TI 2000-043/3
Tinbergen Institute Discussion Paper

The Geography of R & D

Hans Ouwersloot
Piet Rietveld

Tinbergen Institute
The Tinbergen Institute is the institute for economic research of the Erasmus Universiteit Rotterdam, Universiteit van Amsterdam and Vrije Universiteit Amsterdam.

Tinbergen Institute Amsterdam
Keizersgracht 482
1017 EG Amsterdam
The Netherlands
Tel.: +31.(0)20.5513500
Fax: +31.(0)20.5513555

Tinbergen Institute Rotterdam
Burg. Oudlaan 50
3062 PA Rotterdam
The Netherlands
Tel.: +31.(0)10.4088900
Fax: +31.(0)10.4089031

The geography of R&D;
tobit analysis and Bayesian approach
to mapping R&D activities for The Netherlands

Hans Ouwersloot[1]
Piet Rietveld[2]

[1]Universiteit Maastricht
[2]Vrije Universiteit Amsterdam

Abstract:
Regions vary strongly according to the participation of firms in R&D activity. By linking data on R&D activity at the firm level with GIS based data on economic, and other location features of zones we are able to investigate the impact of local factors on R&D involvement for various types of firms.
The relative importance of local factors as determinants of R&D involvement of the firms is estimated by means of a tobit model. Rather strong differences are found between zones in the same urban region. For example modern manufacturing firms located in the centre of the large cities have relatively low levels of R&D, the opposite holds for rings of zones at a certain distance from the cities. Bayesian methods are used for map presentations of the survey data.

Keywords: R&D, innovation, space, Bayesian approach.

Correspondence to:
Hans Ouwersloot
University Maastricht
FdEWB/MW/MMO
PO Box 616
6200 MD Maastricht
The Netherlands

Phone: +31 43 3883813
Fax: +31 43 3884918
e-mail: h.ouwersloot@mw.unimaas.nl

1. Introduction.

One of the key factors in economic development is technological innovation: the introduction of new or improved production techniques, products, and services. The introduction of these new elements is usually preceded by an intensive process of research and development (R&D). In the present paper we address the spatial aspects of R&D activity.

One might argue that the spatial aspects of R&D are not so important because R&D is only an input to the innovation process. The location where R&D takes place is not necessarily the place where the new products will be produced or the new production techniques will be applied. In large firms one may indeed observe clear distinctions between locations where R&D and where actual production takes place (see for example Chapman and Walker, 1992). And in many cases firms do not carry out the R&D itself but simply buy the improved products developed by other firms. In this case 'innovation' just means the adoption of new production techniques or new products developed and produced at places that may be located anywhere in the world. Thus, the economic benefits of R&D have spatial distribution patterns that are not necessarily connected to the spatial distribution of R&D itself. Yet, the spatial aspects of R&D are important for at least two reasons. First, R&D is important as an economic activity per se. The long-run trend towards a knowledge economy (Malecki, 1991, Suarez Villa, 1996) means that a substantial part of value added is created in this type of activity. In addition, although in large firms there is often a separation between R&D and production activities, such a separation is not always absolute and in smaller firms R&D and production take place at the same location. Successful R&D may then have a large impact on the development of the area involved.

R&D oriented firms offer well paid jobs, are generally considered to produce little environmental nuisance and are expected to have positive growth perspectives. Therefore it is no surprise that many regional and local governments are interested in attracting R&D oriented firms to there area (see Howells, 1984, Roger and Cote, 1987 and Camagni, 1991). An important question is of course what are the locational profiles preferred by these firms, and to what extent can these profiles be influenced by local or regional governments. These points will be addressed by the present paper.

In earlier phases, studies on the spatial aspects of R&D have been hampered by the lack of databases on the local production environment. With the development of GIS and large databases with micro data, the opportunities for a detailed analysis of the impact of local factors on R&D activity have substantially improved. In the present paper we will make use of the possibility to connect a database with innovation and R&D activity of firms with data bases of features of the economic and physical environment at the level of four digit postal areas or municipalities. By doing so we can determine which local factors have an impact on the level of R&D activity of firms.

The prime aim of the present paper is to analyse the locational patterns of R&D activity in The Netherlands. By using a tobit model we estimate the contribution of a number of location factors on the spatial distribution of R&D firms for three sectors: traditional manufacturing, modern manufacturing and services.

A secondary aim of the paper concerns the development and application of a methodology to represent spatial data on R&D involvement in the case of small regions. When survey data are used in the context of small spatial units one will find that a good number of the spatial units are empty, and that for most of them the number of observations will be so small that map presentations will easily yield misleading results. As a solution to this problem we use a Bayesian appoach.

The paper is structured as follows. In section 2 we give a further review of the literature on this subject. Some of the research methods used are presented in section 3. Data aspects and estimation results are presented in section 4. Section 5 discusses map presentations of the analysis by means of a Bayesian approach. Section 6 concludes.

2. A review of the literature on spatial aspects of R&D.

The location of R&D activities is influenced by both firm internal and firm external factors. Concerning the firm internal factors, Malecki (1980) focussed on the position of R&D activities of large corporations and identified basically three options for the location of the R&D activities. They can be linked to the location of the head quarter, to the production sites or to separate innovation centres connected to universities and public research institutions. In addition to these intra-firm linkages, which are especially relevant for large firms there are also firm external factors that play a role with both small and large firms.

Based on the literature we arrive at four clusters of locational factors: labour supply, information infrastructure, agglomeration economies and physical infrastructure.

-*Supply of qualified labour*. In knowledge oriented activities such as R&D, qualified labour is obviously of large importance (see for example Davelaar, 1991, Sivitanidou and Sivitanides, 1995). This has implications both for the required level of supply of highly qualified workers, and for the quality of the environmental amenities in a region. An attractive built or natural environment helps in attracting qualified personnel.

-*The informational infrastructure*. The manifest form of information infrastructure is often operationalised as the presence of public research institutes, universities, knowledge transfer centres etc. The information infrastructure works in two directions. First, a well developed infrastructure directly guarantees the presence of expert knowledge nearby. In addition spin-off effects - e.g. graduated students and university personnel that start their own company - can often be encountered. The above-mentioned location of R&D activities near 'innovation centres' (Malecki, 1980) is an example of this. Other authors who stress the importance of this location factor are Luger and Goldstein (1991) and Stough (1999). It is important to note that the information infrastructure is related to the supply of qualified labour mentioned above. An obvious advantage of regions with many university students is that they generate a large number of highly qualified workers in the region (see Beeson and Montgomory, 1990, and Felsenstein, 1999).

-*Agglomeration economies*. These economies are partly related to the two knowledge factors mentioned above. In addition, agglomeration economies are considered as cost-reducing factors that diminish uncertainty and increase production efficiencies (Camagni, 1991, Shefer and Frenkel, 1998, Stough *et al.*, 2000). One of the advantages of a location within a large metropolitan area is that one will have a diversified sectoral structure. Note that since corporate headquarters are typically located in large metropolitan areas, those corporations that link their R&D to headquarters will automatically also benefit from agglomeration economies (cf Malecki, 1980). Agglomeration economies may also function as a proxy for customer/supplier links which are found to be important for R&D (see for example Aydalot and Keeble, 1988).

-*Physical infrastructure* such as the express way density and access to hub airports, or regional airports are also reported to play a role in the locational patterns of R&D activities (Malecki and Nijkamp, 1988, Button et al., 1999, Stough *et al.*, 2000).

Most studies on the location of R&D activities have been carried out at a rather broad spatial level with metropolitan regions as the basic spatial unit. Important topics addressed are: the inequalities in the spatial distribution of R&D activities; the necessary conditions for regions to be an attractive location for these activities; and policy tools to stimulate R&D activities in less developed regions (see for example Howells, 1984, Roger and Cote, 1987, Shefer and Frenkel, 1998).

Studies on the more local aspects of R&D activities have been rare, however, a notable exception being Sivitanidou and Sivitanides (1995) who study the geography of R&D labs within Greater Los Angeles. They distinguish production amenities (for example proximity to local universities), worker amenities (appreciated by the highly qualified workforce) and local institutional constraints related to zoning. They find among others that the positioning of municipalities with respect to the freeway system, and proximity to universities are important factors explaining the attractiveness of municipalities as locations of R&D labs.

Several approaches can be distinguished in empirical research in this field. The first is a series of *case studies* of (spatial clusters) of successful firms like Scott and Angel (1987), Castells (1989, Ch 2), Aydalot and Keeble (1988) and Hilpert and Ruffieux (1991). In these studies the physical and cultural environments of the regions, and the policies of the local and regional governments receive much attention. One of the themes studied is R&D activity. A second group of studies addresses spatial innovation patterns by *cross section data at the regional level*. Indicators of R&D, patents, or innovations are measured at the level of regions; features of the regions are used to explain the level of R&D activity of firms in the region. Examples are Schmandt (1991), Feldman (1994), and Suarez-Villa (1997). A third approach is based on *cross section data at the firm level* (see for example Sanchez, 1992, König et al., 1995, Poot and Brouwer, 1996, Sleegers and Den Ouden, 1998, Shefer and Frenkel, 1998). In these studies usually much attention is paid to the firm specific features whereas the location specific aspects tend to receive less attention.

In the present paper we will follow the third approach (cross section micro data at the firm level) with a special emphasis on the location features of the firms. We combine a rich set of firm data on (amongst others) R&D activity with detailed regional information. Hence we are able to test a number of hypotheses which are put forward in the literature about the impact of the production environment on the level of R&D activity of firms.

3. Data

For our empirical research we have combined information from no less than four different data sources. A complete description of the data used can be found in Poot et al. (1997). Here we will present the main features of our four data sets.

Our core data set is the Innovation survey, commissioned by the Dutch Ministry of Economic Affairs and reported by Brouwer and Kleinknecht (1994). This survey was held among some 8000 Dutch firms in 1992 (with a response rate of approximately 50%), and investigated the firm's attitudes and activities with respect to innovations. Hence, it asked for the firm's R&D endeavors (people and money), the number of new products marketed, formal and informal information networks etc. As a measure of innovativeness we employed the share of employees involved in R&D activities, measured in full time equivalents. For a discussion of alternative innovativeness measures, see Brouwer and Kleinknecht (1996). Since our explanatory model concentrates on spatial data as exogenous variables, no other information from the innovation survey is used.

The innovation survey was stratified with respect to the size of firms, in order to have enough observations from large firms (>50 employees for manufacturing firms, >200 employees for service firms). All sectors are involved, except for the agricultural sector. In our study we distinguish between three sectors, viz. traditional manufacturing, modern manufacturing and the services sector, since it is generally accepted that levels of R&D efforts are different for these three broad sectors. The distinction is based on the two-digits SIC code. The traditional industry consists of the food, clothing, furnishing, construction and graphic industries. The modern industry consists of all other industries in SIC classes 2 and 3. The services industry encompasses SIC classes 4-9. The number of observations in each industry, and the number of observations with positive R&D involvement are given in Table 1. Note that the table refers to number of firms with positive R&D expenditures, which make up about 25% of total innovation expenses (Evangelista et al. 1997). The innovation survey also gives the address of the respondents from which the four-digit postal code can be derived. This postal code is used to infer the spatial characteristics of the zone of location for the observations.

	Traditional Manufacturing	Modern Manu-facturing	Services	Total
Number of observations (1)	819	1168	1923	3910
Number of observations with positive R&D involvement (2)	193	426	237	856
(2) as percentage of (1)	23.6	36.5	12.3	21.9

Table 1. Number of observations and observations with positive R&D in innovation survey.

It is important to realise that our model is set up to relate the level of *R&D involvement* of firms to location factors. The model does not explain the *presence* of firms in certain regions. Hence we can safely neglect physical planning policy influences. Such policy normally tries to direct the location of firms, and consequently may disturb models that try to explain location patterns, but not their level of R&D involvement.

We now turn to the data for the explanatory variables. We used as a starting point the four clusters of variables mentioned in section 2 (quality of labour supply, knowledge infrastructure, agglomeration economies, physical infrastructure). Given the low level of spatial detail the data did not allow us to include the quality of labour supply.

To investigate the quality of labour supply, we concentrate on the potential for worker amenities in the living environment. So, we use the *degree of urbanisation* to reflect the idea that highly qualified labour generally prefers to live in less populated areas, and so it may be attractive for R&D oriented firms to be located there as well. This effect may be mitigated by the fact that the Netherlands is a small country with a well-developed infrastructure. Nevertheless we expect a negative impact of degree of urbanisation: the lower the degree of urbanisation the more attractive the region is for firms with R&D activities.

For the knowledge infrastructure we used data with *distances between spatial units and universities* (there are 13 of these in The Netherlands). In addition we used the *density of service* firms in a narrow sense (SIC 8-9), comprising industries like finance, consultancy, research and the (semi-)public firms as another variable measuring the presence of information. The variable is defined as the number of service firms, divided by the total area of the municipality in which they are located, hence we expect a positive relation between this density and the measured R&D intensity.

The physical infrastructure is represented by three variables. We expect both the *presence of an inter-city railway station* and the *proximity to an express way* to have a positive effect on R&D intensity. For these two variables the motivation is that a good physical infrastructure is beneficial to information exchange. Moreover it gives high quality labour the opportunity to keep living in the desired environment. Presence of an inter-city railway station was measured as a dummy with value 1, when such a station was within a radius of 2 km of the firm. This somewhat arbitrary measure proved to give the best statistical results. Proximity to the express way was defined as the inverse of the log of the number of km to an express way. Consequently for both variables a positive parameter is expected. The third variable measures the *size of an industrial zone*. An industrial zone often gives incentives and opportunities to be involved in R&D activities. For example, in many instances the mere existence of an industrial zone implies the availability of space when the firm wants to extend its activities. Therefore a positive impact is expected.

Agglomeration effects are measured both with respect to the sector composition. The *diversity of the sector composition* is measured as 1 minus the Herfindahl index, a familiar measure of concentration. Diversity often gives incentives to be engaged in R&D activities, hence a positive parameter is expected. Further, the *share of the own sector in total industrial activities* is measured. Here we have two competing hypotheses. Similar to the argument above, others may stimulate a firm's R&D involvement, either as a mere incentive, or as a result of competitive strategy. Alternatively, the presence of other related firms may lead to co-operation, or even free-riding and thus lead to a negative impact on the firm's own involvement in R&D. Therefore we cannot a priori state an expectation concerning the sign of the parameter. Not only the composition, but also the mere presence of a well-developed complex of manufacturing firms may be an incentive to be involved in R&D activities. Therefore we measured the *density of manufacturing firms* (defined as the number of firms divided by the total area, cf. the definition of density of service firms) and expect that this variable will positively influence the R&D involvement.

For the operationalization of the above-described variables we use three alternative data sets. The Living Environment Database from the Dutch Ministry of Housing and Spatial Planning1 gives information on:

- Distance to the nearest connection to the express way network;
- Distance to the nearest inter-city railway station;
- Industrial zone area (km^2), as a percentage of total area of the postal code zone (km^2).

The Postal Code Register of the Dutch Central Bureau of Statistics (1993) gives information on the density of the built environment. This variable is defined using information about the number of physical addresses in the neighbourhood of each address in a given area. The variable appears to be very accurate about the density of the built environment (Van der Stadt, 1994).

The LISA database gives detailed information on the sector composition of regions. The number of firms and number of people employed are given at 1-digit SIC level. Spatial areas are given at the level of municipalities, of which there are – at the time of measurement - about 800 in the Netherlands (compared to approximately 3900 postal code areas). From the

1 We kindly thank the Dutch Ministry of Housing, Spatial Planning and the Environment (VROM/RPD) for providing these data

LISA data we derived the variables describing the sector composition of the zone.

4. Estimating the relation of spatial characteristics with R&D efforts

4.1 Description of the model

To investigate the relationship between the R&D efforts of firms and the spatial characteristics of the region where they are located, we consider a simple linear equation

$$y_i = \alpha + \beta X_i + \varepsilon_i$$

with y_i defined as the level of R&D activity of the firm, X_i a vector of variables capturing the spatial characteristics of the region where firm i is located. α and β are parameters, and ε an error term. We measure the R&D effort of firms as the share of the employees involved in R&D activities. When the firm does not perform any R&D activities this share is 0 by definition. As may be expected, a large number of "zero-observations" results. Consequently the basic assumption of regression, that y_i has a normal distribution is not fulfilled and so OLS estimation is not applicable.

The appropriate way to estimate model (1) with a large number of zero-observations is to interpret it as a Tobit model (see e.g. Maddala, 1983). This implies that we consider y_i as the observed realisation of an underlying latent variable that describes the intention of a firm to be engaged in R&D activities. When this intention y_i^* is positive, we measure $y_i = y_i^*$, thus we equate the observed variable to the latent variable. When the latent variable is negative, our measurement variable equals zero, thus $y_i = 0$. A negative intention to be engaged in R&D activities will lead to no realised R&D efforts, a negative effort is meaningless. So:

$$y_i = y_i^* \qquad \text{if } y_i^* > 0$$
$$y_i = 0 \qquad \text{if } y_i^* \leq 0$$

The interpretation of y_i^* as an intention to be engaged in some activity that is only observed as a non-negative variable is common in Tobit analyses (Amemiya, 1981). In this case the interpretation is even natural. One can think of y_i^* as the outcome of an internal analysis of firm *i*, that is optimising some firm-specific objective function with R&D efforts y_i* as one of the decision variables. When the optimal level y_i* for firm i is negative, it will definitely not be engaged in R&D activities and consequently a zero observation will follow in our data set. Then, while it is reasonable to assume that y_i* follows a normal distribution, only non-negative values y_i will be observed. These are exactly the conditions under which Tobit analysis is appropriate.

The procedures to estimate the Tobit model are well known and extensively described in Maddala (1983). We applied a standard procedure in the SAS software package (procedure LIFEREG), which is based on maximum likelihood estimation.

4.2 Results

As noted in Section 3 we distinguish three broad sectors: traditional manufacturing, modern manufacturing and services. The same model for all three sectors is assumed and therefore we estimate the relationships in one specification, including sector specific intercepts, and the option to include sector specific parameters. The statistical properties of the model dictated us to assume equality of parameters across sectors for some variables, whereas for other variables sector-specific parameters had to be estimated. The estimation result presented in Table 2 is the best we found.

Table 2 Estimation results. Dependent variable: share of employees involved in R&D activities. Tobit model, maximum likelihood estimation

<table>
<tr><th>Estimated coefficient (t-value)</th><th>Traditional manufacturing</th><th>Modern manufacturing</th><th>Services sector</th></tr>
<tr><td>Constant</td><td>-15.60</td><td>-7.58</td><td>1.96</td></tr>
<tr><td>Presence Intercity railway station</td><td>-.79 (.72)</td><td colspan="2">1.87 (3.53)</td></tr>
<tr><td>Proximity to express way</td><td>-.34 (-.92)</td><td>.05 (.17)</td><td>1.02 (3.19)</td></tr>
<tr><td>Degree of urbanization (address density)</td><td>.19 (.52)</td><td colspan="2">-.45 (-2.14)</td></tr>
<tr><td>Dummy extremely low urbanization</td><td colspan="3">-.51 (-1.11)</td></tr>
<tr><td>Area industrial zone</td><td colspan="2">.02 (1.67)</td><td>-.006 (-.31)</td></tr>
<tr><td>Density of manufacturing firms</td><td colspan="3">-36.20 (-2.80)</td></tr>
<tr><td>Density of service firms (SIC 8-9)</td><td colspan="3">5.82 (2.91)</td></tr>
<tr><td>Share own sector in total industry</td><td>-3.65 (-1.67)</td><td>-24.20 (-1.86)</td><td>-3.65 (-1.67)</td></tr>
<tr><td>Diversity of sector composition</td><td colspan="2">9.73 (1.93)</td><td>-2.37 (-.37)</td></tr>
</table>

Sources: Innovation survey 1992; LED, LISA 1992; PCR 93 (CBS)

We briefly discuss the estimation results and interpretation. The statistical properties of the model are modest. Compared to the model with only sector specific intercepts the log-likelihood improved with 27.2 points, for which we used 16 variables. The improvement is statistically significant according to the Likelihood Ratio Test, but the relative improvement of the likelihood function is only 3.8%. Realising that this relative improvement shows some similarity to the familiar R^2 (Amemiya, 1981), demonstrates a modest statistical performance.

The variables representing the physical infrastructure related to transport give the expected results for the services sector (or are insignificant). Hence a good connection to the outer world is an important determinant of R&D activities. The variable measuring the area of an industrial zone is only marginally significant (at the 10% level) for the manufacturing industries. This makes sense as these are the firms that usually have the greatest appetite for physical space. Similarly the presence of many service firms in a narrow stimulates the R&D orientation of firms.

The results for agglomeration effects that concentrate on the sector composition, are also confirming most hypotheses. Diversity helps, but only for the manufacturing firms, for service firms no effect is found. When there is a strong presence of the sector to which a firm belongs, this has a negative impact on the firm's own efforts in R&D. Recall that this was one of two opposing alternative hypotheses. So the data give support to the idea that presence of other firms in the same sector reduces the involvement for individual firms in R&D efforts. A likely explanation may be that the situation leads to co-operation. Alternatively, individual firms may display free-riding behaviour. Surprisingly, the density of manufacturing firms has a negative impact on the R&D efforts of firms. There is no apparent explanation for this finding.

The demographic part of the agglomeration phenomenon gives the hypothesised results. The effect of urbanisation is significant and negative, hence less populated areas (i.e. those with a low address density) lead to higher levels of R&D involvement. The explanation is that the highly qualified labour force looks for environmental amenities, that are supposed to be found in less populated areas. However, note that the dummy for extremely low urbanisation gives a counterweight to this above observation. Hence, the labour force is not looking for isolated districts.

Notice that we did not include knowledge infrastructure variables in Table 2. We included various operationalisations of proximity to knowledge institutions in the regression equation but never found a positive result: in all cases proximity had the wrong sign, suggesting that R&D levels are higher the further away a firm is located from a university. Thus, the result as found for example by Sivitanidou and Sivitanides (1995) for Los Angeles that proximity to universities matters is not confirmed for The Netherlands. The negative sign for this proximity variable implies that it is not consistent with theory (the null hypothesis is that it has a zero impact, the alternative hypothesis is a positive impact) it has been excluded in the final estimation. The issue of the importance of proximity to knowledge institutions for R&D activities has already been discussed earlier in The Netherlands. Davelaar (1991) mentions it as an important factor, but Vaessen and Wever (1990) argue that the density of knowledge centres is so high in The Netherlands that it does not matter how far a firm is removed from it.

Overall we find modest support for our hypotheses. The statistical properties of the model are relatively weak, however. Remarkable is the finding that in the Netherlands there is no statistical support for a stimulating role of universities on R&D involvement of firms.

5. Map presentation of the results

The estimation results give valuable information about the determinants of R&D efforts. They do not tell us, however, where exactly the favourable or unfavourable zones are located. Do they cluster, or are they "randomly" distributed over the country? Are they situated at what are generally believed to be R&D favourable production environments? And finally, are R&D minded firms located in zones that are R&D friendly, according to the analysis of Section 4?

To show how such questions can be addressed we introduce a cartographic representation of the results obtained. So we concentrate on the question how such issues can be addressed, and illustrate our approach for the case of the Netherlands, in particular the Western part of it, also known as the Randstad (see Figure 1). In addition this illustration will only be done for the modern manufacturing sector. For a full analysis of the results of our study - for which we lack the space here - we refer to Poot *et al.* (1997), where the results for all 12 provinces in the Netherlands, and for all three sectors are given and discussed in detail.

The next subsections give the maps that result from our analysis. In subsection 5.2 we give a presentation of the spatial distribution of R&D minded firms, as they are observed in our sample. To arrive at this map we applied a Bayesian approach to our data, which is explained in subsection 5.1. Subsection 5.3 shows the distribution of R&D friendly zones as they emerge from our analysis in Section 4. Subsection 5.4 discusses our findings.

5.1 A Bayesian approach to the representation of R&D data.

We want to represent data on the level of R&D activity of firms in our postal code zones. Given the large number of zones (about 3800) and the 'limited' size of the sample of firms interviewed (about 4000) we find that for a substantial number of zones we have no observation at all, and that for most other zones we have only one or a small number of observations. Especially when we are interested in the pattern of R&D activity for specific subsectors the problem of low numbers of observations per zone is evident.

A straightforward way to follow would be to compute the share of firms in a zone that is involved in R&D activities. In a substantial number of zones we have observations on only 1 or 2 firms. Depending on the probability that a firm is involved in R&D, this approach may lead to a substantial number of zero outcomes (when the 1 or 2 firms observed in a zone are not involved in R&D activity; see Table 3).

Share of firms involved in R&D		Number of firms in a zone				
		0	1	2	3	4
Number of firms in zone, involved in R&D activities	0	no result	0.00	0.00	0.00	0.00
	1		1.00	0.50	0.33	0.25
	2			1.00	0.67	0.50
	3				1.00	0.75
	4					1.00

Table 3. Share of firms involved in R&D, based on a sample of N firms in a zone of which r are involved in R&D.

Yet, in this way we are losing information because it makes a difference whether the result of a zero share of R&D oriented firms in a region is based on only 1 or on (say) 4 firms. In the former case one can be less confident that a zone is an unfavourable location for R&D oriented firms than in the second case. Thus, we want to be able to represent the degree of R&D involvement of firms in zones, while taking into account the strength of the information basis for this.

It appears that a Bayesian approach can be used to reach this purpose. Suppose we have a sample of N firms, and R of these firms are involved in R&D. This means that on average in each of the zones a share of R/N firms is involved in R&D. When we do not have further information (as is the case for zones for which no observations are available) this is the final result for firms in such zones.

How to proceed with zones where we do have observations? In a Bayesian approach we assume that the parameter p, indicating the a priori probability that a firm is involved in R&D is distributed according to a beta distribution[2] with expected value equal to R/N. Let Z be the number of zones. Then the density function is:

$$f(p) = B(R,N)\, p^{R/Z-1}\, (1-p)^{(N-R)/Z-1} \qquad 0 \le p \le 1$$

where B(R,N) is the beta function, implying that the integral of f(p) on the total interval equals 1. The expected value of p equals (R/Z)/[(R/Z)+(N-R)/Z]=R/N (cf. Zelner, 1971).

This a priori information has to be combined with information based on observed data. Consider a zone where we observe n firms of which r are involved in R&D. Assuming that we know the parameter p, the probability that this combination of r and p occurs can be computed by means of the binomial distribution:

$$P(r \mid n, p) = \binom{n}{r} p^{r} (1-p)^{n-r}$$

Application of Bayes' rule can be used to determine the posterior distribution of the parameter p in a zone. Application of the appropriate operations (cf. Zelner, 1971) leads to:

$$g(p) = c\, p^{R/Z+r-1}(1-p)^{(N-R)/Z+n-r-1}$$

where c is a constant to ensure that g is a density function.

The expected value of p according to the posterior distribution is:

2 The beta distribution satisfies the condition that p is between 0 and 1. The density in these extremes is zero. Other distribution functions could be used as well. The beta distribution is a common choice in this context because it is convenient to use. In Bayesian analysis the beta is called the natural conjugate of the binomial distribution (see e.g. Raiffa and Schlaifer, 1961).

$E(p) = (R/Z+r)/(N/Z+n)$.

Note that when we do not have observations (r=n=0) the expectation is equal to the a priori expectation. In addition, when we have a large number of observations the weight of the prior is relatively small. This formula can be used to compute the posterior expectation of the probability that firms in a zone are involved in R&D for any combination of n and r. Consider the case in our study where we have Z=3846 zones, sample size N=1923 (this is the number of firms in the service sector) and the number of these firms involved in R&D equals R=237, implying an a priori probability of R&D involvement of 12%. In Table 4 we give the posterior probabilities for some observed combinations of n and r.

Bayesian posterior probability of randomly selected firm involved R&D activities		Number of firms in a zone				
		0	1	2	3	4
Number of firms in zone, involved in R&D activities	0	.12	.04	.02	.02	.01
	1		.71	.42	.30	.24
	2			.82	.59	.46
	3				.87	.68
	4					.90

Table 4. Bayesian posterior probability that a service sector firm in a zone is involved in R&D, based on a sample of n firms in the zone of which r are involved in R&D (a priori probability is 12%)

The table shows that without observations (r=n=0) the posterior probability equals the prior probability. Note that with n=r=1 the posterior probability equals .71, whereas with n=r=2 we have .82. In Table 3 we have a score of 1.00 for both cases. When the number of observations per zone is substantial the differences between Tables 3 and 4 are rather small (see the column for n=4). However, for small values of n Table 4 dearly gives a more refined picture of the zone specific R&D probabilities than Table 3.

5.2 Observed R&D friendly zones for the modern manufacturing sector

Our presentation of the observed R&D friendliness of zones is based on the Bayesian approach described above. The parameter p which results from this exercise is formally equal to the probability that a randomly selected firm in the zone is involved in R&D activities. Obviously, we associate a high value for this parameter with an R&D-friendly production environment.

We calculate the parameter for all zones in the Netherlands. Because there is a large number of zones with no observations, these are not included in the successive procedure, despite the fact that the Bayesian approach permits to calculate the chance parameter. The remaining zones are classified in five classes, which count approximately the same number of zones. Figure 2 shows the zones in the Randstad that fall in the 5 distinguished classes. Note that, whereas for the Netherlands as a whole, each class counts approximately 20% of the zones, this does not necessarily hold for the part of the Netherlands represented in Figure 2. Moreover, for almost 75% of the zones no modern manufacturing firms were present in the sample. The five classes 1 to 5 each count approximately 20% of the zones with observations in the sample, so this means that these five classes capture about 5% of all zones shown on the map! The actual location of firms engaged in R&D can be inferred from Figure 2, since they are typically in the classes 3 to 5. This means that classes 1 and 2 consist of those zones that are present in the sample, but where no modern manufacturing firms involved in R&D are found.

The large number of empty zones clearly illustrates the point that for a large number of zones no data are given. Further, the figure suggests that the R&D friendly zones are quite dispersed over the Randstad. Nevertheless, some regularity can be found in two concentric circles around two cities, a relatively narrow one around Amsterdam and a relatively wide one around Rotterdam (cf. Figure 1).

Figure 2 also reveals an important drawback of working with postal code zones. These zones are based on the number of physical addresses. The Dutch postal service has designed its postal code zoning system so, that each zone counts approximately the same number of addresses. This immediately implies that zones with a low population density are physically larger than zones with high population density. The result for these large, but scarcely populated zones, tends to dominate the visual interpretation of maps like Figure 2.

5.3 Predicted R&D orientation

A similar procedure as in the previous section was applied to the results of the estimated model. This means that on the basis of the estimation results, as presented in Section 4, we calculated the predicted R&D friendliness for each zone, for a firm in the modern manufacturing sector. Next we divided all predictions for the Netherlands into five classes of approximately equal size. The map that results for the Randstad is given in Figure 3. Again, it is not necessarily so that a count of the five classes would result in approximately the same numbers (cf. the remark in the previous section). In this figure there are only a few empty zones, i.e. those for which the data of the exogenous variables in Table 2 are not available.

The figure shows true clusters of zones with a predicted high level of R&D activity, and of zones with a low predicted level of R&D activity. In particular, the region south-west of Rotterdam, the region of greater The Hague extending along the coast to the north, the region along the south border of greater Amsterdam bending in the south-east direction to Utrecht, and the region directly east of Utrecht, are all areas with a clear clustering of R&D-friendly zones. Notice further that the cities of Amsterdam and Rotterdam themselves are distinctly R&D-unfriendly for the modern manufacturing sector.

5.4 Discussion

Figures 2 and 3 give useful information about the location of R&D-friendly zones in the Randstad. Figure 2 reveals that the actually R&D minded firms are scattered over the region, although some regularities seem to emerge. Figure 3 on the other hand, makes clear that, based on the analysis of R&D data, clusters of R&D-friendly zones exist. The regularities of Figure 2 and the clusters of Figure 3 agree to a limited extent only. However, because of the large number of empty zones in Figure 2, the evidence is not conclusive.

The visual presentation of our analysis adds valuable insight to our understanding. It is equally clear that this insight is complementary to the estimation results of Section 4. Estimation tells us which production environment factors are important in explaining differences in level of R&D activity, the map of Figure 3 tells us where zones with favourable combinations of characteristics can be found, and how they are located in relation to each other. Finally, Figure 2 has the potential to visualise the errors in estimation what may be helpful in understanding the reasons for the errors.

When the maps are combined with still other information, we can again improve upon our understanding the spatial aspects of R&D orientation. For example, in Figure 3, the region east of Utrecht was identified as a potentially R&D-friendly area. However, in physical planning this very area is appointed as a region with high environmental quality, so that manufacturing activity is actually discouraged. Hence our results suggest that for this specific region there is a conflict between the goals of environmental preservation and support of facilitating R&D activity. Such a conflict obviously needs to be solved in the policy domain. The contribution of our analysis is to show that such potential conflicts exist.

6. Conclusion

The degree of involvement of firms in R&D activity depends on many factors like scale and relationships with sellers and buyers. In the present paper we focus on the spatial aspects of R&D intensity of firms. We study the spatial dimension at a high level of spatial detail, because indecisive results on the spatial aspects of R&D found in earlier research may be the consequence of a rather aggregate spatial approach.

A detailed analysis of the spatial aspects of R&D is possible by combining various data sources related to R&D activity of individual firms and accessibility as well as sector composition of zones. We find that for R&D activity in traditional manufacturing the sector composition of the zones is the most important local variable. Diversity of sector composition is beneficial, but the presence of firms of the same sector has a negative impact. A well-developed sector of services firms again helps to become actively engaged in R&D. Moreover there is some indication that physical space in the form of an industrial zone may be helpful as well.

Similar conclusions hold for modern manufacturers. In addition, an inter-city railway station and/or a lowly populated area are supportive of R&D activities.
For R&D activity in services the physical infrastructure is most important. We conjecture that this compensates for the absence of an effect of information infrastructure. Proximity to university is not important as long as the information can flow freely and comfortably over express ways or railways. If this conjecture is correct, the emergence of internet may have further reduced the importance of proximity to knowledge centres. For the services sector we find again that the presence of other firms in the same sector negatively influences the R&D activity, but the services sector in the strict sense helps to be involved in R&D.

Concerning policy recommendations some suggestions can be formulated. Sector composition is important. A diverse sector composition in an area with enough physical space appears to be attractive for R&D orientation of manufacturing firms in general. Also the help of specialised services firms is helpful, and this holds for all firms. Finally, good infrastructure is beneficial for R&D orientation of service firms.

A striking difference with results of other countries is that in The Netherlands, proximity to universities does not have a positive effect on R&D orientation of firms. Another point where the Dutch experience differs from that in many other countries is that the proximity to express ways does not play a significant role for R&D orientation in the manufacturing sector. It only matters for R&D in the service sector. Proximity to inter-city railway stations is more important in this respect: it has a significant impact for R&D orientation in both services and modern manufacturing.

The resulting spatial patterns indicate rather strong differences between zones within the same urban region. This underlines the importance of using a disaggregate approach to the spatial units of analysis. For example modern manufacturing firms located in the centres of the large cities have relatively low levels of R&D, the opposite holds for rings of zones at a certain distance from the cities.

Acknowledgment:

This research has been carried out with financial support of the Ministry of Economic Affairs. The authors thank Niels Vlaanderen, Tom Poot, Natasja Brouwer, Marcel Peek, Jeroen Heijs and Emiel Schelvis for various contributions and suggestions to the project. They also thank three anonymous referees for constructive comments

References:

Amemiya, T, 1981, "Qualitative Response Models, a survey" *Journal of Economic Literature*, **19** 1483-1536

Aydalot P, Keeble D, (eds) 1988 *High Technology Industry and Innovative Environments: the European Experience*, London, Routledge

Beeson P, Montgomory E, 1990, "The effects of colleges and universities on local labour markets" *Review of Economics and Statistics* **75** 753-761

Brouwer E, Kleinknecht A H, 1994 *Innovation in Dutch Manufacturing and Service Industries (1992)* (in Dutch), SEO report to the Ministry of Economic Affairs, The Hague, Beleidsstudies Technologie Economie, vol 27

Brouwer E, Kleinknecht A H, 1996, "Determinants of innovation: a micro econometric analysis of three alternative innovation output indicators" in, Determinants of innovation, the message from new indicators, ed. Kleinknecht A H, London, MacMillan Press, pp 99-124.

Button K J, Lall S, Stough R, Trice M, 1999, "High technology employment and hub airports" *Journal of Air Transport Management*, **5** 53-59

Camagni R, 1991 *Innovation networks; spatial perspectives* London, Belhaven Press

Castells M, 1989 *The Informational City*, Oxford, Blackwell

Chapman K, Walker D F, 1992 *Industrial location*, Oxford, Blackwell

Davelaar E J, 1991 *Regional Analysis of Innovation and Incubation*, Aldershot, Avebury

Evangelista R, Sandven T, Sirilli G, Smith K, 1997 *Innovation expenditures in European industry*, STEP report R-05 to the EC, DG XIII, Oslo (available on the internet 15-02-2000:

http://www.step.no/report/Y1997/0597.pdf (consulted 15-02-2000)).

Feldman M P, 1994 *The geography of innovation* Dordrecht, Kluwer

Felsenstein D, 1999 "Expenditure and knowledge based regional impacts associated with a university; some empirical evidence", in: *Regional Development in an Age of Structural Economic Change*, Eds. Rietveld P, Shefer D, Aldershot, Ashgate pp 73-95.

Hilpert U, Ruffieux B, 1991, "Innovation, policies and regional development:technology parks and regional participation in high tech in France and West Germany", in: *Regional innovation and decentralization*, Ed Hilpert U, London, Routledge, pp 61-88.

Howells J, 1984, "The location of research and development: some observations and evidence from Britain" *Regional Studies* **18** 13-29.

König H, Buscher H S, Licht G, 1995 "Employment, investment and innovation at the firm level" in: *Investment, productivity and employment* Paris, OECD, pp. 67-84.

Luger M, Goldstein H, 1991, *Technology in the Garden, research parks and regional economic development* Chapel Hill, University of North Carolina Press

Maddala G S, 1983, *Limited-dependent and qualitative variables in econometrics*, Cambridge, Cambridge University Press

Malecki E, 1980, "Corporate organization of R and D and the location of technological activities" *Regional Studies* **14** 219-234.

Malecki, E J, Nijkamp P, 1988, "Technology and regional development: some thoughts on policy" *Environment and Planning C* **6** 383-395.

Malecki E J, 1991 *Technology and economic development*, New York, Wiley

Miller R, Cote M, 1987 *Growing the next Silicon Valley: a guide for succesful regional planning* Lexington, Lexington Books, Lexington

Poot T, Brouwer E, 1996, “Meer spreiding van minder R&D intensiteit in Nederland” *(in Dutch)* in *Technologie en de regionale arbeidsmarkt* Eds Atzema, O, Dijk J van, Assen Van Gorcum, pp 29-40.

Poot A P , Brouwer N M, Ouwersloot J, Rietveld P, 1997 *Innovatie in de regio; provinciale innovatieprofielen*, (in Dutch), Amsterdam, SEO, Amsterdam

Raiffa H, Schlaifer R, 1961, *Applied Statistical Decision Theory* Boston Harvard University

Sanchez A M, 1992 “Regional innovation and small high tech technology firms in peripheral regions” *Small Business Economics* **4** 153-168.

Schmandt J, 1991 “Regional technology policy; a new role for the American States” in: *Regional innovation and decentralization*, Ed. Hilpert U, London, Routledge, pp. 157-176.

Scott A J, Angel D P, 1987 “The US semiconductor industry: a locational analysis” *Environment and Planning A* **17** 875-912.

Shefer D, Frenkel A, 1998 “Local milieu and innovations, some empirical results” *The Annals of Regional Science* **32** 185-200.

Sivitanidou R, Sivitanides P, 1995 “The intramatropolitan distribution of R&D activities, theory and empirical evidence” *Journal of Regional Science* **35** 391-415.

Sleegers W F, den Ouden A C, 1998 “Industriele R&D activiteiten en de vestigingsplaats" *(in Dutch)* in: *Innovatie in bedrijf en regio*, Eds Van Dijk J, Boekema F, Assen, Van Gorcum, pp. 81-104.

Stough R, Lall S V, Trice M P, 2000 “Infrastructure and technology in US metropolitan regions" *International Regional Science Review* (forthcoming)

Suarez-Villa L, 1996, "Innovative capacity, infrastructure and regional policy" in: *Infrastructure and the complexity of economic development* Eds Batten D F, Karlsson C, Berlin, Springer, pp. 251-270.

Suarez-Villa L, 1997 "Innovative capacity, infrastructure and regional inversion: is there a long term dynamic?" in: *Innovative behaviour in space* Eds Bertuglia C S, Lombardo S, Nijkamp P, Berlin, Springer pp. 291-305.

Vaessen P, Wever E, 1990, *Bedrijf en Omgeving* (in Dutch), Assen, Van Gorcum

Van der Stadt H, 1994 "The Netherlands, Country of Cities?" *(in Dutch)*, *Economische Statistische Berichten*, 466-469.

Zelner A, 1971, *An introduction to Bayesian inference in econometrics* New York, Wiley

Captions to figures

Figure 1 The region of the Randstad in the Netherlands with major cities

Figure 2 Classification of R&D-friendly postal zones in the Randstad in the Netherlands for the modern industry, based on sample information and Bayesian analysis.

Figure 3 Classification of R&D-friendly postal zones in the Randstad in the Netherlands for the modern industry, based on parameter estimation.