Papers in Evolutionary Economic Geography

# 08.04

**From Average to the Frontier:**
**A Nonparametric Frontier Approach to the**
**Analysis of Externalities and Regional Innovation Performance:**

Tom Broekel

Utrecht University
Urban & Regional research centre Utrecht

http://econ.geo.uu.nl/peeg/peeg.html

# FROM AVERAGE TO THE FRONTIER: A NONPARAMETRIC FRONTIER APPROACH TO THE ANALYSIS OF EXTERNALITIES AND REGIONAL INNOVATION PERFORMANCE

Tom Broekel §

Department of Economic Geography,
Faculty of Geosciences, Utrecht University, The Netherlands

12th March 2008

**Abstract**

Although a rich literature has emerged analyzing the impact of localization, urbanization, and Jacobs externalities on regional innovativeness, the findings are still contradictory. Traditional studies differ mainly in the employed data but rely on similar empirical approaches. This paper argues in favor of using in this context production frontier approaches instead of the commonly employed production function approaches. In addition, a nonparametric frontier approach is used to empirically examine the influence of the externalities on regions' innovativeness. For four different industries positive effect of localization and urbanization externalities are found. In contrast, with the exception of the transport equipment industry, Jacobs externalities seem to be of minor importance.

**JEL codes:** R12, O18, O31

**Keywords:** regional innovation performance, nonparametric frontier analysis, German electrics and electronics industry

§ Department of Economic Geography, Urban & Regional Research Centre Utrecht (URU), Faculty of Geosciences, Utrecht University, Heidelberglaan 2, 3584 CS Utrecht, The Netherlands. *Phone*: +31 30 253 3206. *Fax*: +31 30 253 2037. *E-mail*: t.broekel@geo.uu.nl

* I greatly thank Thomas Brenner for helpful comments and suggestions, and Cinzia Daraio for her assistance in the implementation of the conditional order-m analysis.

# 1 Introduction

Following the ideas of Marshall (1890); Arrow (1962); Romer (1986) and Jacobs (1969) researchers have tried to provide empirical evidence for the importance of localization (Marshall-Arrow-Romer externalities) and Jacobs externalities for regional economic activities. In addition, Hoover (1937, 1948), and Isard (1956) argue in favor of a third externality, namely urbanization, which is also said to take effect on regional economies' development. Besides investigating the relationship between the externalities and regional economic growth (e.g., Glaeser et al., 1992; Henderson et al., 1995) researchers also analyzed the externalities with respect to their influence on regional innovation activities (e.g., Feldman and Audretsch, 1999; Paci and Usai, 1999b).

While the theoretical arguments for positive effects on regional innovativeness of localization, and urbanization, as well as Jacobs externalities are straightforward the empirical evidence is mixed (see for a meta-study de Groot et al., 2007). For example, positive effects of localization externalities are confirmed by van der Panne and van Beers (2006). They find however no positive effects of Jacobs externalities. In contrast, the result of Feldman and Audretsch (1999) indicate that localization externalities seem to be of minor importance. Jacobs externalities are however identified as being crucial.

In order to investigate the externalities' influence on regional innovativeness the majority of the empirical studies relies on production function approaches. Recently, Bonaccorsi and Daraio (2006) argued however that nonparametric production frontier approaches are more appropriate for analyzing R&D systems because strong assumptions that are inherent to the production function approaches are relaxed.

The aim of the present paper is twofold. First, a discussion on the differences between both approaches in the context of the analysis of regional innovativeness is provided. From this, a number of arguments are made in favor of using nonparametric production frontier approaches for analyzing externalities and their impact on regional innovativeness instead of the commonly employed parametric production function approaches.

Second, the paper demonstrates the usefulness of robust nonparametric production frontier approaches by analyzing the impact of externalities on regional innovation processes. It is shown that although this approach requires less assumptions to be made, the obtained results are not less informative.

For four industries that are investigated separately, it is found that localization, and urbanization externalities positively influence industries' regional innovation performance. The evidence for the existence of Jacobs externalities is comparatively weaker and thus they seem to be of smaller importance. With respect to the strength and shape of these influences significant differences are observed between industries. Moreover, in particular the impact of localization externalities is characterized by turning points and non-linearities.

The paper is organized as follows. In Section 2 the existing literature on the externalities' impact on regional innovation processes is briefly reviewed. The theoretical discussion on the usefulness of the two different empirical approaches in this context is subject to this section as well. The specific *robust* nonparametric production frontier approach used for analyzing the effects of the externalities on regional innovation performance is presented in Section 3. Section 4 provides the description of the data used in the empirical assessment. This is followed by a brief presentation and discussion of the findings in Section 5. Section 6 concludes.

# 2 Two ways of analyzing regional innovativeness

## 2.1 Localization, urbanization, and Jacobs externalities

According to Marshall (1890) firms profit from the agglomeration of other firms active in the same industry and region. These benefits show as an increase in the availability of a skilled labor force and the possibility to easily exchange necessary goods. Geographic proximity additionally is argued to foster the potential of knowledge (innovations) to spill over. Hence, as more relevant actors are located in a region, the potential to benefit from spillovers increases accordingly.

Using the terminology of Henderson et al. (1995) I may refer to these externalities as *localization externalities*. They represent the impact on an industry's innovation activities that stems from its agglomeration in a region. The resulting advantages are purely industry specific in nature. In the context of this paper the potential of intra-industry knowledge spillovers is most likely the major source of these benefits.

In contrast to localization externalities two other forms of externalities have been put forward in the literature: urbanization externalities and Jacobs externalities. In contrast to localization externalities *urbanization externalities* differ in that they are not industry-specific. They are caused by the agglomeration of general economic activities, and are argued to yield benefits to all (or a number of) other industries' firms in a region (Hoover, 1937, 1948; Isard, 1956). Among other, these benefits can show as rich local labor markets, well developed infrastructure, strong local demand, and the presence of specific public facilities (research institutes, universities, etc.). That is, all factors that enhance regional firms' economic activities and that tend to go along with increasing degrees of urbanization.

In addition to urbanization externalities, Jacobs (1969) emphasized that the variety of the regional economy can foster firms' growth and innovativeness. Similarly as in the case of localization externalities the positive effects of Jacobs externalities on innovation activities result from firms' possibility to absorb knowledge spillovers. However, not intra-industry spillovers are argued to be crucial, but inter-industry ones. In this respect firms' innovativeness is stimulated by interacting with other industries' actors characterized by complementary knowledge, technologies, and skills. Therefore, actors may profit from being located in regions which are characterized by a diverse but complementary knowledge structure.

While the arguments for the positive effects of localization, urbanization, as well as Jacobs externalities are straightforward, and there exists a great number of studies addressing this issues, the empirical evidence is still mixed. For example, using data from the Netherlands, van der Panne and van Beers (2006) find that firms located in specialized regions tend to show higher levels of innovations, i.e. localization externalities foster innovativeness. In contrast, Paci and Usai (1999b,a) and Greunz (2004) employing Italian and European data respectively, report positive findings for specialization as well as diversification on regional innovativeness with the latter being more relevant in case of high-tech industries and densely populated regions. An even stronger support for the importance of diversification is provided by Feldman and Audretsch (1999) using U.S. data. However, contrasting the other studies' findings, their results "indicate that diversity across complementary economic activities sharing a common science base is more conducive to innovation than is specialization" (Feldman and Audretsch, 1999, p. 427). This is confirmed for manufacturing industries by van Oort (2002) and a data set on the Netherlands.

Except for the use of different data bases and approximation variables most of the studies are similar in the way in which the relationships between regional factors (e.g., R&D efforts, externalities) and the outcome measures of innovation activities are modeled. For example, Feldman and Audretsch (1999) use a Poisson regression which is a special from of the Generalized Least Square approach; Paci and Usai (1999a) rely on an Ordinary Least Square approach that accounts for dependencies between regions; and van der Panne and van Beers (2006) employ a negative binomial regression approach. Although these are different types of regression analyses, in the way they have been employed in these studies all of them represent kinds of *production function* approaches.

In the following section, the production function approach is presented in more detail.[1] In addition, an alternative way of analyzing regional innovativeness, namely the *production frontier* approach, is presented. It is argued that this alternative approach is more appropriate for investigating this issue and hence, may lead to more coherent results on the impact of externalities on regional innovation processes in the future.

## 2.2 Parametric production functions vs. nonparametric frontiers

In the following the production function approach and the production frontier approach are presented in more detail. Furthermore, they are compared with respect to their assumptions, use, and results. Bonaccorsi and Daraio (2006) provide already an extensive comparison of the two approaches and discuss their appropriateness for the evaluation of scientific and technological systems. In contrast to this rather general and technical work, the present paper aims at motivating the use of nonparametric production frontier approaches particularly for analyzing *regional* innovation activities.

This is achieved by restricting the discussion to the two most interesting approaches. First, the parametric production function approach (PPF) is introduced. It represents the most common types of analyses found in empirical studies investigating quantitatively regional innovativeness. Second, the

[1] Performance and production frontier approach are used interchangeably in this paper. However, the latter takes reference to a specific type of statistical technique while the first covers a general concept.

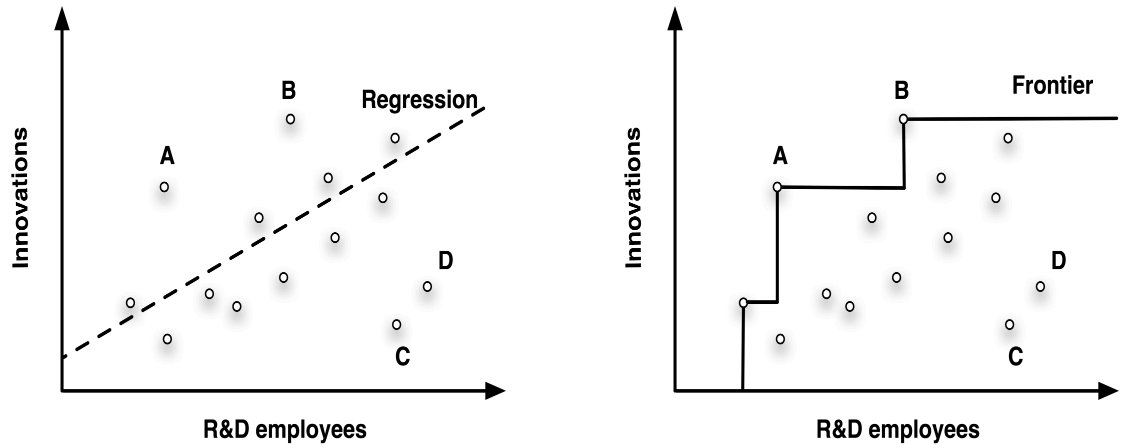
more "exotic" approach, the non-convex and nonparametric production function approach (NNPF) is presented which has been argued to be a promising alternative by Broekel and Brenner (2007).

In the parametric production function approach (PPF) as well as in the non-convex nonparametric production frontier approach (NNPF) researchers start from the idea that innovations, in particular industrial innovations, do not "fall from heaven" but that it needs people and resources for their creation. Hence, it can be argued that some of the variance in the level of regional innovativeness results from regions' endowment with factors that affect the regional actors' ability to come up with novelty. Commonly, the investigation of this variance and the factors causing it, e.g. the effect of localization, urbanization, and Jacobs externalities, is done by relying on (mostly parametric) production function approaches.

A production function "is a mathematical function (a relation) which associates (relates) a vector of input X with the maximum level of output Y" (Bonaccorsi and Daraio, 2006, p. 54). In PPF approaches this mathematical function is defined ex-ante. By setting the explanatory variables (inputs), e.g., firms' R&D efforts and variables approximating externalities into a functional relation with explained variables (outputs), e.g. innovation numbers, a production function approach is employed. Moreover, the actually observed innovativeness is allowed to vary to both sides of the predicted innovativeness, i.e. it can show larger and smaller values than the model predicts. Hence, the function is fitted such that the function *intersects* the data. Thereby the variance that is not explained by the function is minimized. This applies to most models used in the studies cited before. Because regional innovation processes are different from classical production processes (not deterministic) a different terminology is used in the following, i.e. *(regional) input factors* instead of 'inputs' and *(regional) innovativeness* instead of 'outputs'.

With respect to the specific production function defined, it is estimated to which statistically significant extent some (or all) of the considered regional factors 'explain' the variance in the observed regional innovativeness. Or in other words, whether regions' endowments with these factors correlate to the level of innovativeness achieved by the actors located in the regions.

In general, it is acknowledged that it is impossible to consider all factors that influence the level of regional innovativeness. This is in particularly true for factors representing organizational aspects, e.g., the level of interactivity and the institutional set-up. Such implies that the proposed function can not explain all but only parts of the variance. Thus, some regions may depart from the trend represented by the estimated function. These regions show different levels of innovativeness, i.e. more or less innovations than what is 'predicted' by the estimated function and their factor endowment. The dashed (linear) regression curve in Figure 1a illustrates this for the case of the number or R&D employees located in a region and the levels of these regions' innovativeness. For most of the observations a linear relationship seems to describe the relationship between the two variables very well because there is little variance left which is not 'explained' by the regions' endowment with R&D employees. However, there is also a number of regions that ratios between R&D employees and the number of innovations depart strongly from the predicted trend (regions **A**, **B**, **C**, and **D**). If this is true for few observations the explanatory power of the model will not be effected. Hence, in general such observations are of rather little relevance in PPF approaches. In contrast, these observations are crucial in the production frontier approach.




(a) Regression approach

(b) Frontier approach

In contrast to the parametric production function approach in which the production function intersects the data, in the non-convex nonparametric production function approach, a nonparametric frontier function is estimated that envelops the data on the extreme positive border. Then regions are evaluated on the base of their position in relation to the frontier.

The frontier function shown in Figure 1b represents the type of frontier that is used in the Free Disposal Hull (FDH) approach developed by Deprins et al. (1984). It is popular in standard production analysis because it is very general and relies on few assumptions. It becomes apparent that it is not an *average* trend in the data that is relevant in the frontier approach. Instead, the existence of 'extreme' positive observations shapes the frontier. The frontier represents the *best-practice*, or *most innovative*, regions with respect to a certain level of regional factor endowment, e.g., in this example the number R&D employees. In order to identify these 'most innovative' regions, i.e. those regions that form the production frontier, in the NNPF the principle of *dominance* is employed. On the basis of this 'most innovative' regions are characterized by that no other region shows a higher level of innovativeness and has at the same time an equal or lower level of input factor endowment. For example, in Fig. 1b, there is no region with more innovations than region **A** and with equal or less R&D employees. Hence, this region dominates all other regions in terms of innovation performance given its level of R&D employment. In the case that more than one input factor are considered or more than one measure of innovativeness are employed, the vectors of the regions' input factor endowment and innovativeness are compared on the basis of whether they are *weakly dominated* or not (see, e.g., Deprins et al., 1984).

If in comparison to this reference group a region's innovativeness is higher (higher innovativeness) the region becomes part of the frontier and it is declared well performing. If its level of innovativeness is below the ones in the reference group, i.e. it is located below the frontier and this region is declared less performing. The distance to the frontier indicating its degree of (less) performance. In the presented approach as well as in the later used order-$m$ approach this distance (the performance measure) is represented by the vertical (euclidean) distance between the observations and the frontier.[2] The larger the distance the less performing this region is. In the context of this paper, this less performing may be, among others, a result of the effects of externalities.

There are different methods available for estimating the frontier function as well as the distances to the frontier (see, e.g., Scheel, 2000; Daraio and Simar, 2007). While the FDH approach is more suitable for the illustration of the general idea, its *robust* version - the order-$m$ frontier approach - will be presented and applied later in the paper.

In the following subsection a number of arguments are put forward that motivate the use of this rather unfamiliar way of analyzing regional innovation performance.

## 2.3 Arguments in favor of non-convex nonparametric frontier approaches

Why using this "exotic" approach when production function approaches have been successfully applied to this context? Whether to use production function or production frontier approaches is to some extent a matter of the context in which they are applied. Bonaccorsi and Daraio (2006) argue that for analyzing science and technology systems (S&T systems) production frontier approaches in general, and nonparametric production frontier approaches in particular, are more appropriate.

Regional innovation activities are certainly about S&T (sub-)systems. Therefore much of the arguments by Bonaccorsi and Daraio (2006) in favor of (nonparametric) production frontier techniques apply to this context as well. The following arguments are based on the discussion presented in their work. The focus lies however on those aspects that are most important in the light of this paper's theoretical background, i.e. the analysis of externalities and their impact on regional innovativeness. In addition, the discussion tries to be as non-technical as possible and rather focusses on the fit between empirical methodology and theoretical underpinning. In order to shorten the argumentation, I concentrate on comparing the two previously introduced approaches, parametric production function approach (PPF) and the non-convex, nonparametric, production frontier approach (NNPF).

**Frontier technique** In comparison to production function approaches, there are a two conceptual advantages to frontier approaches in the context of evaluating regional innovativeness. Firstly, in PPF a function is estimated that intersects the data. By this, the obtained results always represent *expected* or *average values*. In contrast, when employing frontier techniques benchmarks comprise *best practice*. As Bonaccorsi and Daraio (2006) argue "[b]est practice is not just better than average practice, it may also be structurally different ..." (p. 56). Hence, at the least, analyses based on frontier approaches may therefore lead to different (new) results than analyses employing the production function idea. It is "a common procedure in creation of policy measures and institutional settings to ask for the 'best practice' example what exists already in order to use this as a proposal or even like

[2] This view corresponds to the output-oriented type of analysis which has been argued to be most appropriate in this context by Broekel and Brenner (2007).

a blueprint for other regions" (Gracia et al., 2005, p. 5). To know which regions are best practice and why these regions achieve higher levels of innovation performance will therefore certainly be interestingly from a policy perspective.

Secondly, in light of the emphasize of highly innovative regions that can be found almost through this literature, a shift from *average practice* to *best practice* brings the empirical side closer to the theoretical conceptualizations. In the literature on regional innovativeness marginal differences in innovativeness or even average practice do not play a significant role. Instead, in terms of innovativeness extremely successful regions have been in the focus of the research: "the concept of innovation as a partly territorial phenomenon is, to a great extent, based on the successes of some specialized industrial agglomerations or regionally concentrated networks" (Doloreux and Parto, 2005, p.135). In particular the concepts of regional innovation systems, innovative milieus, and learning regions seem to be examples of *best practice* than of *average practice*. They account however for large parts of the theoretical conceptions used in this literature.

Hence, the use of frontier approaches and the shift in perspective from *average* to *best-practice* corresponds nicely to the theoretical approaches in the field of regional innovativeness which seem to have been mainly concentrated upon the latter.

**Nonparametric nature** In addition to these rather conceptual points favoring frontier approaches in this context, there exist also a number of concrete methodological advantages of the NNPF approach. The first group of these advantages result from its nonparametric nature. The *knowledge production function* (KPF) by Griliches (1979), representing a specific form of the PPF that orients on the neoclassical production function, is often used in the analysis of regional innovativeness. In its basic form the variables approximating the regional factor are multiplicatively connected. However, its specific functional form lacks any empirical evidence. In fact, Broekel and Brenner (2005) even argue that from a theoretical basis a simple linear regression approach is more appropriate in this context than a KPF style empirical model. Because of the need to specify ex-ante a functional relationship between the variables, using the PPF approach is always connected to the danger of miss-specification which can cause unreliable results. Lacking more theoretical and empirical research on the type of functional relationships between the input factors and innovativeness measures that might be appropriate, nonparametric approaches seem to be the better choice as they reduce miss-specification problems.

In addition, most researchers agree on that innovation processes are characterized by non-linearities (see, e.g., Lundvall, 1992; Storper and Scott, 1995). However, in order to account for this in the PPF it has to be decided beforehand which form the non-linearity takes. For example, Feldman and Audretsch (1999) account for non-linearities in the effects of industry specialization and localized competition on regional innovativeness by including the according variables as quadratic terms into their regression function. This accounts though only for one out of many non-linear relationships that could exist between the factors. With respect to this point, nonparametric approaches have a

clear advantage as the estimated functions can take almost any forms.
The particular non-convex version of the nonparametric frontier approach (the NNPF) relaxes further assumptions that are yet widely undiscussed and under-investigated. However, they might be important for the empirical outcomes. These assumptions regard the substitutability of regional factors as well as the presence of economies of scale in the regional innovation processes. With respect to the latter, research on agglomeration (dis-)advantages provides some insights (see, e.g., Mukkala, 2004). Nevertheless, it is widely unclear which regional input factor is subject to what kind of economies of scale. In the NNPF this is of minor importance as the employment of the criteria of *weak dominance* for defining the frontier functions ensures that no assumptions have to be made regarding economies of scale as well as the substitutability and divisibility of regional factors and innovativeness measures.[3] The only assumption made regards that all the input factors are *freely disposable* (Scheel, 2000). This seems to be uncritical in this context.
Related to this, Bonaccorsi and Daraio (2007) point toward another fundamental methodological point that is in particular relevant for the analysis of innovation performance: knowledge may be considered an output as well as an input in R&D processes. This is true for innovations as well as they are newly created knowledge. Hence, on the one hand, it is likely that they serve as input for further R&D processes. On the other hand, they also comprise an output of successful R&D processes. Thus, when analyzing the performance of innovation systems on the basis of time periods (e.g. yearly), innovations generated at the beginning of the period might already be used as input to R&D processes conducted and completed in the remainder of the period. As Bonaccorsi and Daraio (2007) note, this is problematic in production function approaches in which the independent variables (inputs) are assumed to be *independent* of the *dependent* variables (output). In the NNPF approach this assumption is not required and hence, the problem of the mutually influencing relationship between input factors and innovativeness measures is not existing.
This is however not to say that nonparametric approaches are per se advantageous. A clear empirically and theoretically confirmed parametric model allows a much more precise estimation of the significance of regional factors' effects on innovativeness. Despite recent advancements in the robust versions of nonparametric frontier techniques, parametric models (e.g. stochastic frontier functions) are also still less sensitive to noise in the data. Hence, given more knowledge about the correct set-up of the model with respect to the functional relationships, substitutability, economies of scale, and the types of non-linearities involved, parametric models are worthwhile alternatives.

**Uniqueness of regions** One aspect that is closely related to the nonparametric nature of the presented NNPF approach is that for each region an individual frontier function is estimated. The reasons for this is found in that a region's reference group, i.e. the regions which it is compared to, is defined separately for each region. Moreover, in case of an output-orientation, it is conditional on the region's input factor endowment.[4] For example, if a region has no R&D employees of a certain

[3] In fact, variable returns to scale are estimated locally and globally.
[4] In an input-oriented setting it is conditional on the the regions' values in the innovativeness measures.

type that have been defined as part of the input factor endowment it will only be compared with other regions that have no R&D employees of this type as well. This implies that in terms of its factor endowment the unique situation of each region is taken into consideration. In addition, because regions are compared to different reference groups, the frontier function against which it is evaluated can differ for each region.[5] This contrasts the PPF approach in which just *one* function is estimated that intersects the entire sample of regions' input factors and innovativeness relations. Given the uncertain, changing, and unpredictable nature of regional innovation processes the accounting for regional uniqueness is an important requirement for empirical analyses in this context. Or in other words, the use of a NNPF approach allows to relax the 'one-function-fits-all' assumption inherent to PPF approaches.

**Real observations as well as multiple input factors and innovativeness measures** Another point is that in contrast to the PPF, the non-convex nonparametric production function approach takes into account real observations only. For example, in the PPF regression curves might be fitted without directly intersecting a single observation, see, e.g., Fig. 1a. In this case it can be criticized that this represents a purely 'artificial' model as the regression function does not describe the situation of a single 'real' observation (region). Instead, the regression function describes only 'virtual' input factor - innovativeness relationships. In the NNPF approach this critique is less applicable as the frontier always intersects real observations.
In addition, real world observations are often difficult to be described in a single dimension. One of the clear strength of the NNPF is that it allows for an easy handling of multiple input factors as well as multiple innovativeness measures. In contrast, the consideration of innovativeness measures as multiple dependent variables particularly is difficult to achieve relying on PPF approaches (see Bonaccorsi and Daraio, 2006).

**Results** While the NNPF relaxes a number of significant assumptions that are inherent to PPF approaches, it is shown in the remainder of the paper that the obtained results are not less informative. To the contrary, the approach taken allows the identification of turning points as well as non-linearities which add to the understanding of the influence of externalities on regional innovation processes.

Having discussed the theoretical differences of the PPF and NNPF approaches and highlighting some advantages of the latter, the remainder of the paper provides an empirical application of the NNPF to the context of externalities' influences on regions' innovativeness.

[5] This does not mean however that a specific frontier cannot be identical for two or more regions.

# 3 Analyzing externalities' impacts on innovation performance

## 3.1 Unconditional and conditional order-$m$ frontier approaches

In the previous section theoretical arguments in favor of the NNPF approaches have been made using the Free Disposal Hull (FDH) analysis as an example. In the FDH approach the frontier is determined by the extreme positive (most innovative) observations making the performance analysis very sensitive with respect to the existence of outliers and noise in the data (see, e.g., Wilson, 1993). This drawback has been overcome by the introduction of *robust* nonparametric frontier techniques (see for an introduction Daraio and Simar, 2007). One of the *robust* versions of the FDH approach is the order-$m$ frontier approach developed by Cazals et al. (2002)[6].

In contrast to the FDH approach, the idea behind the order-$m$ approach is that instead of evaluating a region's innovation performance with respect to the performance of *all* other regions, Cazals et al. (2002) propose to compare a region with a randomly drawn (sub-) sample of regions.[7] This makes the nonparametric frontier function a *partial* frontier because not all observations are enveloped but only a sub-sample. Based on these partial frontier the evaluation of regions' innovation performance as well as the estimation of the performance scores are done in an identical manner as in the FDH approach, i.e. the principle of *weak dominance* is applied in order to estimate the frontiers.[8] Cazals et al. (2002) show that the resulting order-$m$ performance measure shares most of the characteristics of the FDH performance measure. In addition, because the (partial) frontier is not enveloping all observations, it is less sensitive to outliers and noise in the data.

The estimation of the influence of the externalities on the obtained order-$m$ performance scores is done using the *conditional* order-$m$ approach by Daraio and Simar (2005a,b). Again, it is presented only briefly.[9]

The idea proposed by Daraio and Simar (2005a) is to estimate two performance measures. The first, the *unconditional* performance measure, is calculated as described above: regions are evaluated with respect to a randomly drawn sub-sample of other regions which are characterized by equal or lower levels of factor endowment.

The second measure, the *conditional* performance measure, is estimated similar to the unconditional but in this case the sub-sample of regions used for the comparison is not drawn randomly but conditional on the value of the variable approximating the effect of the externality ('externality variable' in the following). The conditional drawing is done in a way that the sample of regions by which a region's performance is evaluated is positively biased towards those regions with similar values in the externality variable. In other words, the likelihood that a region is part of another region's comparison group depends among others negatively on the difference between the values of the regions' externality variables.

[6] Other approaches are available as well, as e.g. the order-$\alpha$ approach by Aragon et al. (2005).

[7] The sub-sample's size has to be specified by the researcher and is denoted by $m$, giving the name to the procedure.

[8] For technical details see Appendix A.

[9] For a more detailed description see Appendix A as well as Daraio and Simar (2005a,b, 2007).

Further, Daraio and Simar (2005a) suggest to estimate the ratio between the regions' two performance measures, conditional and unconditional. The value of these ratios ($Q_z$) can then be set into relation with the regions' values in the externality variable. By comparing the size of this ratio of regions showing different values in the externality variable the effect for the externality on the regional innovation performance can be analyzed. For this analyses I follow Daraio and Simar (2005a) in using scatterplots and nonparametric regression curves highlighting potential trends in the relation between ratios and externality variables. Note that the regression curve is used for graphical illustration only, the estimation of the influence of the externalities on the innovation performance is done with the mentioned frontier techniques.

From the shape of the nonparametric regression curve the following inference can be made. The ratio between conditional and unconditional performance scores increases as the value in the externality variable rises. This means that the conditional performance score grows faster than the unconditional when the value of the externality variable becomes larger. Given that large performance scores represent low performance, such a pattern reflects that when taking the externality into account, regions' innovation performance decreases as the externality variable increases. If this effect becomes stronger as the externality variable increases, it can be inferred that the externality has a positive influence on region's innovation performance. The opposite is true in case for a negative effect of the externality on the innovation performance. Hence, an increasing regression curve indicates a positive influence, while a decreasing curve hints at a negative impact.

## 3.2 Setting-up the performance analyses

In the unconditional and conditional performance analyses firms' R&D employees are defined as input factors and the externalities as external factors. This is motivated by the view that firms' R&D efforts are the most important factors in innovation processes (Feldman and Audretsch, 1999).

Using only R&D employees as input factors in a regional innovation performance analysis implies that the obtained performance measures represent firms' innovation performance aggregated to the regional level. Its variance is shaped by all other factors except firms' R&D efforts, e.g. externalities. Or in other words, the impact of the R&D employees' spatial distribution is excluded from the resulting innovation performance measure. If assuming that they also approximate firms' total investments into R&D, the resulting innovation performance represents an innovativeness measure that is 'free' from the effect of regional differences in firms' R&D investments. In an industry specific set up, as in this paper, this is likely to be the case.[10]

By this means the externalities' effects are modeled as one out of many factors that can cause differences in firms' utilization of R&D investments on a regional level.[11] In contrast to R&D employees,

[10] Other approaches could also take a regional innovation system perspective. In this case the regions' endowment with non-industrial actors and their resources are defined as additional inputs factors. However, this raises the problem of which factors to consider and which are rather seen as external factors (see Broekel and Brenner, 2007).

[11] Other factors that can result in firms' R&D performance to differ are, e.g. their embeddedness into regional networks, higher quality of R&D staff, more efficient working routines.

externalities are not necessary for innovation processes but rather take effect on the their innovation performance. Hence modeling R&D employees as input factors and the regional externalities as external factors seems to be a very well fitting setting.
Regarding the empirical performance analysis, positive effects of externalities show the following. Regions that are characterized by weak externalities (small values of the externalities variables) will be dominated in terms of innovativeness by regions with similar input factor endowments but strong externalities (large values of externalities variables). In case of negative effects the opposite is true.
With respect to the empirical assessment, in a first step the performance scores are separately estimated for each year (1999-2005), each industry (CHEM, ELEC, INSTR, TRANS), and each externality variable that is presented in Section 4. As has been put forward above, in a second step, scatterplots are used to show the relation between the ratio of conditional and unconditional performance and the values of the externality variables. Nonparametric regressions are used to indicate potential trends in the point clouds.
While the performances analyses are conducted separately for each year, the analyses of the influence of externalities is done using the pooled results of the seven years. In practice this implies that all ratios of conditional and unconditional performance scores estimated for each region and each year are depicted in the same scatterplot. Hence, in the plots each region is represented seven times (number of years). The reason for this is that by using the pooled ratios the impact of short term change which might be considered as noise in the data is reduced. Moreover, the robustness of the nonparametric regression that is used to illustrate potential trends in the data increases.
From a methodological point of view such an endeavor is appropriate if the underlying mechanisms determining regions' innovation performances do not change significantly within the time period under consideration. The theories on localization, urbanization, and Jacobs externalities provide only little reason for why the externalities' influences should change on such short term basis.[12] In order not to bias the interpretation by distortion on the extreme borders of the externality variables these are cut-off from the scatterplots. However at least 95 % of the obtained ratios are depicted.
As argued above an increasing regression curve indicates a positive impact of the externality on the regional innovation performance. In case of a decreasing regression curve the opposite is true. While most often the ratios between unconditional and conditional order-$m$ performance measure are in the range between zero and a positive value larger than one, sometimes it takes a value of one. Such can be the result of two different settings. On the one hand this can indicated that there are no reference regions for a certain region, e.g. in the case that a region shows a minimum value in one input factor.[13] On the other hand, a regions' ratio between conditional and unconditional performance scores is one if both values are identical. This is the case when the externality variable does not affect the conditional performance score. In the results section it is refrained from discussing those parts of the

[12] In fact the performance scores, as well as the regions' ranking with respect to their performance scores, changed only little between the years. This indicates that the underlying effects, i.e. influence of externalities, did not change significantly in this time period.

[13] See Appendix A.

regression curves which are shaped by the existence of large numbers of such observations because of their unclear meaning.[14]

There has been only one study using production frontier techniques for investigating the impact of localization, urbanization, and Jacobs externalities on regions' innovation performance. Fritsch and Slavtchev (2007) employ a two stage approach in which in a first stage a parametric, (quasi) deterministic as well as a stochastic, frontier approach is used for estimating performances. The resulting performance scores serve then as dependent variable on which a number of regional variables are regressed in a second stage (see for further details Fritsch and Slavtchev, 2007).

The approach taken in this paper differs from their study in a number of points. First, in order to account for sectoral and industrial differences, Fritsch and Slavtchev (2007) include some industries' employment shares as independent variables in the second stage regression. In contrast, the analyses in this paper are conducted separately for four industries, i.e. the obtained performance measures are industry specific in the first place. Second, unlike Fritsch and Slavtchev (2007), the present paper makes use of separate measures for localization, urbanization, and Jacobs externalities. Third, the effects of regional factors on the performance scores are estimated directly (one stage approach) instead of explaining the variation in the scores in a second stage regression. Fourth, by using a nonparametric approach mis-specification problems are avoided.

# 4 Employed data

## 4.1 Data on innovativeness and R&D activities

The 270 German labor market regions are chosen as units of analysis, because they seem to fit best to the theoretical arguments of a regional dimension of innovation processes (see, e.g., Broekel and Binder, 2007).[15] As it is common in innovation research innovativeness is approximated by patent applications. The data on patent applications for the years 1999-2005 are published by the *Deutsches Patent- und Markenamt (German Patent Office)* in Greif and Schmiedl (2002) and Greif et al. (2006) (called *Patentatlas* in the following). The applications by public research institutes, e.g., universities and research societies (e.g. Max Planck Society) as well as the patent applications by private inventors are not included. The latter is because the corresponding R&D employment data covers only industrial R&D. Hence, only the patent applications of industrial R&D should be considered.

Data on R&D employees is obtained from the German labor market statistic. Following Bade (1987) the R&D personnel is defined as the sum of the occupational groups: agrarian engineers (032), engineers (60), physicists, chemists, mathematicians (61) and other natural scientists (883).

Conducting industry specific analyses requires definitions of industries that, in the context here,

[14] Here such observations result mainly from that one or more input variables are characterized by a large number of zeros. Hence, to a large extent this is a data problem.

[15] I use the up-to-date definition of labor market regions in contrast to the older definition used in Greif and Schmiedl (2002) and Greif et al. (2006). For the analysis of externalities' influences on regions' performances labor market regions are commonly chosen as regional units (see, e.g., Combes, 2000).

cover the input factor as well as the innovativeness measure side. In other words, the industries' R&D employees need to reflect those firms to which the patent applications of the *Patentatlas* correspond. This is an important issue because in the *Patentatlas* the patent applications are classified according to 31 technological fields (TF). In contrast, the industry specific R&D employment is organized according to the German Industry Classification ('Deutsche Wirtschaftskzweig Klassifikation') which is the German equivalent to the international NACE classification. Thus, the technological fields classification in the *Patentatlas* as well as the German Industry Classification need to be matched.

I use the concordance between these two classifications developed by Broekel (2007). The concordance defines five 'sectors' for which it is possible to assign a number of the *Patentaltas'* technological fields to a number of industries defined by the German Industry Classification (see for further details Broekel, 2007).

This paper concentrates on four of these 'sectors': Chemistry (CHEM), Transport equipment (TRANS), Electrics & electronics (ELEC), and Instruments, medical & optical equipment (INSTR).[16] Their definitions in terms of assigned technological fields and industries is presented in Table 4.

For most of these industries patenting represents an important property rights protection mechanism (Arundel and Kabla, 1998). This ensures that the innovativeness measure captures most, or at least a significant share of, innovations in this industry.

The patent applications of the technological fields that are assigned to the industries are added up to obtain a single variables approximating innovativeness. The reason for adding up the patent applications is that there are a number of technological fields that cover few positive observation. Eventually this reduces the variance in the performance measures making the detection of an influence of cooperativeness statistically more difficult. Thus, the adding-up increases the analyses' explanatory power. Moreover, this reduces the number of dimensions of innovativeness measure (output space) reducing the 'sparsity problem' meaning that regions are deemed well-performing simply because there exist no peers with which to compare it (see on this Witt, 2001). However, this is achieved at the costs of not taking into account some diversity in the technological structure of the industries' innovations.[17]

On the input factor side the organization of the R&D employees into different two-digit industries is kept. Table 4 shows which two-digit industries are used to define the four industries.[18]

Summarizing, each industry is defined by a single innovativeness variable (patent applications) and two to three input factor variables (R&D employees) which enter the performance analysis.

[16] Please note that the terms 'sector' and 'industry' can be arbitrarily used in this context. For the sake of readability the term 'industry' is used in the remainder of the paper.

[17] Please note that the analysis can handle multiple output variables. However, the data does not allow to make use of this.

[18] The analysis is restricted to regions with at least one positive values in one input factor variables (R&D employees). This eliminates less than four percent of the observations.

## 4.2 Data on externalities

Following Burger et al. (2007), urbanization externalities are approximated by population density (POP_DEN). It is estimated as the number of inhabitants per square kilometer. Note that in contrast to the other variables, the data for this variable is only available for 1999-2004 and hence, the analyses are conducted for this time period. The data on the population density is obtained from the German statistical office.

It is common to approximate localization externalities by the degree of regional specialization (Feldman and Audretsch, 1999; van der Panne and van Beers, 2006). A frequent way to measure the latter is the 'production structure specialization index' (PS). It is estimated as follows.

$$PS_{s,r} = \frac{X_{s,r}/\sum\limits_{s} X_{s,r}}{\sum\limits_{s,r} X_{s,r}/\sum\limits_{s}\sum\limits_{r} X_{s,r}}$$

$X_{s,r}$ is the employment of sector $s$ in region $r$. The numerator is a sector's share on a region's employment and the denominator represents the sector's share on total (national) employment. If $PS_{s,r}$ is equal to unity for a sector in a region, the specialization is identical to the average of all regions. A PS above one indicates above average specialization, while for a $PS$ below unity the opposite is true.

This index is however non-symmetric, i.e. in case of below average specialization the index takes values between zero and one, and in case of above average specialization its values range between one and infinity. This makes it "basically not comparable on both sides of unity" (Laursen, 1998, p. 3). Therefore, the index (PS) is made symmetric as proposed by Laursen (1998) in a different context by calculating $\frac{PS-1}{PS+1} + 1$

This index, denoted as SPEC in the following, ranges from $0$ to $+2$. One is added to it in order to keep some similarity to the traditional PS. Unity represents that there is no difference between a region's degree of specialization and the national average.

In the literature Jacobs externalities are approximated by a wide range of different diversity indices. I follow Henderson et al. (1995) in employing an inverted Hirschman-Herfindahl index. As suggested by Combes (2000) this index is normalized at the value it takes at the national level. The estimation of the diversity index $DIV^{n}_{r,s}$ is conducted using employment data with $emp_{r,s'}$ referring to the employment of sector $s$ in region $r$; $emp_r$ represents the total employment in region $r$; $emp_s$ the sector's total employment.[19] Its exact estimation shows the following.

$$DIV_{r,s}^{n} = \frac{1/\sum_{s'=1,s'\neq s}^{S} (emp_{r,s'}/(emp_r - emp_{r,s}))^2}{1/\sum_{s'=1,s'\neq s}^{S} (emp_{s'}/emp - emp_s)^2} \tag{1}$$

$S$ represents the total number of sectors, i.e. the number of two-digit manufacturing industries. The numerator is maximal in the case that all industries (except the one under investigation) are of identical size.

The advantage of this diversification index is that it estimates the "sectoral diversity faced by sector $s$ in this zone [region] and is therefore not necessarily negatively linked with the own local specialization of sector $s$" (Combes, 2000, p. 337). Hence, it is an industry specific diversification index which is not just reflecting 'negative specialization'. It indicates the diversification in the manufacturing sector.

As in the case of the production specialization index, this index is not symmetric. Again, the index is made symmetric in an identical manner as SPEC by calculating

$$DIV_{man} = \frac{DIV_{r,s}^{n} - 1}{DIV_{r,s+1}^{n}} \quad . \tag{2}$$

The data on employment has been collected from the German labor market statistics.

The three variables (POP_DEN, SPEC, and $DIV_{man}$ are only weakly correlated with each other, see Table 5 in the Appendix. Therefore, a univariate approach, i.e. the externality variables are tested separately seems to be valid. The results for the impact of each variable on the regional innovation performance are presented in the following section.

# 5 Results

## 5.1 Urbanization externalities

The findings on the influence of urbanization externalities, approximated by population density, are unambiguous. In all investigated industries urbanization is positively associated with regional innovation performance, see Fig. 2 in the Appendix.

With respect to the specific shape of POP_DEN's influence on RegIP in most industries an almost linear increasing regression curve is found up to a level of 1,000 people / $km^2$. In CHEM the increase seems to be above linear, i.e. a power function with an exponent somewhat larger than one. For values of POP_DEN larger than 1,000 people / $km^2$ there are not enough observations in order

[19] Further, $^n$ indicates the non symmetric character of this variables.

| Factor | CHEM | TRANS | ELEC | INSTR |
|---|---|---|---|---|
| **POP_DEN*** | positive | positive | positive | positive |
| *Range* | $POP\_DEN < 1t$ | $POP\_DEN < 1t$ | $POP\_DEN < 1t$ | $POP\_DEN < 1t$ |
| * POP_DEN in thousand (t) inhabitants per $km^2$. | | | | |

Table 1: Results on localization externalities

to make a solid interpretation. However, it can be speculated that the positive influence weakens or even turns negative.

## 5.2 Localization externalities

In contrast to the findings on urbanization economies, the results on localization externalities approximated by the degree of employment specialization, need some more discussion. Figure 3 in the Appendix shows the according scatterplots and nonparametric regression curves.

When catching a glimpse on the scatterplots it seems that in all industries a more or less developed U-shaped regression curve describes the influence of SPEC on the industries' RegIP best. However in all cases the decreasing trend is driven by the pattern described before: a number of observations with low values in SPEC are characterized by $Q_z = 1$. As SPEC is increasing $Q_z$ takes values below one. As argued in Section 3 it is refrained from discussing such relationships.

In ELEC the decreasing trend changes to an increasing one at value of about 0.5. In CHEM the degree of specialization needs to exceed a value of $SPEC = 1$ in order for the curve to change its slope from negative to positive. In INSTR a positive slope is found only for values larger than $SPEC = 1.2$, however this is backed by few observations. Similar holds in case of TRANS with a strong increase in the regression curve for SPEC values above one, but few observations supporting it. Thus, for all industries, after exceeding an industry specific turning point, the degree of specialization works in favor of regional innovation performance. For values of SPEC below this turning point, no positive influence can be detected. In case of INSTR and TRANS the positive impact of the degree of specialization shows only for few regions that are very highly specialized.

With respect to the importance of firm size effects, the presence of some 'extreme' positive observations in INSTR and TRANS (in terms of the ratio between conditional and unconditional performance) hint at a positive impact of localization economies on RegIP that may results from the presence of large multinational corporations. For example, in the case of TRANS the regions with the highest ratio between conditional and unconditional performance scores are Stuttgart and München, and for ELEC München and Erlangen. All these regions do not only show high degrees of specialization in the according industries, but they are also well known for being the home of large multinational corporations and their manufacturing facilities, e.g. Daimler AG, BMW AG, and SIEMENS AG. One can therefore speculated that parts of the positive relationship between the degree of specialization and RegIP may be caused by the presence of these multinationals.

| Factor | CHEM | TRANS | ELEC | INSTR |
|---|---|---|---|---|
| **SPEC** | weakly positive | strong positive | positive | weakly positive |
| *Range* | $SPEC > 1$ | $SPEC > 1$ | $SPEC > 0.5$ | $SPEC > 1.2$ |

Table 2: Results on localization externalities

Table 2 summarizes the findings for localization externalities. In general positive effects stemming from localization externalities on a region's innovation performance are of importance in all four industries. Hence, this paper confirms the findings on positive effects of localization economies for the German case. In particular, in ELEC and TRANS the evidence points to a strong importance of localization economies. The results for all industries are characterized by the existence of turning points and non-linearities.

## 5.3 Jacobs externalities

In case of the diversification measure it is important to note that the diversity measure is small in case of strong diversification and large in case of no diversification. In contrast to the previous variables, a negative regression curve therefore indicates a positive influence of diversity on industries' regional innovation performance.

In case of $DIV_{man}$ in three industries, ELEC, INSTR, and TRANS a negative, almost linear regression curve is found, see Fig. 4 in the Appendix. This implies that in these industries a high diversification in the manufacturing sector increases the innovation performance, i.e. Jacobs externalities are positively effective. Similarly as for the other industries, for CHEM the curve is almost linear. In contrast, it is characterized by a negative slope suggesting that the manufacturing sectors' diversity influences this industries' innovation performance negatively. However, the slope departs little from zero, indicating a comparatively weak influence. Only in case of TRANS the decreasing trend in the regression curves is more pronouncedly developed. This indicates that diversity in the manufacturing sector is of larger importance for this industry's firms. The analyses reveal that significant difference exist in the importance of Jacobs externalities for different industries. In particular, while technological spill-overs are of importance for the innovative activities of TRANS and to a smaller extent in ELEC, they seem to be of minor importance in the other two industries. In the case of CHEM furthermore, diversity in the manufacturing sector takes a negative effect on this industry's RegIP. At the same time it has been shown before that this industry profits the strongest from urbanization economies. Taking both findings together this seems to suggest that it is not the tech-

| Factor | CHEM | TRANS | ELEC | INSTR |
|---|---|---|---|---|
| $DIV_{man}$ | weakly negative | positive | weakly positive | weakly positive |

Table 3: Results on Jacobs externalities

nological spillover potential of other manufacturing industries that matters for CHEM's firms. This indicates rather that, in the case of CHEM, the positive effects of urbanization are not connected to a more diversified manufacturing structure. As CHEM is a science based industry which firms rely strongly on public research (Pavitt, 1984), these positive effects are likely to stem from geographic proximity to relevant public research facilities. These are strongly concentrated in urban centers.
According to Pavitt (1984) TRANS is a *scale-intensive* industry and its most important sources of technological know-how are suppliers and consulting engineers. Hence, it can be argued that to some extent the positive effects of diversification in the manufacturing sector result from the presence of supplier industries within the same region. Note that this does not cancel out the previous findings on the importance of localization economies, because the diversification measure is not just the opposite of the specialization index and both (SPEC and $DIV_{man}$) show a low correlation with each-other, see Table 5 in the Appendix. Contrasting the other industries TRANS is thus positively influenced by all three types of externalities.
In general the regression curves for all industries do not depart strongly from a linear trend. Summarizing, the evidence for effects of Jacobs externalities is for all industries, with the exception of TRANS, rather weak and differs between industries. This is in line with previous findings in the literature (see, e.g., van der Panne, 2004). For TRANS, Jacobs externalities have a comparatively strong positive influence.

# 6 Conclusion

The impact of localization (Marshall externalities), urbanization, and Jacobs externalities on firms' innovation activities has been the focus of a rich literature. Despite the many studies conducted the findings on the importance of these regional factors are still contradictory. In previous work on this issue mainly some sorts of production function approaches have been employed, e.g. the *knowledge production function*.
It is argued in the paper that in this context, the use of production function approaches and in particular the use of parametric production function approaches do not seem to be the most appropriate. Instead the paper advocates the application of nonparametric production function approaches. By comparing a specific non-convex, nonparametric, production frontier approach with the common parametric production function approach, four arguments have been put forward in favor of the first. They regard a shift in the perspective taken (from *average-practice* to *best-practice*); advantages of nonparametric models; the accounting for the uniqueness of regions, and the fit of the empirical model to real world observations. In this way, the paper provided a discussion of the appropriateness of the two approaches for the investigation of regional innovativeness.
In the second part of the paper an application of a non-convex, nonparametric production frontier approach to the context of externalities' influence on regional innovativeness has been presented. By this means it was confirmed that nonparametric production function approaches are applicable in

this context.
In particular the paper contributed to the existing literature in three ways. First, it was argued in the paper that a nonparametric production frontier approach is more applicable for analyzing regional innovativeness than traditional approaches. Most importantly, the nonparametric production frontier approach allows to relax a number of theoretically problematic assumptions inherent to the commonly employed parametric production function techniques.
Second, despite the fewer assumptions imposed in the empirical endeavor, employing German labor market region data, the obtained results confirmed previous findings in the literature. For example, in all considered industries, localization and urbanization externalities influence industries' regional innovation performance. In comparison to these two, the evidence for Jacobs externalities is somewhat weaker but still existing. Furthermore, the strengths and shapes of these influences differ between the considered industries. The applicability of the nonparametric frontier approach has thus not only been motivated theoretically but it was also shown that the approach delivers interesting results that fit the theoretical predictions.
Third, the nonparametric nature of the approach taken allowed to obtain a more detailed picture of how the externalities take effect on regional innovation performance. Evidence was found for the existence of industry-specific turning points and non-linearities in the ways firms' innovation processes are effected by the externalities. However, in case of population density as proxy for urbanization economies as well as the manufacturing sector's diversity approximating Jacobs externalities, a linear relationship seems to be well describing their impact on regions' innovation performance. This contrasts the assumption of log-linear relationships which are often found in the literature. In case of the impact of the degree of specialization as measure for localization economies on innovativeness, the results suggested that rather non-linear functions are most appropriate.

# Appendix

## A Conditional and unconditional order-$m$ frontier analysis

The following is a brief technical representation of the conditional and unconditional order-$m$ frontier approach as proposed by Daraio and Simar (2005a,b). For a detailed introduction into robust production frontier techniques see Simar and Wilson (2006) and for an extensive treatment on the conditional and unconditional order-$m$ approach see Simar and Wilson (2006).
The main idea of the unconditional order-$m$ is simple: For a multivariate case consider $(x_0, y_0)$ as the inputs and outputs of the unit of interest. $(X_1, Y_1), ..., (X_m, Y_m)$ are the inputs and outputs of $m$ randomly drawn units that satisfy $X_i \leq x_0$. $\tilde{\lambda}_m(x_0, y_0)$ measures the distance between point $y_0$ and the order-$m$ frontier of $Y_1, ..., Y_m$. It can be written as:

$$\tilde{\lambda}_m(x_0, y_0) = \underset{i=1,...,m}{max} \left\{ \underset{j=1,...,q}{min} \left(\frac{Y_i^j}{y^j}\right) \right\} \tag{3}$$

with $Y_i^j$ ($y_0^j$) as the $j$th component of $Y_i$ (of $y_0$ respectively). The order-*m* efficiency measure of unit $(x_0, y_0)$ is defined as

$$\lambda_m(x_0, y_0) = E[\tilde{\lambda}_m(x_0, y_0)|X \leq x_0] \ . \tag{4}$$

The obtained performance measure represents the radial distance of the unit to the order-$m$ frontier. Note that in any case a unit is at least compared to itself which results in a performance score of one.
In order to calculate the order-*m* frontiers Cazals et al. (2002) suggest to employ a Monte-Carlo approximation with 200 replications which is followed in this paper as well. An open issue is the choice of parameter $m$. The literature does not yet provide a definite rule on how to choose the value of $m$. "Experience has shown that in many applications, qualitative conclusions are little affected by particular choices of $m$, provided the value of $m$ are somewhat less than the sample size, $n$" (Simar and Wilson, 2006, p. 78). In the paper I follow Bonaccorsi et al. (2005) in setting the level of robustness to ten percent, i.e. ten percent of the units are outside the frontier. The larger the performance measure is for a region, the less well does this region perform.

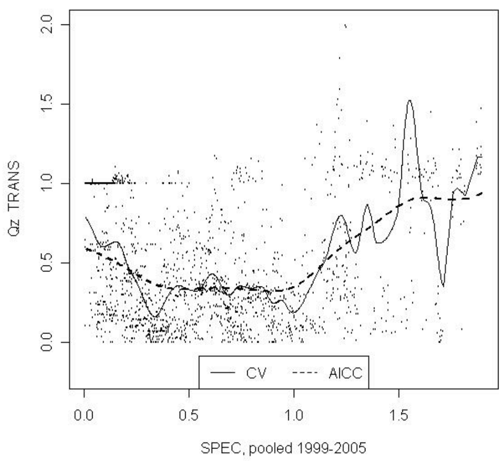



Figure 1: Comparison of bandwidths

The idea behind the conditional performance approach is that in the estimation of a region's performance more weight is given to the comparison of this region with regions with similar values in an external factor $z$, e.g. the externality variable. Thus, this 'conditional'

(conditional on the external factor's value) obtained order-$m$ performance measure is "biased" towards the comparison with other regions that show similar values in the external factor. In order to achieve this, the $m$ observations are not drawn randomly but conditional on the external factor. For unit 0, with a input level $x_0$ Daraio and Simar (2005b) propose a probability of $\frac{K(z-z_i)/h}{\sum_{j=1}^{n} K((z-z_j)/h)}$ to draw (with replacement) among those $Y_i$ such that $X_i \leq x_0$, $m$ comparison regions, whereby $K$ is a symmetric kernel with compact support[20], $z$ the external factor, and $h$ a chosen bandwidth for this particular kernel. The choice of the bandwidth is done as suggested by Daraio and Simar (2005a, 2007) using their k-nearest neighbor method which is based on likelihood cross-validation for the density of $z$. The estimation of the performance scores is conducted identically to the order-$m$ approach presented before.

By estimating the ratio between the two performance measures, conditional order-$m$ ($\lambda_m^z$) and unconditional order-$m$ ($\lambda_m$), the influence of the external factor on the regional innovation processes can be analyzed. This ratio $\frac{\lambda_m^z}{\lambda_m}$ is denoted as $Q_z$ in following. In this univariate setting (only one external factor $z$ is considered at a time) the relationship between $Q_z$ and the value of $z$ can be illustrated by a scatterplot. Moreover, a nonparametric regression helps to describe the external factors influence on the innovation processes.

With respect to the employed nonparametric regression technique that helps to interpret the scatterplots, I follow Daraio and Simar (2007) in using a simple Nadaraya-Watson (Nadaraya, 1964; Watson, 1964) estimator with a Gaussian kernel. A crucial aspect in nonparametric regression is the choice of the degree of smoothing, i.e. the choice of the appropriate bandwidths (see, e.g., Bowman and Azzalini, 1997). In the context of the employed analysis Daraio and Simar (2007) suggest to use a least-squares cross-validation (CV) automatic procedure.

However, it seems that that the bandwidths obtained by the CV method are too small for the data used in the paper. In Figure 1 the solid line represents the regression curve estimated on the base of the CV-bandwidths. Looking at the bumpy curve it seems that this method 'undersmoothes' the relationship. Therefore, the bandwidths are chosen using the *Improved Akaike Information Criterion* (AICC) developed by Hurvich et al. (1998) in the following. In Figure 1 the dashed line represents the smoother regression curve obtained by using the AICC-bandwidths.

[20] An epanechnikov kernel is used here.

# B Tables

| Sector | Technological fields* | Industries** | Control *** |
|---|---|---|---|
| Chemistry | TF5, TF12, TF13, TF14, TF15 | DG24, DI26 | TF6 ,TF20, DF23 |
| Machine building | TF1, TF2, TF3, TF7, TF8, TF9, TF11, TF17, TF18, TF19, TF20,TF21, TF23, TF24, TF25 | DA15, DA16, DB17, DB18, DC19, DC20, DE21, DE22, DH25, DJ27, DJ28, DK29, DN36 | TF6, TF22, DM34 |
| Transport equipment | TF10, TF22 | DM34, DM35 | TF23, TF20 |
| Electrics & electronics | TF27, TF28, TF29, TF30, TF31 | DL30, DL31, DL32 | DL33 |
| Medical & optical equipment | TF4, TF16, TF26 | DL33, DF23 | TF6, TF15, DL30 |
| * As defined in Greif and Schmiedl (2002) ** According to the GIC DESTATIS (2002) *** Technological fields of industries which have to be controlled for | | | |

Table 4: Definitions of industries

| CHEM | $DIV_{man}$ | $SPEC$ | $POP\_DEN$ |
|---|---|---|---|
| $DIV_{man}$ | – | 0.033 | −0.028 |
| $SPEC$ | 0.033 | – | −0.051 |
| $POP\_DEN$ | −0.028 | −0.051 | – |
| ELEC | $DIV_{man}$ | $SPEC$ | $POP\_DEN$ |
| $DIV_{man}$ | – | −0.211*** | −0.005 |
| $SPEC$ | −0.211*** | – | 0.178*** |
| $POP\_DEN$ | −0.005 | 0.178*** | – |
| INSTR | $DIV_{man}$ | $SPEC$ | $POP\_DEN$ |
| $DIV_{man}$ | – | −0.137** | −0.019 |
| $SPEC$ | −0.137** | – | 0.232*** |
| $POP\_DEN$ | −0.019 | 0.232*** | – |
| TRANS | $DIV_{man}$ | $SPEC$ | $POP\_DEN$ |
| $DIV_{man}$ | – | −0.055 | −0.037 |
| $SPEC$ | −0.055 | – | 0.177*** |
| $POP\_DEN$ | −0.037 | 0.177*** | – |
| All values represent Spearman's rank correlation coefficients. | | | |

Table 5: Correlations of employed variables

# C Figures

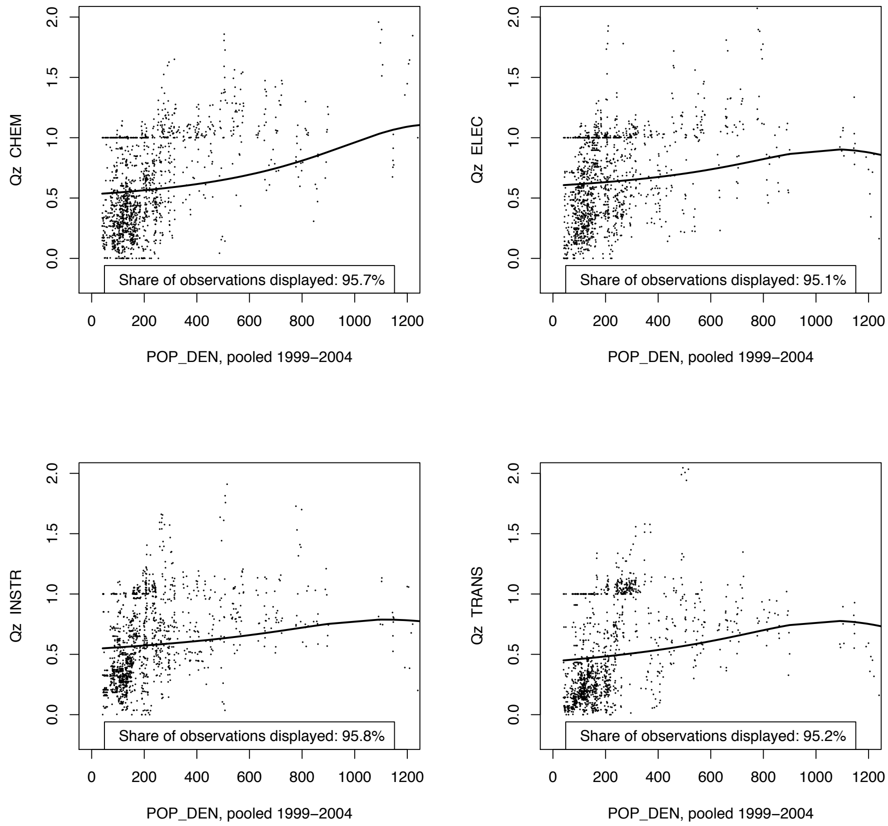



Figure 2: Results urbanization

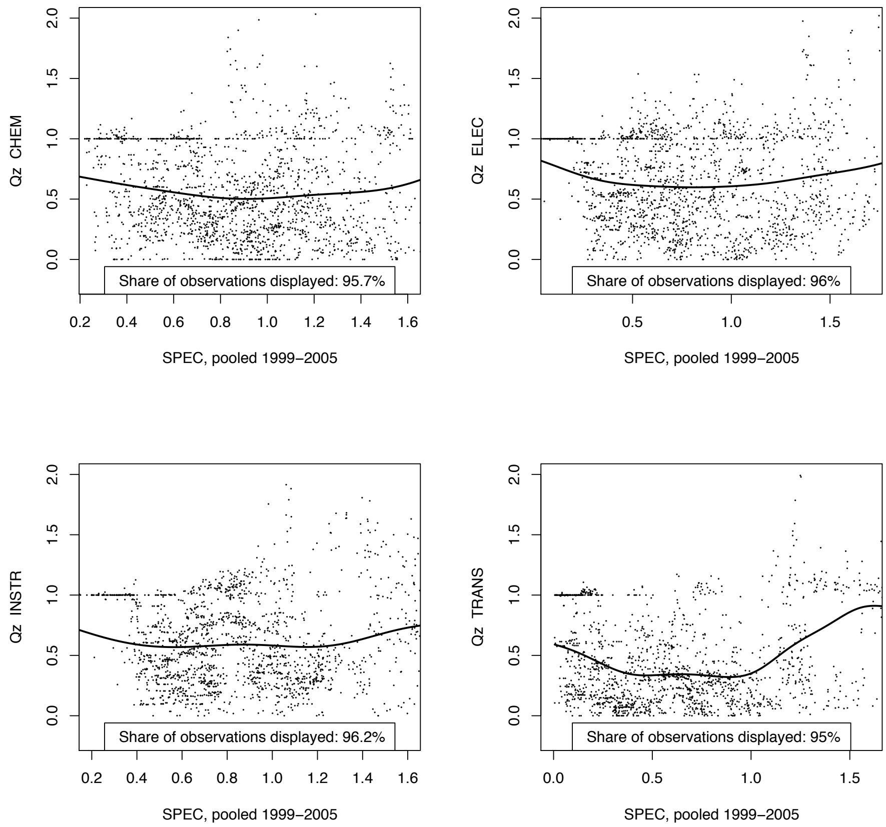



Figure 3: Results SPEC

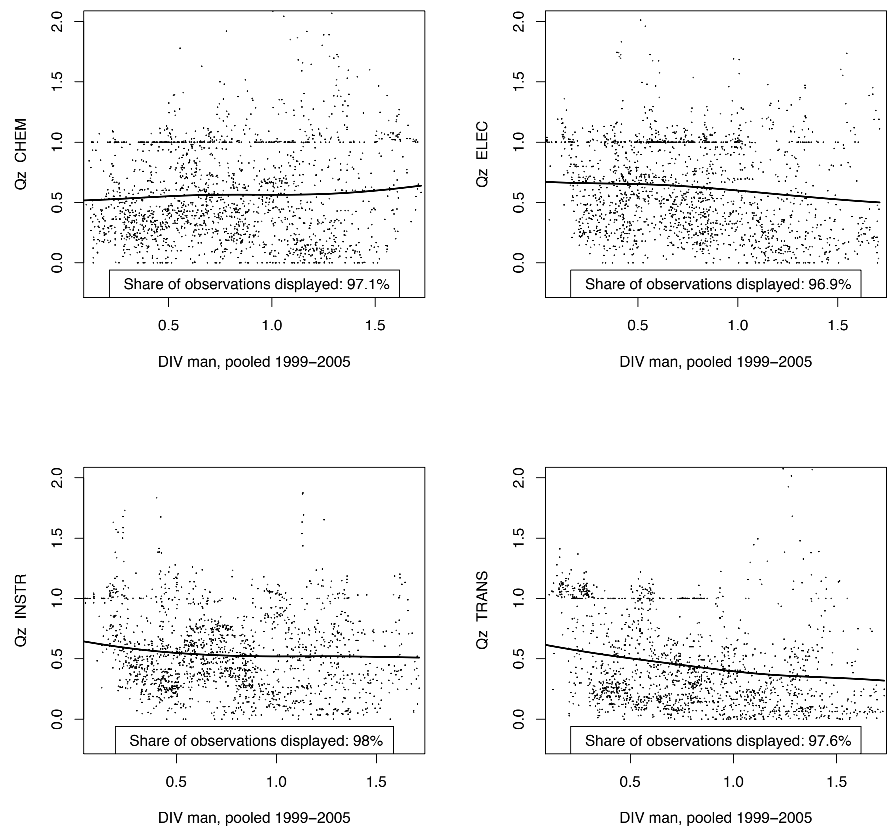



Figure 4: Results Jacobs externalities

# References


Aragon, Y., Daouia, A., and Thomas-Agnan, C. (2005). A Conditional Quantile-Based Efficiency Measure. *Econometric Theory*, 21:358–389.

Arrow, K. (1962). The Economic Implications of Learning by Doing. *Review of Economic Studies*, 29:155–173.

Arundel, A. and Kabla, I. (1998). What Percentage of Innovations are Patented? Empirical Estimates for European Firms. *Research Policy*, 27(2):127–141.

Bade, F.-J. (1987). *Regionale Beschäftigungsentwicklung und produktionsorientierte Dienstleistungen*. Sonderheft 143. Deutsches Institut für Wirtschaftsforschung, Berlin.

Bonaccorsi, A. and Daraio, C. (2006). Econometric Approaches to the Analysis of Productivity of R&D Systems. In *Handbook of Quantitative Science and Technology Research Handbook of Quantitative Science and Technology Research - The Use of Publication and Patent Statistics in Studies of S&T Systems*, chapter 2, pages 51–74. Springer Netherland.

Bonaccorsi, A. and Daraio, C. (2007). Measuring Knowledge Spillover Effects via Conditional Nonparametric Analysis. *Paper presented at the Workshop on Agglomeration and Growth in Knowledge-based Societies in Kiel, Germany, April 20-21, 2007.*

Bonaccorsi, A., Daraio, C., and Simar, L. (2005). Advanced Indicators of Productivity of Universites. An Application of Robust Nonparametric Methods to Italian Data. *Scientometrics*, 66(2):389–410.

Bowman, A. W. and Azzalini, A. (1997). *Applied Smoothing Techniques for Data Analysis - The Kernel Approach with S-Plus Illustrations*. Oxford University Press Inc., New York.

Broekel, T. (2007). A Concordance between Industries and Technologies - Matching the technological fields of the Patentatlas to the German Industry Classification. *Jenaer Economic Research Papers*, #2007-013.

Broekel, T. and Binder, M. (2007). The Regional Dimension of Knowledge Transfers - A Behavioral Approach. *Industry and Innovation*, 14(2):151–175.

Broekel, T. and Brenner, T. (2005). Local Factors and Innovativeness - An Empirical Analysis of German Patents for Five Industries. *Papers on Economics & Evolution, revised version, available upon request from the authors*, 509.

Broekel, T. and Brenner, T. (2007). Measuring Regional Innovativeness - A Methodological Discussion and an Application to One German Industry. *DIME Working Paper*, 2007-13.

Burger, M., van Oort, F., and van der Knaap, B. (2007). A Treatise on the Scale-Dependency of Agglomeration Externalities and the Modifiable Areal Unit Problem. *Paper presented at the Workshop on Agglomeration and Growth in Knowledge-based Societies in Kiel, Germany, April 20-21, 2007.*

Cazals, C., Florens, J.-P., and Simar, L. (2002). Nonparametric Frontier Estimation: A Robust Approach. *Journal of Econometrics*, 106(1):1–25.

Combes, P.-P. (2000). Economic Structure and Local Growth: France, 1984-1993. *Journal of Urban Economics*, 47(3):329–355.

Daraio, C. and Simar, L. (2005a). Conditional Nonparametric Frontier Models for Convex and Non Convex Technologies: a Unifying Approach. *LEM Working Paper Series*, 12.

Daraio, C. and Simar, L. (2005b). Introducing Environmental Variables in Nonparametric Frontier Models: A Probabilistic Approach. *Journal of Productivity Analysis*, 24(1):93–121.

Daraio, C. and Simar, L. (2007). *Advanced Robust and Nonparametric Methods in Efficiency Analysis - Methodology and Applications*. Kluwer Academic Publishers, Boston / Dordrecht / London.

de Groot, H. L. F., Poot, J., and Smit, M. J. (2007). Agglomeration, Innovation and Regional Development. *Tinbergen Institute Discussion Papers*, 079(3).

Deprins, D., Simar, L., and Tulkens, H. (1984). Measuring Labor Efficiency in Post Offices. In Marchand, M., Pestieau, P., and Tulkens, H., editors, *The Performance of Public Enterprises: Concepts and Measurements*, pages 345–367. Elsevier.

DESTATIS (2002). *Klassifikation der Wirtschaftszweige, Ausgabe 2003 (WZ2003)*. Statistisches Bundesamt, Wiesbaden.

Doloreux, D. and Parto, S. (2005). Regional Innovation Systems: Current Discourse and Unresolved Issues. *Technology in Society*, 27(2):133–153.

Feldman, M. P. and Audretsch, D. B. (1999). Innovation in Cities: Science-based Diversity, Specialization and Localized Competition. *European Economic Review*, 43:409–429.

Fritsch, M. and Slavtchev, V. (2007). What Determines the Efficiency of Regional Innovation Systems? *Jenaer Economic Research Papers*, 6.

Glaeser, E., Kallal, H., Scheinkam, J., and Shleifer, A. (1992). Growth in Cities. *The Journal of Political Economy*, 100:1126–1152.

Gracia, A. G., Voigt, P., and Iturriagagoitia, J. M. Z. (2005). Evaluating the Performance of Regional Innovation Systems. *5th Triple Helix Conference on "'The Capitalization of Knowledge: Cognitive, Economic, Social & Cultural Aspects", Turin, Italy, May 18.-21. 2005*.

Greif, S. and Schmiedl, D. (2002). *Patentatlas 2002 Dynamik und Strukturen der Erfindungstätigkeit*. Deutsches Patent- und Markenamt, München.

Greif, S., Schmiedl, D., and Niedermeyer, G. (2006). *Patentatlas 2006. Regionaldaten der Erfindungstätigkeit*. Deutsches Patent- und Markenamt, München.

Greunz, L. (2004). Industrial Structure and Innovation - Evidence from European Regions. *Journal of Evolutionary Economics*, 14(5):563–592.

Griliches, Z. (1979). Issues in Assessing the Contribution of R&D to Productivity Growth. *Bell Journal of Economics*, 10:92–116.

Henderson, V., Kuncoro, A., and Turner, M. (1995). Industrial Development in Cities. *The Journal of Political Economy*, 103(5):1067–1090.

Hoover, E. M. (1937). *Location Theory and the Shoe and Leather Industry*. Harvard University Press, Cambridge.

Hoover, E. M. (1948). *The Location of Economic Activity*. McGraw Hill, New York.

Hurvich, C. M., Simonoff, J. S., and Tsai, C.-L. (1998). Smoothing Parameter Selection in Nonparametric Regression Using an Improved Akaike Information Criterion. *Journal of the Royal Statistical Society, Series B*, 60(2):271–293.

Isard, W. (1956). *Location and Space-Economy*. MIT Press, Cambridge, MA.

Jacobs, J. (1969). *The Economy of Cities*. Random House, New York.

Laursen, K. (1998). Revealed Comparative Advantage and the Alternatives as Measures of International Specialisation. *DRUID Working Papers*, 98-30.

Lundvall, B. A. (1992). *National Systems of Innovation: Towards a Theory of Innovation and Interactive Learning*. Pinter Publishers. London. UK.

Marshall, A. (1890). *Principles of Economics*. MacMillian, London (8th edition 1920).

Mukkala, K. (2004). Agglomeration Economies in the Finnish Manufacturing Sector. *Applied Economics*, 36:2419–2427.

Nadaraya, E. A. (1964). On Estimating Regression. *Theory of Probability Applications*, 9:141–142.

Paci, R. and Usai, S. (1999a). Externalities, Knowledge Spillovers, and the Spatial Distribution of Innovation. *Geojournal*, 49:381–390.

Paci, R. and Usai, S. (1999b). The Role of Specialization and Diversity Externalities in the Agglomeration of Innovative Activity. *Working Paper CRENoS*, 199915.

Pavitt, K. (1984). Sectoral Patterns of Technical Change: Towards a Taxonomy and a Theory. *Research Policy*, 13(6):343 – 373.

Romer, P. M. (1986). Increasing Returns and Long-Run Growth. *Journal of Political Economy*, 1002-1037(94).

Scheel, H. (2000). *Effizienzmaße der Data-Envelopment-Analysis*. Deutscher Universitäts-Verlag, Schriftenreihe: Gabler Edition Wissenschaft, Wiesbaden.

Simar, L. and Wilson, P. W. (2006). Statistical Inference in Nonparametric Frontier Models: Recent Developments and Perspectives. In H. Fried, C. L. and Schmidt, S., editors, *The Measurement of Productive Efficiency*, chapter 4. Oxford University Press.

Storper, M. and Scott, A. J. (1995). The Wealth of Regions. *Futures*, 27:505–526.

van der Panne, G. (2004). Agglomeration Externalities: Marshall versus Jacobs. *Journal of Evolutionary Economics*, 14(5):593–604.

van der Panne, G. and van Beers, C. (2006). On the Marshall-Jacobs Controversy: It Takes Two to Tango. *Industrial and Corporate Change Advance Access published on August 8,2006, DOI.*

van Oort, F. (2002). Innovation and Agglomeration Economies in the Netherlands. *Journal of Economic and Social Geography*, 93:344–360.

Watson, G. S. (1964). Smooth Regression Analysis. *Sankhya Series A*, 26:359–372.

Wilson, P. W. (1993). Detecting Outliers in Deterministic Nonparametric Frontier Models with Multiple Outputs. *Journal of Business & Economic Statistics*, 11(3):319–323.

Witt, U. (2001). Learning to Consume- A Theory of Wants and the Growth of Demand. *Journal of Evolutionary Economics*, 11:23–36.