The geography of R&D: tobit analysis and a Bayesian approach to mapping R&D activities in the Netherlands

Hans Ouwersloot

University Maastricht, FdEWB/MW/MMO, PO Box 616, 6200 MD, Maastricht, The Netherlands;
e-mail: h.ouwersloot@mw.unimaas.nl

Piet Rietveld

Department of Economics, Free University, De-Boelelaan 1105, 1081 HV, Amsterdam,
The Netherlands; e-mail: prietveld@econ.vu.nl

Received 6 March 2000; in revised form 25 April 2000

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 regions, 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 the R&D involvement of 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 large cities have relatively low levels of R&D, and the opposite holds for rings of zones at certain distances from the cities. Bayesian methods are used for map presentations of the survey data.

1 Introduction

One of the key factors in economic development is technological innovation: the introduction of new or improved production techniques, products, and services. This is usually preceded by an intensive process of research and development (R&D). In this 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 where the new production techniques will be applied. In large firms, one may indeed observe clear distinctions between those locations where R&D and those where actual production takes place (see, for example, Chapman and Walker, 1992). And, in many cases, firms do not carry out the R&D themselves but simply buy improved products which have been 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. 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 their area (see Camagni, 1991; Howells, 1984; Miller

and Cote, 1987). 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 in this 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 microdata, the opportunities for a detailed analysis of the impact of local factors on R&D activity have substantially improved. In this paper, we make use of the fact that it is possible to connect one database that contains information on the innovation and R&D activity of firms with another containing 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 this paper is to analyse the locational patterns of R&D activity in the Netherlands. By using a tobit model, we estimate the influence of a number of location factors on the spatial distribution of R&D firms in three sectors: traditional manufacturing, modern manufacturing, and the service sector.

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, we find that a substantial number of the spatial units are empty and that, for most of these units, the number of observations of R&D activity will be so small that map presentations can easily yield misleading results. As a solution to this problem we use a Bayesian approach.

In section 2 we present a review of the literature on this subject. Some of the research methods used are discussed in section 3, and data aspects and estimation results are presented in section 4. In section 5 we discuss the map presentations of the results by means of a Bayesian approach, and we summarise our findings in section 6.

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

The location of R&D activities is influenced by factors which are both internal and external to the firm. With respect to the internal factors, Malecki (1980) focused on the position of R&D activities of large corporations and identified three basic options linking R&D to the location of the corporation headquarters, to the production sites, or to separate innovation centres connected to universities and public research institutions. In addition to these intrafirm linkages, especially relevant for large firms, there are also external factors that play a role within both small and large firms. Based on the literature, we arrive at four categories of external 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 great 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 attractively 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—for example, graduate students and university personnel that start their own company—can often be encountered. The aforementioned 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 et al (2000). It is important to note that the information infrastructure is related to the supply of qualified labour. 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 Montgomery, 1990; Felsenstein, 1999).

Agglomeration economies. These economies are partly related to the two knowledge factors discussed 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 there will be a diversified sectoral structure. Note that since corporate headquarters are typically located in large metropolitan areas, those corporations that link their R&D to their headquarters will automatically also benefit from agglomeration economies (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. The expressway density and access to hub or regional airports are also reported to play a role in the locational patterns of R&D activities (Button et al, 1999; Malecki and Nijkamp, 1988; 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 have been: inequalities in the spatial distribution of R&D activities; the conditions necessary for regions to be attractive locations for R&D; and policy tools to stimulate R&D activities in less developed regions (see, for example, Howells, 1984; Miller and Cote, 1987; Shefer and Frenkel, 1998).

Studies on the more local aspects of R&D activities have been rare, a notable exception, however, being that by Sivitanidou and Sivitanides (1995) who studied the geography of R&D labs within greater Los Angeles. They distinguished production amenities (for example, proximity to local universities), worker amenities (appreciated by the highly qualified workforce), and local institutional constraints related to zoning. They found, among other things, that the positioning of municipalities with respect to the freeway system and the proximity to universities were important factors in 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 the spatial clusters of successful firms (Aydalot and Keeble, 1988; Castells, 1989, chapter 2; Hilpert and Ruffieux, 1991; Scott and Angel, 1987). 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 using *cross sectional data at the regional level*. Indicators of R&D—patents or innovation—are measured at the level of the region; features of the regions are used to explain the level of R&D activity of firms within the region (Feldman, 1994; Schmandt, 1991; Suarez-Villa, 1997). A third approach is based on *cross sectional data at the firm level* (see, for example, König et al, 1995; Poot and Brouwer, 1996; Sanchez, 1992; Shefer and Frenkel, 1998; Slegers and den Ouden, 1998). In these studies much attention is usually paid to firm-specific features whereas location-specific aspects tend to receive less attention.

In this paper we will follow the third of these approaches (cross-sectional micro-data at the firm level) with a special emphasis on the location features of firms. We combine a rich set of data on (amongst other aspects) R&D activity within several firms, with detailed regional information. Hence, we are able to test a number of

hypotheses which have been put forward in the literature about the impact of the production environment on the level of a firm's R&D activity.

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 new innovations. It asked for the firm's R&D endeavours (in terms of both people and money), the number of new products marketed, formal and informal information networks, and so on. We used the share of employees involved in R&D activities, measured in full-time equivalents, as a measure of innovativeness. 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 was used.

The Innovation survey was stratified with respect to the size of the firms in order to have enough observations from large firms (>50 employees for manufacturing firms, >200 employees for service firms). All sectors other than the agricultural sector, were involved. In our study, we distinguish between three sectors—traditional manufacturing, modern manufacturing, and the service sector—since it is generally accepted that levels of R&D efforts are different for these broad sectors. The distinction is based on the two-digit SIC code. Traditional manufacturing consists of the food, clothing, furnishing, construction, and graphic industries. Modern manufacturing consists of all other industries in SIC classes 2 and 3. The service sector encompasses SIC classes 4–9. The number of observations in each industry and the number of observations which indicated a positive R&D involvement are given in table 1. Note that the table refers to the number of firms with a positive R&D expenditure; in other words, R&D expenditure which makes up about 25% of total innovation expenses (Evangelista et al, 1997). [The remaining 75% is spent on 'buying' innovations (see section 1).] The Innovation survey also gives the addresses 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 in which each firm is located.

It is important to realise that our model is set up to relate the level of *R&D involvement* of the 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 policies normally try 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.

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

	Sector			Total
	traditional manufacturing	modern manufacturing	service sector	
Total number of observations	819	1168	1923	3910
Number of observations with positive R&D involvement	193	426	237	856
Observations with positive R&D involvement as a percentage of total	23.6	36.5	12.3	21.9

We now turn to the data for the explanatory variables. We use as a starting point the four clusters of variables mentioned in section 2 (quality of labour supply, knowledge infrastructure, agglomeration economies, and physical infrastructure). Given the low level of spatial detail of the 'quality of labour supply' data, we did not use this variable.

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 the 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 that gave the *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, using only SIC classes 8–9), comprising industries like finance, consultancy, research, and the (semi)public firms as another variable measuring the presence of information. This 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 intercity railway station* and the *proximity to an expressway* 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. The presence of an intercity railway station was measured as a dummy with a value of 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 of the expressway was defined as the inverse of the log of the number of kilometres to an expressway. 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 with respect to the sectoral composition. The *diversity of the sectoral composition* is measured as '1 minus the Herfindahl index', a familiar measure of concentration. Diversity often gives incentives for firms to be engaged in R&D activities, hence a positive parameter is expected. Furthermore the *share of the firm's own sector* in total industrial activities is measured. Here we have two competing hypotheses. Similar to the argument above, the presence of other firms 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 cooperation 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; compare this with the definition of the density of service firms) and expect that this variable will positively influence the R&D involvement.

For the operationalisation of the aforementioned variables we use three alternative data sets. The Living Environment Database from the Dutch Ministry of Housing and Spatial Planning⁽¹⁾ gives information on: the distance to the nearest connection to the expressway network; the distance to the nearest intercity railway station; and the industrial zone area (km²), as a percentage of the total area of the postal code zone (km²).

The Postal Code Register of the Dutch Central Bureau of Statistics (data collected in 1993) gives information on the density of the built environment. This variable is defined by 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 sectoral composition of regions. The number of firms and the number of people employed are given at individual SIC digit 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 with approximately 3900 postal code areas). From the LISA data we derived the variables describing the sectoral 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:

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

where y_i is defined as the level of R&D activity of the firm; X_i as a vector of variables capturing the spatial characteristics of the region where firm i is located; α and β as the usual linear equation parameters; and ε_i as 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, by definition, zero. 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 an ordinary least squares estimation is not applicable.

The appropriate way to estimate such a model (with a large number of zero-observations) is to interpret it as a tobit model (see, for example, 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:

$$\begin{aligned} y_i &= y_i^*, & \text{if } y_i^* > 0; \\ y_i &= 0, & \text{if } y_i^* \leq 0. \end{aligned}$$

The interpretation of y_i^* as an intention to be engaged in some activity that is only observed as a nonnegative 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

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

the R&D effort, 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, although it is reasonable to assume that y_i^* follows a normal distribution, only nonnegative 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 we therefore estimate the relationships in one specification, including sector-specific intercepts and the option to include sector-specific parameters. The statistical properties of the model led 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. Results obtained by applying a tobit model to the relationship between the R&D efforts and the spatial characteristics of three broad industrial sectors [source: Innovation survey (Brouwer and Kleinknecht, 1994); Living Environment Database; LISA database; Postal Code Register].

Spatial characteristic of the region	Estimated coefficients, with <i>t</i> -values in parentheses		
	traditional manufacturing	modern manufacturing	service sector
Constant	-15.60	-7.58	1.96
Presence of an intercity railway station	-0.79 (0.72)	1.87 (3.53)*	1.87 (3.53)*
Proximity to an expressway	-0.34 (-0.92)	0.05 (0.17)	1.02 (3.19)
Degree of urbanisation (address density)	0.19 (0.52)	-0.45 (-2.14)*	-0.45 (-2.14)*
Extremely low urbanisation (dummy)	-0.51 (-1.11)*	-0.51 (-1.11)*	-0.51 (-1.11)*
Area of the industrial zone	0.02 (1.67)*	0.02 (1.67)*	-0.006 (-0.31)
Density of manufacturing firms	-36.20 (-2.80)*	-36.20 (-2.80)*	-36.20 (-2.80)*
Density of service firms (SIC 8-9)	5.82 (2.91)*	5.82 (2.91)*	5.82 (2.91)*
Share of own sector in total industry	-3.65 (-1.67)*	-24.20 (-1.86)	-3.65 (-1.67)*
Diversity of sector composition	9.73 (1.93)*	9.73 (1.93)*	-2.37 (-0.37)

* These coefficients were set equal to each other across the sectors indicated.

We will briefly discuss the estimation results and interpretation. The statistical properties of the model are modest. Compared with the null 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%. A modest statistical performance is demonstrated when we realise that this relative improvement shows some similarity to the familiar R^2 (Amemiya, 1981).

The variables representing the physical infrastructure related to transport give the expected results for the service sector (or are insignificant). Hence a good connection 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 sense stimulates the R&D orientation of firms.

The results for agglomeration effects that concentrate on the sector composition also confirm most of our expectations. 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, there is a negative impact on the firm's own efforts in R&D. We recall from section 3 that this was one of two opposing alternatives. So the data give support to the idea that the presence of other firms in the same sector reduces the involvement of individual firms in R&D efforts. A likely explanation may be that the situation leads to cooperation. 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, therefore less populated areas (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 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 a negative sign, suggesting that R&D levels are higher the further away a firm is located from a university. Thus, the result found 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), and it has therefore been excluded in the final estimation. The issue of the importance of proximity to knowledge institutions for R&D activities in the Netherlands has already been discussed. 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 them.

Overall we find modest support for our hypotheses. The statistical properties of the model are, however, relatively weak. A remarkable result is the lack of statistical support for a stimulating role of universities on the 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, exactly where 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. We concentrate on the question of how such issues can be addressed, and illustrate our approach for the case of the Netherlands and, in particular, the western part 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 following subsections show the maps that result from our analysis. In section 5.2 we give a presentation of the spatial distribution of R&D-minded firms, as they are

Figure 1. The region of the Randstad in the Netherlands showing the major cities.

observed in our sample. To obtain this map we applied a Bayesian approach to our data, which is explained in section 5.1. Section 5.3 shows the distribution of R&D-friendly zones as they emerged from our analysis in section 4. Section 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’ sample size of the 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. The problem of a low number of observations per zone is especially evident when we are interested in the pattern of R&D activity for specific subsectors.

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 for only one or two 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 one or two firms observed in a zone are not involved in R&D activity—see table 3, over).

However, in this way we are losing information because it makes a difference whether a zero share of R&D-oriented firms in a region is based on a total of only one or on (say) four 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.

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

Number of firms in the zone that are involved in R&D activities	Total number of firms in a zone				
	0	1	2	3	4
0	–	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

Note: the table entries represent the proportion of firms involved in R&D for each combination of N and R ; –, no result.

It appears that a Bayesian approach can be used for 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.

However, how do we 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:

$$f(p) = B(R, N)p^{\frac{R}{Z}-1}(1-p)^{\frac{N-R}{Z}-1}, \quad 0 \leq p \leq 1,$$

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

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

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

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

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

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

$$E(p) = \frac{R/Z + r}{N/Z + n}.$$

Note that when do 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

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

observations the importance of the a priori distribution 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 $N = 1923$ (the number of firms in the service sector), and $R = 237$ (the number of these firms involved in R&D), 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 .

The table shows that without observations ($r = n = 0$) the posterior probability is the same as equals the a priori probability. Note that with $n = r = 1$ the posterior probability equals 0.71, whereas with $n = r = 2$ we have 0.82. 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 clearly gives a more refined picture of the zone-specific R&D probabilities than table 3.

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 a zone of which r are involved in R&D (the a priori probability is 12%).

Number of firms in the zone that are involved in R&D activities	Total number of firms in a zone				
	0	1	2	3	4
0	0.12	0.04	0.02	0.02	0.01
1	–	0.71	0.42	0.30	0.24
2	–	–	0.82	0.59	0.46
3	–	–	–	0.87	0.68
4	–	–	–	–	0.90

Note: the table entries represent the Bayesian posterior probability of a randomly selected firm being involved in R&D activities; –, no result.

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 calculated the parameter p 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 calculation of the chance parameter. The remaining zones are categorised into five classes which contain approximately the same number of zones. Figure 2 (see over) shows the zones in the Randstad that fall into the 5 distinguished classes. Note that whereas for the Netherlands as a whole, each class counts for 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 presented in the sample. As the remaining 25% is split equally into 5 large classes, this implies that each of these 5 classes contains 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 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.

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.

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 wider one around Rotterdam (compare 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 contains approximately the same number of addresses. This immediately implies that zones with a low population density are physically larger than zones with a 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 was applied to the results of the estimated model. This means that on the basis of the estimation results, presented in section 4, we calculated the predicted R&D-friendliness of 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 the case that a count of the five classes would result in approximately the same numbers (see our remark in section 5.2). In this figure, there are only a few empty zones; zones for which the data of the exogenous variables in table 2 are not available.

Figure 3 shows true clusters both of zones with a high predicted level of R&D activity and of zones with a low predicted level of R&D activity. In particular, the region southwest 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

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

the southeast direction to Utrecht, and the region directly east of Utrecht are all areas with a clear clustering of R&D-friendly zones. Notice, also, 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 observed R&D-friendly firms are scattered over the region, although some regularities seem to emerge. On the other hand, figure 3 makes it clear that, based on the analysis of R&D data, clusters of predicted R&D-friendly zones do exist. The regularities of figure 2 and the clusters of figure 3 agree only to a limited extent. 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 the level of R&D activity, and 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, which may be helpful in understanding the reasons for these errors.

When the maps are combined with yet more information, we can again improve upon our understanding of 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 area is actually appointed as a region with high environmental quality, so that manufacturing activity is discouraged.

Hence, our results suggest that for this specific region there is a conflict between the goals of environmental preservation and 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 such as scale and relationships with sellers and buyers. In this paper, we focused on the spatial aspects of the R&D intensity of firms. We studied 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 have been the consequence of a rather aggregate spatial approach.

A detailed analysis of the spatial aspects of R&D was made possible by combining various data sources related to the R&D activity of individual firms, and the accessibility and sectoral composition of the zone in which the firm is located. For R&D activity in traditional manufacturing, we found that the sectoral composition of the zones was the most important local variable. Diversity of sectoral composition was beneficial, but the presence of firms in the same sector had a negative impact. A well-developed service sector also helps traditional manufacturing firms to become actively engaged in R&D. Moreover, there is some indication that physical space in the form of an industrial zone may also be helpful.

Similar conclusions held for modern manufacturers. In addition, an intercity railway station and/or a sparsely populated area were 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 information infrastructure effect. Proximity to a university is not important as long as the information can flow freely and comfortably over expressways or railways. If this conjecture is correct, the emergence of the Internet may have further reduced the importance of proximity to knowledge centres. For the service sector we again find that the presence of other firms in the same sector negatively influences the R&D activity, but the service sector in general helps other firms to be involved in R&D.

As far as policy recommendations are concerned, some suggestions can be formulated. Sectoral composition is important. A diverse sectoral composition in an area with plenty of physical space appears to be attractive for the R&D orientation of manufacturing firms in general. Also, the presence of specialised service firms is helpful for firms in all other sectors. Finally, good infrastructure is beneficial for the R&D orientation of service firms.

A striking difference with the results of other countries is that, in the Netherlands, proximity to universities does not have a positive effect on the R&D orientation of firms. Another point where the Dutch experience differs from that in many other countries is that the proximity to expressways does not play a significant role for R&D orientation in the manufacturing sector. It matters only for R&D in the service sector. Proximity to intercity 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 large cities have relatively low levels of R&D, and the opposite holds for rings of zones at a certain distance from the cities.

Acknowledgements. This research has been carried out with financial support from the Ministry of Economic Affairs. The authors wish to 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* (Routledge, London)
- Beeson P, Montgomery 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)* SEO report to the Ministry of Economic Affairs (in Dutch), volume 27, Beleidsstudies Technologie Economie, The Hague
- 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. A H Kleinknecht (Macmillan, London) 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* (Belhaven Press, London)
- Castells M, 1989 *The Informational City* (Blackwell, Oxford)
- Chapman K, Walker D F, 1992 *Industrial Location* (Blackwell, Oxford)
- Davelaar E J, 1991 *Regional Analysis of Innovation and Incubation* (Avebury, Aldershot, Hants)
- Evangelista R, Sandven T, Sirilli G, Smith K, 1997, "Innovation expenditures in European industry", report R-05 to the European Commission, STEP Group, Storgaten 1, N-0155, Oslo, Norway; available at <http://www.step.no/reports/Y1997/rep1997.html>
- Feldman M P, 1994 *The Geography of Innovation* (Kluwer, Dordrecht)
- 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 P Rietveld, D Shefer (Ashgate, Aldershot, Hants) 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. U Hilpert (Routledge, London) 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* (OECD, Paris) pp 67–84
- Luger M, Goldstein H, 1991 *Technology in the Garden, Research Parks and Regional Economic Development* (University of North Carolina Press, Chapel Hill, NC)
- Maddala G S, 1983 *Limited-dependent and Qualitative Variables in Econometrics* (Cambridge University Press, Cambridge)
- Malecki E, 1980, "Corporate organization of R&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: Government and Policy* **6** 383–399
- Malecki E J, 1991 *Technology and Economic Development* (John Wiley, New York)
- Miller R, Cote M, 1987 *Growing the Next Silicon Valley: A Guide for Successful Regional Planning* (Lexington Books, Lexington, MA)
- Poot T, Brouwer E, 1996, "Meer spreiding van minder R&D intensiteit in Nederland" [More spread of less R&D intensity in the Netherlands], in *Technologie en de Regionale Arbeidsmarkt* Eds O Atzema, J van Dijk (Van Gorcum, Assen) pp 29–40
- Poot A P, Brouwer N M, Ouwersloot J, Rietveld P, 1997 *Innovatie in de Regio; Provinciale Innovatieprofielen* [Innovation in the region; provincial innovation profiles] (SEO, Amsterdam)
- Raiffa H, Schlaifer R, 1961 *Applied Statistical Decision Theory* (Harvard University Press, Cambridge, MA)
- 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. U Hilpert (Routledge, London) pp 157–176

-
- Scott A J, Angel D P, 1987, "The US semiconductor industry: a locational analysis" *Environment and Planning A* **19** 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 intrametropolitan distribution of R&D activities, theory and empirical evidence" *Journal of Regional Science* **35** 391–415
- Sleegers W F, den Ouden A C, 1998, "Industriële R&D activiteiten en de vestigingsplaats" [Industrial R&D activities and place of location], in *Innovatie in Bedrijf en Regio* Eds J van Dijk, F Boekema (Van Gorcum, Assen) 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 D F Batten, C Karlsson (Springer, Berlin) 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 C S Bertuglia, S Lombardo, P Nijkamp (Springer, Berlin) pp 291–305
- Vaessen P, Wever E, 1990 *Bedrijven Omgeving* [The firm and environment] (Van Gorcum, Assen.)
- van der Stadt H, 1994, "Nederland, stedenland?" [The Netherlands, country of cities?] *Economische Statistische Berichten* **79** 466–469
- Zelner A, 1971 *An Introduction to Bayesian Inference in Econometrics* (John Wiley, New York)