

Schaffer, Axel; Rauland, Jan

Conference Paper

Regional efficiency in generating technological knowledge

50th Congress of the European Regional Science Association: "Sustainable Regional Growth and Development in the Creative Knowledge Economy", 19-23 August 2010, Jönköping, Sweden

Provided in Cooperation with:

European Regional Science Association (ERSA)

Suggested Citation: Schaffer, Axel; Rauland, Jan (2010) : Regional efficiency in generating technological knowledge, 50th Congress of the European Regional Science Association: "Sustainable Regional Growth and Development in the Creative Knowledge Economy", 19-23 August 2010, Jönköping, Sweden, European Regional Science Association (ERSA), Louvain-la-Neuve

This Version is available at:

<https://hdl.handle.net/10419/119145>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

Regional efficiency in generating technological knowledge

Axel Schaffer, Jan Rauland

Institute for Economic Policy Research (IWW), Karlsruhe Institute of Technology (KIT), Germany,
schaffer@iww.uni-karlsruhe.de, jan.rauland@iww.uni-karlsruhe.de

1 Introduction

There is broad consensus among economists that international competitiveness of modern economies heavily relies on their ability to generate knowledge (Romer 1990, Grossman and Helpman 1991, Barro and Sala-i-Martin 1997, Los and Verspagen 2006). In order to be more specific, many studies further define separate forms of knowledge: implicit versus explicit, tacit versus codified or local versus global knowledge (e.g. Jensen et al. 2007). Notwithstanding the importance of all forms of knowledge, the presented study focuses on the generation of codified knowledge as a main driver of technological knowledge and subsequently regional growth.

Following the relevant literature, a region's ability to generate technological knowledge relies, among other factors, on the educational achievements of its workforce, the degree of intramural R&D expenditures and the existence of regional knowledge spillovers. However, the pure existence might be of little help, if these factors are not used efficiently (Wang and Huang 2007). The paper at hand particularly deals with this issue and foresees to analyse the regions' efficiency in generating technological knowledge in further detail. Therefore, the first step foresees to identify total (output-) efficiency of EU regions at NUTS 2 level in section 2. For this purpose, we apply an outlier robust enhancement of the data envelopment analysis (DEA), the so-called order- α -frontier analysis with patent applications as output and human capital as well as R&D expenditures as input variables (Daouia and Simar 2005).

The analysis points to the existence of a spatial factor. This conflicts with a basic assumption of DEA, which indeed presumes a strong degree of independency of the operating units. Therefore, section 3 introduces a geoaddivitive regression analysis based on markov fields, which allows for decomposing the total efficiency into a smoothed spatial and a non-spatial part. The smoothed part shows spatial dependence and identifies regions with a comparative (dis)advantage due to their location in space. The non-spatial effect gives an idea on a region's efficiency compared to the neighboring and nearby regions and can be interpreted as structural efficiency.

Section 4 concludes and discusses some policy implications.

2 Regional efficiency

2.1 *Theoretical framework and related literature*

Regional efficiency analysis traces back to several studies on the economic performance of Asian regions. Macmillan (1986) and Hashimoto and Ishikawa (1993) applied DEA to rate the efficiency of Chinese and Japanese cities respectively. Charnes et al. (1989) used the same methodology to identify urban industrial performance and assess regional planning tools in China, and Seifert and Zhu (1998) applied DEA to monitor the productivity growth of China's industries over several decades.

More recently, DEA is also used to analyze efficiency in the context of R&D activities. Wang and Huang (2007), for example, apply the technique for a cross-country study on relative R&D efficiency. Guan and Chen (2010) as well as Zhang et al. (2003) follow a similar approach for Chinese regions.

Clearly DEA is not the only way to measure efficiency. Among others stochastic frontier analysis (SFA) could be mentioned as a more complex approach, which requires, compared to DEA, more indepth knowledge on the functional relationship of inputs and outputs. As a consequence the approach reduces the uncertainty of the results (Aigner et al. 1977, Li 2009).

DEA, in contrast, is not bound to a given theory but the results are only determined by the data. This is considered an advantage for the presented study, as the main aim is to receive directly observable evidence with respect to regions' efficiency in generating technical knowledge.¹

At the same time the strong orientation on the data comes along with a major drawback. Since the estimated frontiers generally envelop all data, the sensitivity to outliers is problematic for most DEA models. The presented paper deals with this problem by using an outlier robust enhancement of DEA, the so-called order- α -frontier approach (Daouia and Simar 2005).² As a consequence, the frontier does not reflect the maximum achievable output anymore, but some outliers are allowed to lie above the frontier.

2.2 *Methodological remarks*

The estimation of the regions' efficiency in generating technological knowledge follows the basic principles of DEA.³ These comprise the selection of decisionmaking units (DMUs), the definition of inputs and outputs and the formulation of the mathematical model.

¹ See Guan and Chen for a comprehensive discussion on the advantages of DEA in the R&D context.

² See section 2.2 for details on the approach.

³ In the following we assume a basic familiarity with DEA. Among others, Charnes et al. (1994) or Cooper et al. (2006) provide a good introduction into DEA.

Definition of DMUs

DMUs are characterized by a uniform production function to transfer a set of inputs into one or multiple outputs. Technological efficiency, for example, is often analyzed at the level of firms that produce the same goods or services. If the focus is on regions' efficiency in using available factors to generate certain outputs, DMUs are generally defined as spatial entities.

Following the territorial system of the European Union – the so-called Nomenclature Territorial Statistical Units (NUTS) – efficiency is analyzed at the NUTS 2 level, which can be considered the basic administrative unit chosen by the EU for a broad set of regional policies. In total, the analysis accounts for 192 NUTS 2 regions of continental Europe.

Definition of inputs and outputs

There is broad consensus among regional scientists that regions' ability to generate technical knowledge is driven, on the one hand, by region-specific and, in the medium term, immobile factors as well as, on the other hand, by their ability to attract (or provide) funds for R&D activities.

Immobile factors include, but are not limited, to the level of human capital, the existence of major scientific institutions (e.g. universities) and the overall economic structure. Funds could either be private or public in nature and might be used for rather specific scientific projects or provided for basic research. While a certain environment with immobile factors might indeed affect the structure and dimension of funds, they are in principal mobile and could (at least in part) easily be removed to other regions within short time.

The chosen inputs, which aim to reflect both categories, are defined according the regions' level of human capital and total intramural R&D expenditures. For the sake of comparability, R&D expenditures are considered as a percentage of the regional GDP of the year 2002. Human capital, also defined in relative terms, is given as the percentage of a region's labor force that belongs to the human resources in science and technology (HRST). Following the Canberra Manual, HRST comprise all members of the labor force, who either dispose of tertiary education in the field of science and technology, or are employed in a corresponding occupation where the above qualifications are normally required (OECD 1993). In this context, the requirements are in line with internationally harmonized standards: the International Standard Classification of Education (ISCED) and the International Standard Classification of Occupation (ISCO).

On the output side, the regions' patent applications to the EPO per person employed are chosen as a proxy for the ability to generate knowledge.⁴ The indicator is frequently applied in the literature despite its potential shortcomings (Grupp and Schmoch 1999, Pavitt 1985): First, patents are just granted to products and not to services or immaterial goods. Thus, not all innovations are patentable. Second, not all manufacturing industries are technology-driven and rely on R&D. Third, not all innovations are filed as a patent even if they are patentable. Firms sometimes decide against a temporary monopoly of a patent, just to keep their inventions secret (Rammer 2003). Hence, not all innovations are reflected in patents. The clear focus on generating technological knowledge lessens the relevance of the first and the second issue for the study at hand. In contrast, the third shortcoming also holds for the presented analysis. At the same time, patents are still rich of information and can be considered one of the most powerful indicators for measuring technological output (Griliches 1990, Audretsch 1995).

Mathematical formulation

Traditional DEA describes one of the most popular tools to measure efficient production boundaries of firms or regions. However, the models are often found to be rather sensitive to outliers. The presented order- α -frontier analysis, which is described in full detail by Daouia and Simar (2005), aims to overcome this shortcoming by defining a frontier function that leaves out extreme observations. Presuming that regions improve their efficiency more likely by growing outputs rather than decreasing inputs, the output-oriented version of the model is applied for this study.

We start with a brief specification of the traditional DEA model (Equations (3) to (5)), which defines the baseline for the extension according equations (6) and (7).

Any region disposes of a set of inputs $x \in \mathbb{R}_+^p$ to generate a set of outputs $y \in \mathbb{R}_+^q$. Feasible combinations of (x, y) are defined as:

$$(3) \quad \Psi = \{(x, y) \in \mathbb{R}_+^{p+q} \mid x \text{ can produce } y\}.$$

The frontiers of Ψ reflect maximum outputs that can be produced with given inputs. Thus, the regions' efficient frontier can be defined in the following way:

$$(4) \quad Y^\theta(x) = \{(x, y^\theta(x)) \mid y^\theta(x) \in Y(x) : \lambda y^\theta(x) \notin Y(x), \forall \lambda > 1\}$$

where $Y(x)$ describes the set of technologically feasible outputs, and $y^\theta(x)$ denotes the

⁴ The data refer to an average between 2002 and 2004.

maximum achievable output of a unit that produces at input level x . This, in turn, allows the definition of a unit's efficiency score as:

$$(5) \quad \lambda(x,y) = \sup\{\lambda | (x, \lambda y) \in \Psi\} = \sup\{\lambda | \lambda y \in Y(x)\}$$

where $\lambda(x,y) \geq 1$ is the proportionate increase of output y a region operating at input level x has to attain to be efficient.

The frontiers of Ψ , which are unknown in practice, can be determined by nonparametric estimators, such as the free disposal hull (FDH) estimator (Deprins et al. 1984). However, these estimators often envelop all data points, which in turn makes traditional DEA rather sensitive for outliers.

The problem can be solved by using more robust estimators which deal with extreme observations in a different way. Instead of defining the efficient boundary according the uppermost technically achievable output (for any given input), extreme observations can for example be allowed to lie above a partial frontier (Cazals et al. 2002). For this case, Aragon et al. (2005) introduced the concept of the order- α partial frontier ($\alpha \in [0,1]$) which is applied for the presented study.

With $S_{Y|X}(y|x)$ defined as the probability $\text{Prob}(Y \geq y | X \leq x)$ and $F_X(x)$ as the probability $\text{Prob}(X \leq x)$, Daouia and Simar (2003) define the order- α -quantile output efficiency score for each unit $(x,y) \in \Psi$ as:

$$(6) \quad \lambda_\alpha(x,y) = \sup\{\lambda | S_{Y|X}(\lambda y | x) > 1 - \alpha\} \text{ for } F_X(x) > 0 \text{ for } \alpha \in [0,1]^5$$

Note that, following this approach, each unit is only compared with units disposing of equal or worse input levels. A unit (x,y) is efficient at level α , if it lies on the calculated frontier, this means if $\lambda_\alpha(x,y) = 1$. In this case, the unit is dominated by a unit with lower input with a probability $\leq 1 - \alpha$.⁶ Units below the efficient boundary ($\lambda_\alpha(x,y) > 1$) are considered inefficient. This means a unit with similar or worse input delivers higher output with a probability $> 1 - \alpha$. On the contrary, units above the frontier ($\lambda_\alpha(x,y) < 1$) could reduce their outputs but would remain output efficient.

The applied nonparametric estimator of $\lambda_\alpha(x,y)$ is obtained by substituting $S_{Y|X}(y|x)$ with its empirical correspondent $\bar{S}_{Y|X,n}(y|x)$, based on the samples $(X_1, Y_1), \dots, (X_n, Y_n)$ where X is the observed input and Y the output. This results in the empirical efficiency score $\bar{\lambda}_{\alpha,n}(x,y)$.

⁵ The efficiency score converges from below to the Debreu-Farell output efficiency measure $\lambda(x,y)$.

⁶ In case a unit's input equals the minimum level, which means $F_X(x) = 0$, the unit cannot be dominated by a unit with less input and is therefore considered to be efficient ($\lambda_\alpha(x,y) = 1$).

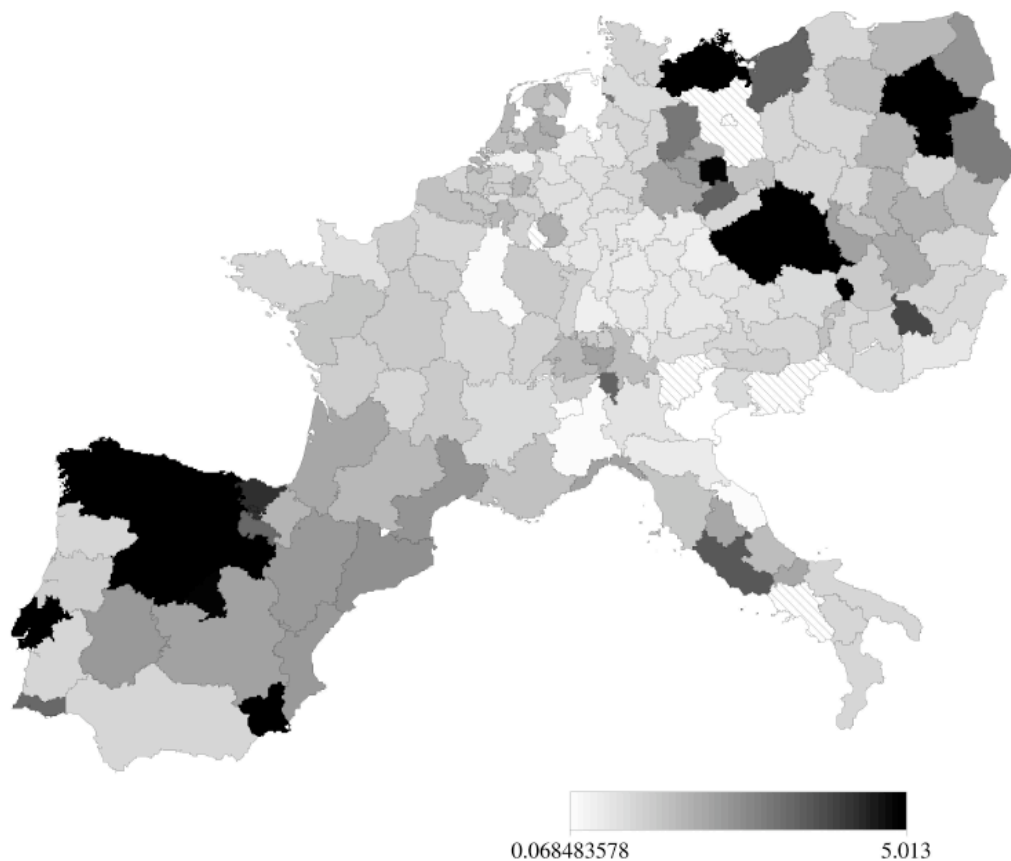
$$(7) \quad \bar{\lambda}_{\alpha,n}(x,y) = \sup\{\lambda \mid \bar{S}_{Y|X,n}(\lambda y \mid x) > 1 - \alpha\}$$

The order- α -frontier analysis is particularly charming for samples with extreme observations as they easily might occur for a sample of regions. Furthermore the approach accounts to some extent for the heterogeneity of the sample, as each region is only compared with regions whose input levels are equal or worse.

2.3 Results

The results, illustrated by figure 1, derive from the application of the order- α -frontier analysis for $\alpha = 0.05$.⁷ A region is considered comparatively inefficient in using its intramural R&D investments and human capital, if one or more other regions equipped with a similar or worse level of these inputs generate(s) higher levels of output.

Figure 1. Efficiency in generating knowledge, selected EU regions, efficiency scores 2002



Source: Own calculation, regions with missing observation marked by stripes

⁷ Efficiency scores have been calculated for $\alpha=0,01$, $\alpha=0,05$ and $\alpha=0,1$. See annex for detailed results.

Dark-grey shaded regions show rather low efficiency levels⁸ and can be considered comparatively disadvantaged with regard to their productivity in generating knowledge. On the contrary, the white- and light grey-shaded regions show the highest level of efficiency, as they deliver comparatively high patent applications per employee with the given inputs.

Northern Italian and southern German regions turn out to generate knowledge in a rather efficient way. In contrast northern Spanish and Czech regions' productivity is among the lowest. More general, the results point to the existence of regional spillovers, as neighboring regions' efficiency scores seem to be rather similar. At the same time, we also observe a (national) border effect. Thus, knowledge generation seems, at least to some extent, be determined by the national factors such as language or institutional settings (Keller 2002, Bottazzi and Peri 2003). The next step foresees to analyse the results in more detail.

3 Decomposition of regional efficiency

3.1 Knowledge spillovers

The efficiency scores of the order- α -frontier approach point to a region's ability to transform the existing human capital and the intramural R&D expenditures into (codified) technological knowledge. Following the logic of maximizing outputs, highly innovative firms will have a high preference for locating their R&D division in regions on (or at least close to) the frontier in the long run. Alternatively firms could opt for a place of location in a neighboring region. This would be reasonable, if the firms expect knowledge spillovers between the two regions. While there is broad consensus on the existence of such spillovers, the geographical scope of these spillovers is not clear. Bottazzi and Peri (2002, 2003) find that spillovers within European regions are significant up to 300 km of distance (and never significant beyond) but decrease as distance increase. The result still holds, if results are controlled for centrality. "This means that regions in the periphery with R&D intensive neighbors benefit from it just as central regions do" (Bottazzi and Peri 2003, p. 699).

Following these results, the incorporation of regional spillovers into the model gains in importance with decreasing size of the spatial units. Thus, knowledge generation at county (NUTS 3) or even commune (NUTS 4) level heavily relies on the knowledge generated in regions nearby. In contrast, the results for larger units such as member states (NUTS 0) or federal states (NUTS 1) are mainly driven by localized spillovers within the region rather than cross-border spillovers.

With regard to the medium-sized regions at NUTS 2 level, chosen for the presented study, we

⁸ Note that, due to the mathematical formulation of the model, high efficiency scores point to low efficiency levels and vice versa.

expect a balanced mix of intra-regional and external spillovers. Therefore, the main aim of this chapter is to decompose the overall efficiency into a spatial and a non-spatial part. For this purpose we apply a geoaddivitive model, which can be considered an extended kriging approach.

3.2 *From kriging approaches to geoaddivitive models*

The interest in spatial dependence and the availability of spatial prediction has first been of particular interest in the broader field of geology. In fact, various techniques for spatial analyses with a limited number of observations originated in this field. One of the most popular methodologies to identify a spatial pattern prevalent in a given set of data is the so-called kriging.⁹ This approach describes an interpolation technique for spatially dependent variables which is still used in the field of mining and soil research but nowadays goes beyond this field. Interest in and applications for the kriging approach include, for example, data intensive fields such as environmental monitoring or agriculture and forestry management (Tavares et al. 2008, Diodato and Ceccarelli 2004, Jost et al. 2005).

The need to process ever more complex data sets points to one of kriging's major drawbacks: Computational cost increase rapidly with a rising complexity of the application. As a counter measure Hartman and Hössjer (2008) propose to develop a kriging predictor based on Markov random fields, which substantially decrease computation time.

Another drawback of pure kriging is the missing control of covariates. This problem can be addressed if an extended version is applied. The so-called *universal kriging* allows for the control of covariates by incorporating them directly into the model (Cressie 1993). However, universal kriging is limited in a sense that covariate effects are presumed linear in nature. If instead non-linear relationships of the covariate with the response variable can be expected, Kammann and Wand (2003) propose a fusion of kriging and additive models to *geoaddivitive models*. Thus, geoaddivitive models provide a particularly suitable tool, if the response variable can be explained by a spatial and one or more covariates whose effects on the response variable are non-linear in nature.¹⁰ Recent applications of geoaddivitive models include the geographical variability of infants' health conditions in Massachusetts (Kammann and Wand 2003), gender-specific health status in Germany and forest health outcomes in Bavaria (Brezger and Lang 2006).

In the presented study, the geoaddivitive approach is used to decompose the regions' efficiency

⁹ See Trangmar et al. (1985) and Goovaerts (1999) for surveys on studies dealing with kriging.

¹⁰ In case the only interest is in the spatial distribution of a certain variable, both kriging approaches and geoaddivitive models can be applied.

(identified by using the order- α -frontier analysis) into a spatially driven and a non-spatial factor. The smoothed spatial factor shows the effect that is primarily determined by the general performance of a regions' greater surrounding area. The non-spatial structural factor gives an idea on a region's efficiency compared to the regions nearby. In so doing, an average performing region in the knowledge-intensive part of the considered economy surrounded by very efficient regions, is marked less efficient than an equally performing region surrounded by rather inefficient regions.

3.3 Methodological remarks

The basic principle of the approach foresees to smooth the observed data, which yields decreasing deviations of the variables assigned to neighboring units. The difference between observed and smoothed data than identifies the non-spatial factor driven by the units' structure rather than their location in space.¹¹

The model set up starts with the definition of neighborhood. While distance between points might serve as a good indicator for neighborhood in the continuous case, the indicator is more problematic for analyzing regions, where localization is discrete (Brezger 2004). Instead, two regions r and s are defined as neighbors $r \sim s$, if they share a border. The smoothing algorithm, which is weighted by the length of the common border (ω_{rs}), presupposes a growing regional interdependence with increasing length of ω_{rs} .

The mathematical formulation of the model follows the principles of structured additive regression analysis and therefore aims at substituting a usual parametric by a flexible non-parametric parameter, containing in this case spatial information (Fahrmeir et al. 2001, Hastie and Tibshirani 1990). Since the presented study concentrates on the spatial distribution of the efficiency as the response variable, no parametric covariables are considered. Therefore, the nonparametric regression model can be defined as (Fahrmeir and Lang 2001):

$$(8) \quad \bar{\lambda}_{\alpha,n}(x_i, y_i) = f_{geo(i)} + f_{rand(i)}$$

where $\bar{\lambda}_{\alpha,n}(x_i, y_i)$ is the empirical efficiency score of region i (defined by equation (7)) and $f_{geo(i)}$ the spatially smoothed factor of region $i = 1, \dots, n$. The remaining term $f_{rand(i)}$, generally considered the normally distributed error e_i , is interpreted as structural factor that cannot be explained by the spatial correlation.¹²

In order to smooth the regions' spatial factor a penalizing term, based on the least square

¹¹ A more detailed introduction into the field of geoaddivitive regression analysis is given by Fahrmeir et al. (2007).

¹² Presuming that the observed data are correct, we consider the remaining error as irrelevant.

method (PLS), is introduced in equation (9). In this preliminary step, the weights ω_{rs} remain unconsidered.

$$(9) \quad PLS(\mu) = \sum_{i=1}^n (y_i - f_{geo(s_i)})^2 + \mu \sum_{s=2}^d \sum_{r \in N(s), r < s} (f_{geo(r)} - f_{geo(s)})^2$$

where $N(s)$ is the set of neighbors surrounding region s and μ is a parameter to control the smoothing intensity. The first term sums up the squared differences of observed data and modeled spatial factor. The smoothing process, defined in the second term, multiplies μ with the cumulated squared differences of spatial factors for all neighborhood relations.

In line with Fahrmeir et al. (2007) the penalizing approach, which includes the minimization of $PLS(\mu)$, can be interpreted in Bayesian way. This, in turn, yields markov random fields. As a consequence, the application of the model follows a Bayesian approach (with fully Bayesian inference) and is simulated by markov-chain-monte-carlo (MCMC) technique.¹³

In a last step, the expected value γ of the nonparametric spatial factor $f_{geo(i)}$ is defined as the average of the expected values of neighboring regions. Given the distribution of γ_r for all neighbors and introducing the weights ω_{rs} the conditional distribution of the expected spatial factor of region s (γ_s) is defined normally distributed as:

$$(8) \quad \gamma_s | \gamma_r \sim N\left(\frac{\omega_{sr}}{\omega_{s+}} \sum_{r \in \delta_s} \gamma_r, \frac{\tau^2}{\omega_{s+}}\right)$$

where δ_s is the set of neighbors of region s and ω_{sr} the weight of neighbor r . ω_{s+} denotes the cumulated weights of all regions neighboring s . The variance parameter τ^2 controls the level of variation between the model result and the expected value.¹⁴

The remaining non-spatial factor is considered normally distributed as well and can be defined as $f_{rand(s_i)} \sim N(0, \tau^2)$.

Based on the conditional expectation $\gamma_s | \gamma_r$ defined by equation (8), the MCMC-simulation results in a common distribution for the vector $\gamma = (\gamma_1, \gamma_2, \dots, \gamma_n)$. Thus, the estimated spatial factor can be identified for all regions in the last step.

3.4 Results

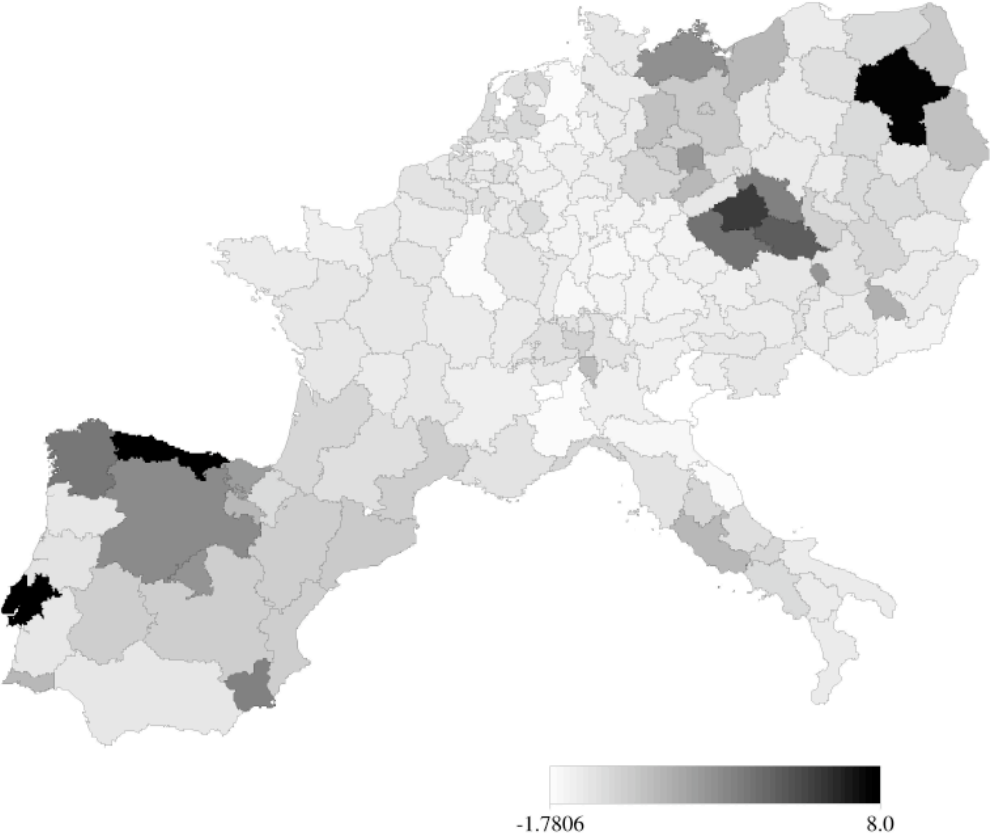
Figure 2 and 3 illustrate the regions' spatial and non-spatial efficiency if the decomposition

¹³ For this purpose the software BayesX has been applied. The algorithm and its computation is described in detail by Brezger et al. (2009).

¹⁴ Note that the model behind equation (8) is equivalent to the penalizing model defined by equation (7) (see Fahrmeir et al. 2007, p. 390)

follows the model outlined in chapter 3.3. Again white and light-grey shaded regions point to high efficiency, whereas dark grey shaded regions identify relatively low productivity in generating knowledge.

Figure 2. Efficiency of selected regions, smoothed spatial factor, 2002



Source: Own calculation.

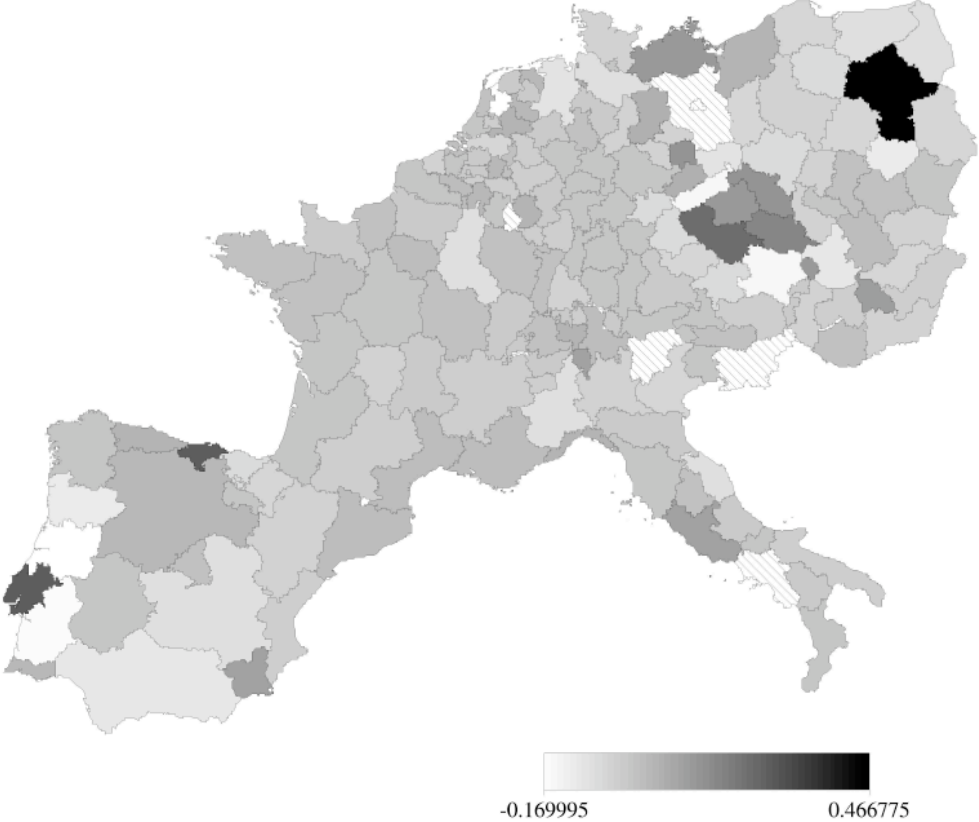
As for figure 1 a small score, which could indeed be negative, points to high efficiency and vice versa. Based on the performance of the surrounding regions, the algorithm estimates the spatial efficiency for all regions of the sample. However, despite the smoothing process, some regions' efficiency still differs significantly from the neighbors' performance. Such an effect occurs, if the regions' overall efficiency, identified in the first step, yields extreme results. In addition the smoothing relies not only on the direct neighbors but accounts for trends in the greater area as well. Lisboa, for example, shows a rather low overall efficiency¹⁵ and furthermore the smoothing process is partly influenced by the weak performance of regions in northern Spain.

¹⁵ In absolute terms, the region of Lisboa is characterized by very high input but only average output levels which, in turn, explains the modest performance in terms of efficiency.

On the one hand, the findings of the geoaddivitive regression show that regions in the centre of the considered sample (Eastern France, Benelux, Western Germany, Northern Italy) seem to benefit from their location in space while Eastern and South Western regions have a comparative disadvantage. On the other hand, results are not clear in any case. In particular results for generally disadvantaged areas are quite heterogeneous. This can partly be explained by the definition of efficiency which identifies regions with very low input levels as comparatively efficient even if they generate little knowledge in absolute terms (as it is the case for some Polish, Hungarian and southern Spanish regions).

Besides the spatial factor, the regions' efficiency is also driven by a non-spatial intraregional effect. This effect, shown by figure 3, can be interpreted as structural effect.

Figure 3. Efficiency of selected regions, non-spatial factor, 2002



Source: Own calculation, regions with missing observation marked by stripes

According to figure 3, a dark grey-shaded region *k* is considered comparatively inefficient in generating knowledge, if regions *in the greater neighboring area of region k* equipped with a similar or worse level of human and infrastructure capital generate(s) higher levels of per-capita-income. By contrast, white or light grey-shaded regions are efficient compared to the

nearby regions. The rather homogeneous picture in the sample's centre shows that performance is rather similar. The higher efficiency of some German regions results from their neighborhood to weak performing czech regions. The picture is more heterogenous for the Iberian Peninsula and Eastern Europe. Since the absolute level of patent application differs significantly in these regions, a general conclusion cannot be drawn at this stage. Some regions in Eastern Europe, however, show similar absolute input and output levels. In this case, the findings could help to identify appropriate funding measures in the future. While relatively efficient regions (in terms of non-spatial efficiency) might indeed benefit by further increasing R&D expenditures, most likely this strategy is less effective for relatively inefficient regions.

4 Conclusions

Taking into account the regions' human capital and intramural R&D expenditures, the presented paper calculates, in a first step, the regional efficiency by using the order- α -frontier analysis. The approach, which can be considered an outlier-robust extension of the DEA, identifies rather high efficiency for the core of continental Europe and generally low efficiency for Eastern Europe and the Iberian Peninsula.

The results indicate the existence of a spatial effect driven by knowledge spillovers. Therefore, efficiency is further decomposed into a spatial and a non-spatial part by applying a geoaddivitive approach. The results confirm the existence of the spatial effect. While regions in Eastern France, the Benelux, Western Germany and Northern Italy benefit from their location in space, Eastern European and Spanish regions seem to have a comparative disadvantage in generating knowledge.

A policy devoted to increase the competitiveness of EU regions by enhancing the regions' ability to generate knowledge could either follow the efficiency argument, according to which return on public investments should be maximized. In this case public R&D expenditures should particularly go to already efficient regions with high spillover potentials. If, however, the focus is on the equity argument, two alternative strategies arise. First, the distribution of R&D expenditures could again be based on the degree of efficiency, which this time yields a funding of the relatively inefficient, spatially disadvantaged regions (Athanasopoulos 1996, Camagni 1990). In this case the effectiveness of funding could still be increased, if the (spatially disadvantaged) regions' structural efficiency would be taken into account as well.

Spatially inefficient regions with relatively high scores compared to their neighbors would benefit more from additional R&D expenditures compared to their spatially and structurally inefficient neighbors. The latter might benefit more by other funding, which could help to

foster structural change towards a more knowledge based structure. Alternatively, the region might for some reasons not be in a position to turn into a knowledge intensive economy but shows other strengths worth to be further developed. Clearly, the decision on appropriate funding patterns in terms of public R&D cannot be based on regional efficiency alone. Instead the absolute dimensions and the national institutional framework must be taken into account as well. Vice versa, a funding scheme that does not account for the regions' productivity level, might easily fall short as well.

References

- Aigner, D., Lovell, C., A., K., Schmidt, P. (1977) Formulation and estimation of stochastic frontier production function models, *Journal of Econometrics* 6/4: 21-37.
- Audretsch, D. (1995), *Innovation and Industry Evolution*. Cambridge, MIT Press.
- Aragon, Y., Daouia, A., Thomas-Agnan, C. (2005) Nonparametric frontier estimation: a conditional quantile-based approach, *Econometric Theory* 21: 358-389.
- Athanassopoulos, A. D., (1996), Assessing the comparative spatial disadvantage (CSD) of regions in the European Union using non-radial data envelopment analysis methods, *Journal of Operational Research* 94: 439-452.
- Barro, R. J., Sala-i-Martin, X. (1997), Technological Diffusion, Convergence and Growth, *Journal of Economic Growth*, 2 (1), 1-27.
- Bottazi, L., Peri, G. (2002), Innovation and spillovers in regions: Evidence from European Patent Data, IGIER Bocconi University Working Paper # 215, Milan.
- Bottazi, L., Peri, G. (2003), Innovation and spillovers in regions: Evidence from European patent data, *European Economic Review* 47/4: 687-710.
- Boussofiane, A., Dyson, G. R., Thanassoulis, E. (1991), Applied Data Envelopment Analysis, *European Journal of Operations Research* 52: 1-15.
- Brezger, A. (2004), *Bayesian P-Splines in Structured Additive Regression Models*, Dissertation, Ludwig-Maximilians-Universität: Munich
- Brezger, A., Lang, S. (2006), Generalized structured additive regression based on Bayesian P-splines, *Computational Statistics & Data Analysis*, 50/4: 967-991.
- Brezger, A., Kneib, T., Lang, S. (2009), BayesX- Software for Bayesian inference in structured additive regression models, *Methodology*,
<http://www.stat.uni-muenchen.de/~bayesx/bayesx.html>
- Cazals, C., Florens, J. P., Simar, L. (2002), *Nonparametric frontier estimation: a robust approach*, *Journal of Econometrics* 106: 1-25.

- Charnes, A., Cooper, W. W., Li, S. (1989), Using data envelopment analysis to evaluate efficiency in the economic performance of Chinese cities, *Socio-economic Planning Science* 23: 325-344.
- Charnes, A., Cooper, W. W., Lewin, A. Y., Seiford, L. M., (Eds.), (1994), *Data Envelopment Analysis: Theory, Methodology, and Applications*, Kluwer Academic Publishers: Norwell.
- Cooper, W. W., Seiford, L. M., Tone, K. (2006), *Introduction to Data Envelopment Analysis and Its Uses: With DEA-solver Software and References*, Springer: Berlin.
- Cressie, N. (1993), *Statistics for Spatial Data*, New York: Wiley.
- Daouia, A., Simar, L. (2005), Robust nonparametric estimators of monotone boundaries, *Journal of Multivariate Analysis* 96: 311-331.
- Deprins, D., Simar, L., Tulkens, H. (1984), Measuring labor inefficiencies in post offices, in: Marchand, M., Pestieau, P., Tulkens, H. (Eds.) *The Performance of Public Enterprises: Concepts and Measurements*, Elsevier: Amsterdam, 243-267.
- Diodato, N., Ceccarelli, M. (2004), Multivariate indicator Kriging next term approach using a GIS to classify soil degradation for Mediterranean agricultural lands, *Ecological Indicators* 4/3: 177-187.
- Fahrmeir, L., Lang, S. (2001), Bayesian inference for generalized additive mixed models based on Markov random field priors, *Journal of the Royal Statistical Society C (Applied Statistics)* 50, 201–220
- Fahrmeier, L., Tutz, G., Hennevogl, W. (2001). *Multivariate statistical modelling based on generalized linear models*, Springer: New York.
- Fahrmeir, L., Kneib, T., Lang, S. (2007), *Regression – Modelle, Methoden und Anwendungen*, Springer: Berlin.
- Goovaerts, P. (1999), Geostatistics in soil science: state-of-the-art and perspectives, *Geoderma* 89: 1–45.
- Griliches, Z. (1990), Patent statistics as economic indicators: A survey. *Journal of Economic Literature*, 27: 1661-1707.
- Grossman, G., Helpman, H. (1991) *Innovation and Growth in the Global Economy*, MIT Press, Cambridge, Mass.
- Grupp, H. and Schmoch, U. (1999), Patent statistics as economic indicators. *Research Policy*, 28: 377-396.
- Guan, J., Chen, K. (2010), Measuring the innovation production process: A cross-region empirical study of China's high-tech innovations, *Technovation* 30/5-6: 348-358.
- Hartman, L., Hössjer, O., (2008) Fast kriging of large data sets with Gaussian Markov random

- fields, *Computational Statistics & Data Analysis* 52/5: 2331–2349.
- Hashimoto, A., Ishikawa, H. (1993), Using DEA to evaluate the state of society as measured by multiple social indicators, *Socio-Economic Planning Sciences* 27: 257-268.
- Hastie T., Tibshirani R. (1990), *Generalized Additive Models*, Chapman and Hall: London.
- Jensen, M. B., Johnson, B., Lorenz, E., Lundvall, B. A. (2007), Forms of knowledge and modes of innovation, *Research Policy*, 36, 680-693.
- Jost, G., Heuvelink, G. B. M., Papritz, A. (2005), Analysing the space–time distribution of soil water storage of a forestnext term ecosystem using spatio-temporal previous termkrigingnext term, *Geoderma* 128/3-4: 258-273.
- Kammann, E. E., Wand, M. P. (2003), Geoaddivitive models, *Applied Statistics* 52/1: 1-18.
- Keller, W. (2002), Geographic Localization of International Technology Diffusion, *American Economic Review*, 92(1): 120–142.
- Li, X. (2009), China’s regional innovation capacity in transition: an empirical approach, *Research Policy* 38/2: 338-357.
- Los, B., Verspagen, B. (2006), The evolution of productivity gaps and specialization patterns, *Metroeconomica* 57 (4), 464–493.
- Macmillan, W. (1986), The estimation and application of multi-regional economic planning models using data envelopment analysis, *Journal of Regional Science* 60: 41-57.
- Pavitt, K. (1985): Patent statistics as indicators of inventive activities: possibilities and problems. *Scientometrics* 7: 77-99.
- Rammer, C. (2003): Patente und Marken als Schutzmechanismen für Innovationen, Studie zum deutschen Innovationssystem. Nr. 11-2003, Zentrum für Europäische Wirtschaftsforschung, Mannheim.
- Romer, P. M. (1990), Endogenous Technological Change, *Journal of Political Economy*, 98, 71-102.
- Tavares, M. T., Sousa, A. J., Abreu, M. M. (2008), Ordinary kriging and indicator kriging in the cartography of trace elements contamination in São Domingos mining site, *Journal of Geochemical Exploration* 98/1-2: 43-56.
- Trangmar, B. B., Yost, R. S., Uehara, G. (1985), Application of geostatistics to spatial studies of soil properties. *Advances in Agronomy* 38, 45-94.
- Wang, E. C., Huang, W. (2007), Relative efficiency of R&D activities: A cross-country study accounting for environmental factors in the DEA approach, *Research Policy* 36: 206-273.
- Zhang, A., Zhang, Y., Zhao, R. (2003), A study of the R&D efficiency and productivity of Chinese firms, *Journal of Comparative Economics*, 31/5-6: 444-464.

ANNEX

Table 1. Efficiency scores of order- α -frontier analysis (high scores point to low efficiency).

Code	Region	alpha=0.9	alpha=0.95	alpha=0.99
AT11	Burgenland	1,000	1,480	2,461
AT12	Niederösterreich	0,629	0,917	1,330
AT13	Wien	1,255	1,872	2,839
AT21	Kärnten	0,983	1,233	2,089
AT22	Steiermark	0,734	0,970	1,828
AT31	Oberösterreich	0,572	0,796	1,219
AT32	Salzburg	0,780	0,977	1,264
AT33	Tirol	0,660	0,819	1,140
AT34	Vorarlberg	0,330	0,526	1,000
BE10	Région de Bruxelles-Capitale	1,309	1,558	2,570
BE21	Prov. Antwerpen	0,849	1,077	1,600
BE22	Prov. Limburg (B)	1,420	1,759	2,267
BE23	Prov. Oost-Vlaanderen	1,147	1,365	2,252
BE24	Prov. Vlaams Brabant	0,966	1,180	1,752
BE25	Prov. West-Vlaanderen	0,729	1,000	1,533
BE31	Prov. Brabant Wallon	0,581	0,822	1,257
BE32	Prov. Hainaut	1,134	1,422	2,325
BE33	Prov. Liège	0,897	1,062	1,649
BE34	Prov. Luxembourg (B)	0,861	1,083	2,050
BE35	Prov. Namur	1,436	1,716	2,501
CH01	Région Lémanique	0,812	1,196	1,729
CH02	Espace Mittelland	1,027	1,652	2,568
CH03	Nordwestschweiz	0,897	1,336	2,026
CH04	Zürich	0,784	1,165	1,668
CH05	Ostschweiz	1,074	1,553	2,080
CH06	Zentralschweiz	1,500	2,197	3,848
CH07	Ticino	2,659	3,373	5,008
CZ01	Praha	6,609	8,289	12,163
CZ02	Střední Čechy	6,107	8,321	15,301
CZ03	Jihozápad	1,142	6,619	11,008
CZ04	Severozápad	1,000	1,000	1,000
CZ05	Severovýchod	1,517	5,829	8,193
CZ06	Jihovýchod	5,192	7,298	13,821
CZ07	Střední Morava	1,389	2,147	11,600
CZ08	Moravskoslezsko	1,165	1,496	1,704
DE11	Stuttgart	0,293	0,484	0,752
DE12	Karlsruhe	0,385	0,636	1,000
DE13	Freiburg	0,384	0,446	0,755
DE14	Tübingen	0,398	0,584	1,000
DE21	Oberbayern	0,408	0,624	1,000
DE22	Niederbayern	0,516	0,701	1,210
DE23	Oberpfalz	0,364	0,466	1,000
DE24	Oberfranken	0,417	0,542	1,000
DE25	Mittelfranken	0,375	0,571	0,873
DE26	Unterfranken	0,405	0,556	1,000
DE27	Schwaben	0,512	0,606	1,025
DE30	Berlin	1,313	1,954	2,797
DE41	Brandenburg - Nordost	2,078	2,484	3,620
DE42	Brandenburg - Südwest	1,736	2,052	2,506
DE50	Bremen	2,613	3,033	5,132
DE60	Hamburg	0,740	0,928	1,133
DE71	Darmstadt	0,521	0,706	1,000
DE72	Gießen	0,699	0,819	1,199
DE73	Kassel	0,753	1,000	1,252
DE80	Mecklenburg-Vorpommern	3,321	5,294	6,762
DE91	Braunschweig	0,709	0,968	1,284
DE92	Hannover	0,790	1,000	1,557
DE93	Lüneburg	0,564	0,938	1,000
DE94	Weser-Ems	0,074	0,080	1,000

Code	Region	alpha=0.9	alpha=0.95	alpha=0.99
DEA1	Düsseldorf	0,583	0,682	1,000
DEA2	Köln	0,549	0,784	1,200
DEA3	Münster	0,387	0,514	1,000
DEA4	Detmold	0,491	0,701	1,000
DEA5	Arnsberg	0,509	0,797	1,000
DEB1	Koblenz	0,423	0,689	1,000
DEB2	Trier	1,686	1,909	2,782
DEB3	Rhein Hessen-Pfalz	0,432	0,586	0,969
DEC0	Saarland	0,432	0,629	1,028
DED1	Chemnitz	2,097	3,430	6,025
DED2	Dresden	1,122	1,460	2,152
DED3	Leipzig	4,200	4,875	8,249
DEE1	Dessau	1,000	2,101	3,495
DEE2	Halle	1,810	2,245	3,956
DEE3	Magdeburg	2,469	3,007	5,192
DEF0	Schleswig-Holstein	0,901	1,117	1,439
DEG0	Thüringen	1,783	2,083	3,057
DK00	Danmark	1,024	1,296	1,928
ES11	Galicia	4,200	6,039	10,042
ES12	Principado de Asturias	6,841	10,124	16,835
ES13	Cantabria	10,558	15,656	17,558
ES21	Pais Vasco	3,221	4,287	5,199
ES22	Comunidad Foral de Navarra	1,591	1,800	2,624
ES23	La Rioja	1,000	3,361	5,590
ES24	Aragón	1,795	2,331	3,015
ES30	Comunidad de Madrid	4,212	4,858	6,939
ES41	Castilla y León	4,218	5,293	8,647
ES42	Castilla-la Mancha	1,620	2,177	2,177
ES43	Extremadura	1,861	2,389	2,389
ES51	Cataluña	1,605	2,557	3,507
ES52	Comunidad Valenciana	1,629	2,341	3,894
ES53	Illes Balears	1,320	1,320	1,320
ES61	Andalucia	0,713	1,000	5,702
ES62	Región de Murcia	1,000	5,795	9,636
FR10	Île de France	0,821	1,223	1,855
FR21	Champagne-Ardenne	0,110	0,125	1,000
FR22	Picardie	0,916	1,000	1,752
FR23	Haute-Normandie	0,890	1,155	2,134
FR24	Centre	1,000	1,302	2,403
FR25	Basse-Normandie	0,206	0,790	1,110
FR26	Bourgogne	0,712	1,000	1,727
FR30	Nord - Pas-de-Calais	1,000	1,502	2,741
FR41	Lorraine	1,034	1,258	1,814
FR42	Alsace	0,725	1,057	1,856
FR43	Franche-Comté	0,766	1,093	1,675
FR51	Pays de la Loire	1,055	1,350	2,558
FR52	Bretagne	0,849	1,053	1,855
FR53	Poitou-Charentes	1,000	1,233	2,050
FR61	Aquitaine	1,551	2,021	3,490
FR62	Midi-Pyrénées	1,176	1,636	3,082
FR63	Limousin	0,208	1,000	1,841
FR71	Rhône-Alpes	0,692	0,900	1,327
FR72	Auvergne	1,000	1,272	1,948
FR81	Languedoc-Roussillon	1,804	2,474	4,452
FR82	Provence-Alpes-Côte d'Azur	1,278	1,519	1,924
FR83	Corse	0,839	0,927	1,000
HU10	Közép-Magyarország	3,242	3,923	4,967
HU21	Közép-Dunántúl	1,000	1,024	1,571
HU22	Nyugat-Dunántúl	1,000	1,000	1,000
HU23	Dél-Dunántúl	0,684	0,878	1,000
HU31	Észak-Magyarország	1,000	1,000	1,000
HU32	Észak-Alföld	0,691	1,000	1,060
HU33	Dél-Alföld	0,571	0,647	1,000
ITC1	Piemonte	0,097	0,136	1,000

Code	Region	alpha=0.9	alpha=0.95	alpha=0.99
ITC2	Valle d'Aosta/Vallée d'Aoste	0,106	0,120	1,000
ITC3	Liguria	1,903	2,131	3,106
ITC4	Lombardia	0,572	0,743	1,083
ITD1	Provincia Autonoma Bolzano-Bozen	0,250	0,961	1,351
ITD2	Provincia Autonoma Trento	0,879	1,070	1,543
ITD3	Veneto	0,060	0,068	1,000
ITD4	Friuli-Venezia Giulia	0,712	0,911	1,184
ITD5	Emilia-Romagna	0,376	0,528	1,000
ITE1	Toscana	1,000	1,358	2,345
ITE2	Umbria	1,656	2,078	3,395
ITE3	Marche	0,100	0,114	1,000
ITE4	Lazio	2,793	3,589	5,181
ITF1	Abruzzo	1,119	1,608	2,675
ITF2	Molise	1,542	2,072	2,072
ITF3	Campania	0,798	0,904	1,000
ITF4	Puglia	0,652	1,000	1,000
ITF5	Basilicata	0,973	1,031	1,594
ITF6	Calabria	0,674	1,000	1,000
ITG1	Sicilia	0,793	1,000	1,225
ITG2	Sardegna	1,000	1,000	1,000
LU00	Luxembourg (Grand-Duché)	1,041	1,219	1,785
NL11	Groningen	1,608	2,016	2,959
NL12	Friesland	1,457	1,628	2,660
NL13	Drenthe	0,664	0,982	1,634
NL21	Overijssel	1,388	2,002	3,393
NL22	Gelderland	1,503	1,920	3,037
NL23	Flevoland	2,910	3,378	4,949
NL31	Utrecht	1,138	1,312	1,875
NL32	Noord-Holland	1,427	1,789	2,626
NL33	Zuid-Holland	1,229	1,542	2,262
NL34	Zeeland	1,298	1,825	2,369
NL41	Noord-Brabant	0,261	0,382	0,669
NL42	Limburg (NL)	1,000	1,161	1,700
PL11	Lódzkie	1,411	1,811	2,779
PL12	Mazowieckie	8,193	10,639	19,647
PL21	Malopolskie	1,000	1,894	1,894
PL22	Slaskie	1,347	1,778	1,778
PL31	Lubelskie	2,923	2,923	2,923
PL32	Podkarpackie	1,583	1,583	1,583
PL33	Swietokrzyskie	1,000	1,000	1,000
PL34	Podlaskie	2,452	2,452	2,452
PL41	Wielkopolskie	1,000	1,000	1,000
PL42	Zachodniopomorskie	3,459	3,459	3,459
PL43	Lubuskie	1,000	1,000	1,000
PL51	Dolnoslaskie	0,567	1,000	1,000
PL52	Opolskie	1,000	1,000	1,000
PL61	Kujawsko-Pomorskie	1,551	1,551	1,551
PL62	Warminsko-Mazurskie	1,643	1,643	1,643
PL63	Pomorskie	1,000	1,000	1,000
PT11	Norte	1,000	1,000	1,000
PT15	Algarve	3,345	3,345	3,345
PT16	Centro (PT)	1,183	1,183	1,183
PT17	Lisboa	12,139	14,965	24,887
PT18	Alentejo	1,000	1,000	1,000
SI00	Slovenija	2,298	3,279	5,024
SK01	Bratislavský	3,948	5,060	7,677
SK02	Západné Slovensko	1,031	1,323	1,323
SK03	Stredné Slovensko	1,488	2,000	2,000
SK04	Východné Slovensko	1,000	1,000	1,000