0.037)
vet_r^k (demand for intermediate products)	-0.1128*** (0.0163)	-0.1413*** (0.0293)	0.0345 (0.024)	-0.1801*** (0.0132)
$prod_r^k$ (regionôs productivity)	0.2968*** (0.0246)	0.2324*** (0.0327)	0.1562*** (0.0211)	0.0776*** (0.0183)
y_r (average incomes in the region)	0.4172** (0.2048)	0.7937*** (0.1862)	0.9197*** (0.2357)	0.6407*** (0.1796)
$sharel_r^k$ (other clusters with low education)	-0.0063 (0.0115)	-0.0167 (0.0127)	-0.0131 (0.0153)	-0.0401*** (0.0114)
$shareh_r^k$ (other clusters with high education)	-0.0081 (0.012)	0.0018 (0.0116)	-0.0175** (0.0073)	-0.047*** (0.007)
cap^k (capital intensity in an industry)	-1.4964*** (0.0445)	-0.5354*** (0.1447)	-0.174*** (0.0285)	0.3272*** (0.0316)
constant	7.516*** (0.4616)	1.0159* (0.5759)	1.4822*** (0.5274)	-7.1604*** (0.521)
Year dummies	yes	yes	yes	yes
n	5417	1911	3483	3933
R^2	0.5693	0.5062	0.6252	0.5415

Notes: Standard errors in parentheses. * Significant at 10 %, ** Significant at 5 %, *** Significant at 1 %.

Appendix

Table A1. Descriptive statistics

Variable	N	Mean	Std dev	Min	Max
lq *low education*	8197	2.7231	1.2192	-1.4030	7.0646
high education	7505	2.0216	1.1431	-1.6440	6.2902
av_r^k (average plant size)	8197	33.7354	55.0365	1.3778	816.6
	7505	33.7136	77.3799	1.2181	1363
$popu_r$ (regionôs population)	8197	120152	228672	6271	1235514
	7505	128802	229054	7129	1235514
un_r (regionôs unemployment rate)	8197	.1387	.0437	.0145	.3023
	7505	.1384	.0413	.0145	.3023
$dwprice_r$ (dwelling price in the region, ú/m^2)	8197	938.9	297.4	385	2361
	7505	972.8	297.1	423	2361
veh_r^k (supply of intermediate products)	8197	.0142	.0114	.0017	.1180
	7505	.0130	.0060	.0013	.0495
vet_r^k (demand for intermediate products)	8197	.0160	.0130	.0002	.1523
	7505	.0105	.0098	.0004	.2683
$prod_r^k$ (regionôs productivity)	8197	.9617	.6336	.0166	19.3385
	7505	.8344	.7235	.0015	31.3209
y_r (average incomes in the region)	8197	10.9857	1.8132	5.3411	20.2632
	7505	11.1080	1.7686	5.3411	20.2632
$sharel_r^k$ (other clusters with low education)	8197	.0395	.0379	.0000	.3921
	7505	.0379	.0340	.0005	.3921
$shareh_r^k$ (other clusters with high education)	8197	.0398	.0413	.0012	.2347
	7505	.0414	.0422	.0000	.2347
cap^k (capital intensity in an industry)	8197	.5707	.1380	.1878	.8382
	7505	.5285	.2082	.1225	.8580

Table A2. Specialization in the private sector for years 1997 ï 2005, GLS, Random effect models

Variables in log form

	treating $prod_r^k$ exogenous	treating $prod_r^k$ exogenous	treating $prod_r^k$ endogenous	treating $prod_r^k$ endogenous
explanatory variable	lql_r^k (with low education)	lqh_r^k (with high education)	lql_r^k (with low education)	lqh_r^k (with high education)
av_r^k (average plant size)	0.7458*** (0.0088)	0.8287*** (0.0086)	0.7442*** (0.0098)	0.8513*** (0.0095)
$popu_r$ (regionõs population)	-0.2492*** (0.0257)	0.1329*** (0.0273)	-0.2363*** (0.0259)	0.1436*** (0.0273)
un_r (regionõs unemployment)	-0.0635* (0.0343)	0.0796* (0.0415)	-0.0548 (0.0384)	0.0611 (0.0455)
$dwprice_r$ (dwelling price in the region)	-0.1581*** (0.0347)	0.0025 (0.0442)	-0.1646*** (0.0377)	0.0010 (0.0474)
veh_r^k (supply of intermediate products)	0.2765*** (0.0321)	-0.0608 (0.0383)	0.3109*** (0.0352)	-0.0713 (0.0396)
vet_r^k (demand for intermediate products)	-0.0787*** (0.0251)	0.0172 (0.0255)	-0.0913*** (0.0264)	0.0246 (0.0263)
$prod_r^k$ (regionõs productivity)	0.0515*** (0.0107)	0.0329*** (0.0085)	0.0467*** (0.0120)	0.0246*** (0.0095)
y_r (average incomes in the region)	0.1524*** (0.0471)	0.0034 (0.0583)	0.1651*** (0.0518)	-0.0707 (0.0629)
$sharel_r^k$ (other clusters with low education)	0.0021 (0.0056)	0.0023 (0.0082)	0.0027 (0.0057)	-0.0005 (0.0085)
$shareh_r^k$ (other clusters with high education)	0.0087 (0.0054)	-0.0040 (0.0032)	0.0082 (0.0057)	-0.0063 (0.0033)
cap^k (capital intensity in an industry)	-1.6223*** (0.0910)	-0.2439*** (0.0534)	-1.6501*** (0.0935)	-0.2578*** (0.0527)
constant	6.1740*** (0.3214)	0.6911* (0.3693)	6.1621*** (0.3368)	0.6987* (0.3797)
n	8197	7507	6899	6306
R^2 within	0.4852	0.5748	0.4735	0.5846
R^2 between	0.4981	0.4716	0.5132	0.4757

Notes: Standard errors in parentheses. * Significant at 10 %, ** Significant at 5 %, *** Significant at 1 %.

Figure 1. Specialized clusters with the low level of education, mean for years 1997-2005

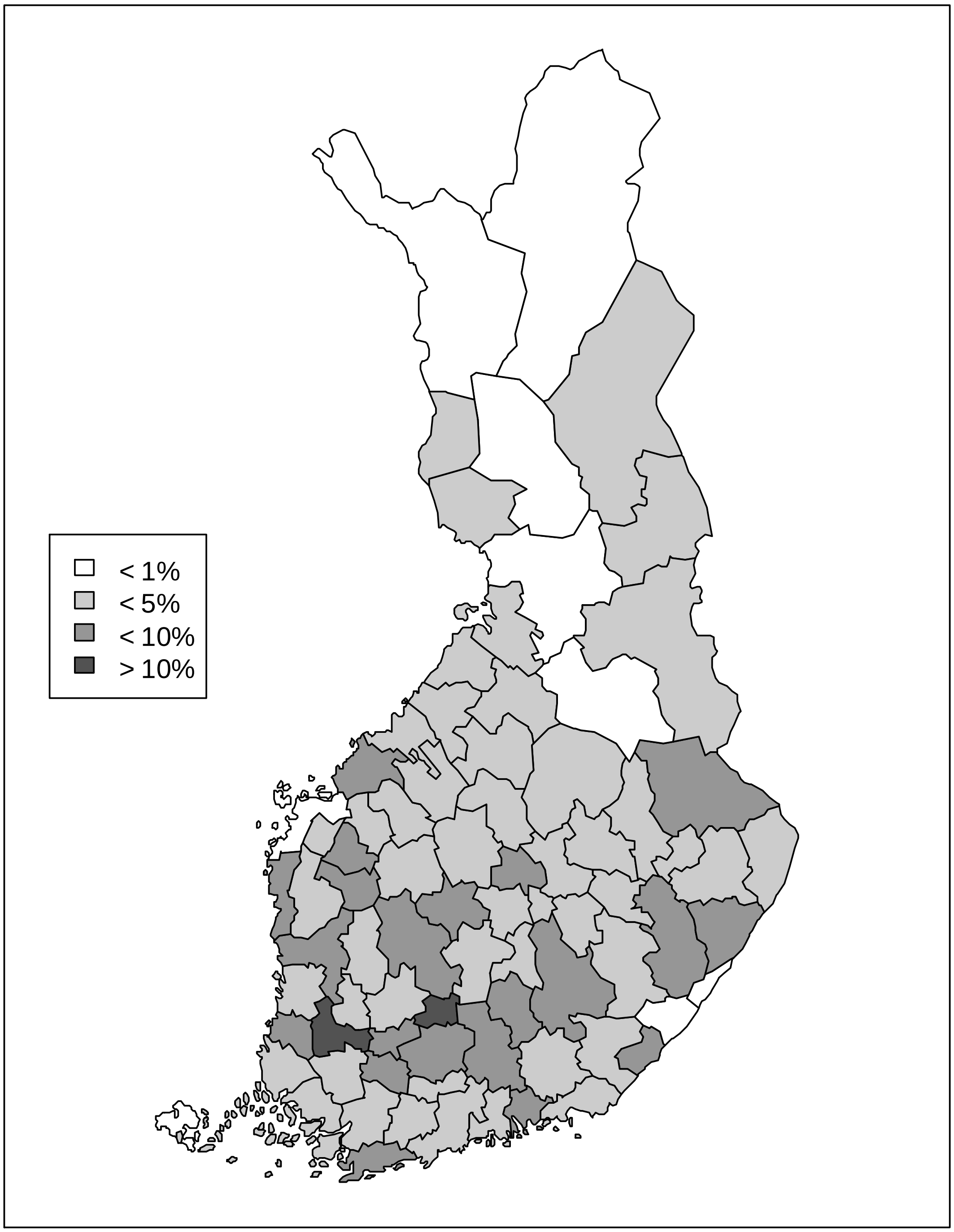

Figure 2. Specialized clusters with the high level of education, mean for years 1997-2005.

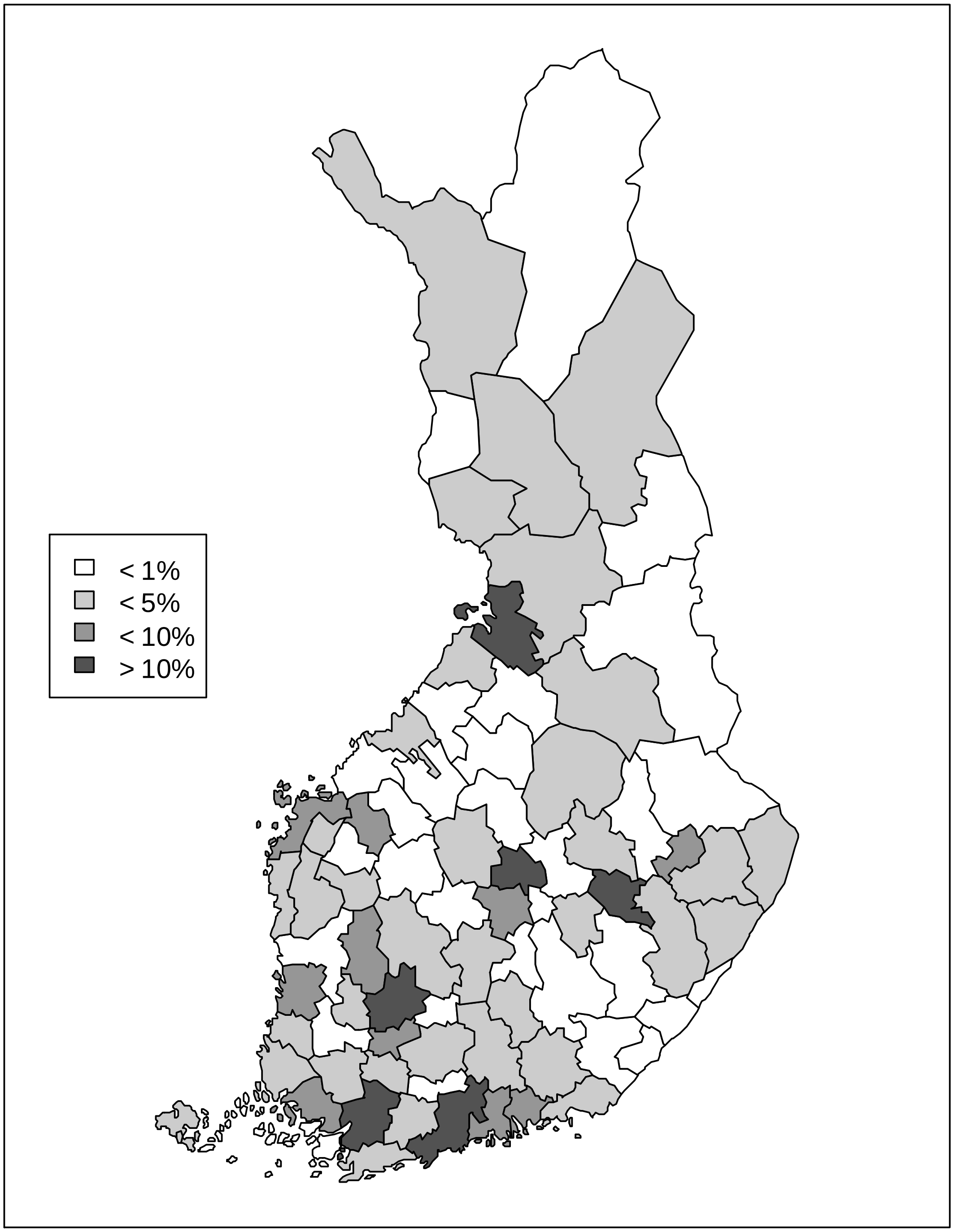

Figure 3. The size of population in year 2005