

Paper no. 2013/43

**Are there Scale Economies in Scientific Production?
On the Topic of Locally Increasing Returns to Scale**

Torben Schubert (torben.schubert@circle.lu.se)
CIRCLE, Lund University, Sweden

and Fraunhofer Institute for Systems and Innovation Research and
Technological University Berlin, Chair of Innovation Economics, Germany

This is a pre-print version of a paper that has been submitted for
publication to a journal.

This version: December 2013

WP 2013/43

**Are there Scale Economies in Scientific Production?
On the Topic of Locally Increasing Returns to Scale**

Torben Schubert

ABSTRACT

In this paper the question of returns to scale in scientific production is analyzed using nonparametric techniques for multidimensional efficiency measurement. Based on survey data for German research groups from three scientific fields it is shown that the multidimensional production possibility sets are weakly non-convex and locally strictly non-convex. This implies that the production functions for the groups in the sample are characterised by increasing returns in some regions and at least constant returns to scale otherwise. This has two implications for organization of scientific research: first, at least some groups in our sample have suboptimal size and could benefit from growing. Second, specialisation on certain tasks in science (e.g. transfer-oriented groups vs. research-oriented groups) would increase output of the overall system.

JEL Code: C14, O30

Keywords: Research Units, Specialisation, Production, Efficiency, Returns to scale, DEA

Disclaimer: All the opinions expressed in this paper are the responsibility of the individual author or authors and do not necessarily represent the views of other CIRCLE researchers.

Are there Scale Economies in Scientific Production?

On the Topic of Locally Increasing Returns to Scale

Torben Schubert

CIRCLE, Lund University

Sölvegatan 16

22100 Lund, Sweden

Email: torben.schubert@circle.lu.se,

Fraunhofer Institute for Systems and Innovation Research

Breslauer Str. 48, 76139 Karlsruhe (Germany)

Also: Technological University Berlin, Chair of Innovation Economics

Abstract: In this paper the question of returns to scale in scientific production is analyzed using nonparametric techniques for multidimensional efficiency measurement. Based on survey data for German research groups from three scientific fields it is shown that the multidimensional production possibility sets are weakly non-convex and locally strictly non-convex. This implies that the production functions for the groups in the sample are characterised by increasing returns in some regions and at least constant returns to scale otherwise. This has two implications for organization of scientific research: first, at least some groups in our sample have suboptimal size and could benefit from growing. Second, specialisation on certain tasks in science (e.g. transfer-oriented groups vs. research-oriented groups) would increase output of the overall system.

Keywords: Research Units, Specialisation, Production, Efficiency, Returns to scale, DEA

JEL: C14, O30

1. Introduction

It is widely believed that efficient regular production systems (e.g. firms or economies) are characterised by division of labour. Indeed, this idea is easily put across when looking at the alternative: the subsistence economy, where everybody produces the goods he needs by himself without trading goods. Although the absence of trade reduces transaction costs, these gains are certainly outweighed by the losses induced by not reaping the benefits of specialisation; in particular, learning effects and accumulation

Although this idea seems natural in the production of regular goods and services, there is little agreement that this insight can be transferred to the production of "scientific goods". Many arguments are put forth to support this scepticism. Some of them are philosophical (for example, the postulated "unity of research and teaching" by Wilhelm von Humboldt, one of the most important university reformers at the beginning of the 19th century in Germany), while some are at least implicitly based on efficiency arguments. For example, sometimes it is assumed that economies of scope exist (Johnes 1997, Cohn et al. 1989), which would render specialisation inefficient.

Many authors have now analysed the question of increasing returns to scale and scope in science (among many others Lloyd et al., 1993, Johnes, 1997, de Groot et al., 1991, Dundar and Lewis, 1995). Though the results with respect to economies of scope are somewhat mixed, the majority of analyses at least at the level of the university demonstrate the existence of increasing returns to scale. These analyses, however, suffer from two limitations. First, most of these studies use aggregated university-level data. Second, employed methodologies build on the estimation of parametrically specified cost-functions which imply not only potentially restrictive functional form assumptions but also assume away the possibility that some research units may be inefficient. This may lead to biases in estimation.

With respect to the first objection, only few papers have analysed the production returns on lower levels of aggregation. One example is Bonaccorsi and Daraio (2005), who use local regression techniques and do not find clear cut evidence of increasing returns to scale. Brandt and Schubert (2013) use parametric regression techniques and find with

respect to publication output increasing returns on the university level and decreasing to constant returns on the level of the research group. This analysis is particularly interesting because it tries to disentangle the production returns on micro and macro-levels of analysis. However, their approach is limited not only by the parametric specification but also by the assumption that scientific output is uni-dimensional – in this case adequately proxied by publications.

The latter two problems can be avoided by the use of nonparametric envelopment estimators. Further, we will show that creative use of these estimators can identify returns to scale characteristics of the production function. Therefore, in this paper we will integrate the neoclassical or parametric understanding of scale economies into a nonparametric framework of efficiency estimation. We analyse whether increasing returns to scale (IRS) exist or not.

If IRS are present, two implications for efficient organisation follow. First, (at least some) research groups would benefit from becoming larger. Second, specialisation of research groups on certain tasks becomes desirable, because a division of labour leads to increased output of the overall system.

The remainder of this article is organised as follows: in Section 2 we will review the literature on returns to scale in science. We will then show under which conditions specialisation is an optimal strategy. In Section 3 we will describe our methodology. Subsequently, in Section 4 the estimation results concerning the returns to scale of scientific production functions are presented. Section 5 concludes.

2. The Nature of Scientific Production

2.1 Returns to scale in scientific production

The literature on returns to scale is largely based on aggregated university level data, where most authors in fact find (at least for some outputs) increasing returns to scale.

This means that groups with IRS can produce over-proportionately more if their inputs are increased.

For example, Worthington and Higgs (2011) find ray economies of scale up to 120 % of the mean in a multi-input, multi-output setting. Comparable results are found by de Groot et al. (1991), Sav (2004), Laband and Lentz (2003), Johnes et al. (2008), as well as Koshal and Koshal (1995). Glass et al. (1995a, b) observe ray economies, but also find product specific economies of scale for undergraduate teaching. Johnes (1999) and Izadi et al. (2002) do not detect ray economies of scale but product-specific economies of scale for undergraduate teaching, postgraduate teaching and research. This is in line with Brandt's and Schubert's (2013) result that there are IRS on the level of the university.

However, they also analyse the returns to scale on the level of the research group and find no evidence of IRS. Indeed the results on the micro-level are less clear, probably also due to the divergence in methods and datasets, where the latter type of analyses is necessarily based on the survey data or case studies.

Van Tunzelmann et al. (2003), who reviewed the existing literature on the level of research groups for size effects on research group productivity, concludes as follows: Evidence across different studies indicates that there appears to be a critical mass threshold for group size, at least in some scientific fields, which hovers around six to eight people. This 'critical mass' threshold may differ among major subject fields, as individual studies show, but no comprehensive picture has emerged so far. A study by Carayol and Matt (2004) focusing on 80 laboratories of the Louis Pasteur University comes to similar conclusions. With regard to the relationship between research group size and the size of the respective department, empirical findings indicate that research groups of sufficient size are able to function well regardless of the size of the department or the university they are affiliated with (Tunzelmann et al. 2003). This latter result, in our terms, indicates the absence of agglomeration effects, while there may be IRS for very low levels of input that turn into DRS if inputs increase. The latter argument would, for example, result from an s-curved cost function. This is congruent with Johnston (1994) to some extent, who, on the level of universities, finds economies of scale for low output

levels and diseconomies of scale for high output levels. The results are mixed, at this level in particular. Adams and Griliches (2000) find constant returns to scale, which implies that size does not matter at all. The same conclusion is drawn by Narin and Hamilton (1996) and by Bonaccorsi and Daraio (2005) for Italian CNR units.

However, all of the analyses have been based either on a cost function or a production function approach. This has the advantage that relatively simple regression techniques can be employed but the disadvantage that these techniques suffer from the parametric assumptions (with the exception of Bonaccorsi and Daraio (2005), who use locally regression techniques) and the elimination of potential inefficiency that the groups might exhibit (with the exception of Johnes (1999) and Izadi et al. (2002), who use parametric frontier estimation). Both types of assumptions can lead to severe estimation bias, if they are not true. We will therefore reinvestigate the topic of IRS and the research group level using more flexible non-parametric frontier estimators, which allow the research groups to display inefficiency in their use of resources.

As argued, the importance of IRS in scientific production derives from two aspects. First and relatively obvious, groups with IRS can produce over-proportionately more if their inputs are increased. More precisely, if all inputs are increased by a factor $\lambda > 1$ then the outputs increase by a factor $\delta > \lambda$ (c.f. Brandt and Schubert 2013 for details). Thus, such groups would benefit from growing. More subtly, if IRS exist among research groups (and total inputs over all groups are fixed), aggregate outputs of all research groups would increase, if they specialised on certain outputs, implying for example that there are graduate teaching or publication or transfer-oriented research groups. Without going into too much technical detail we will explain this latter point now by the means of an illustrative exemplification.

2.2 Specialisation in science

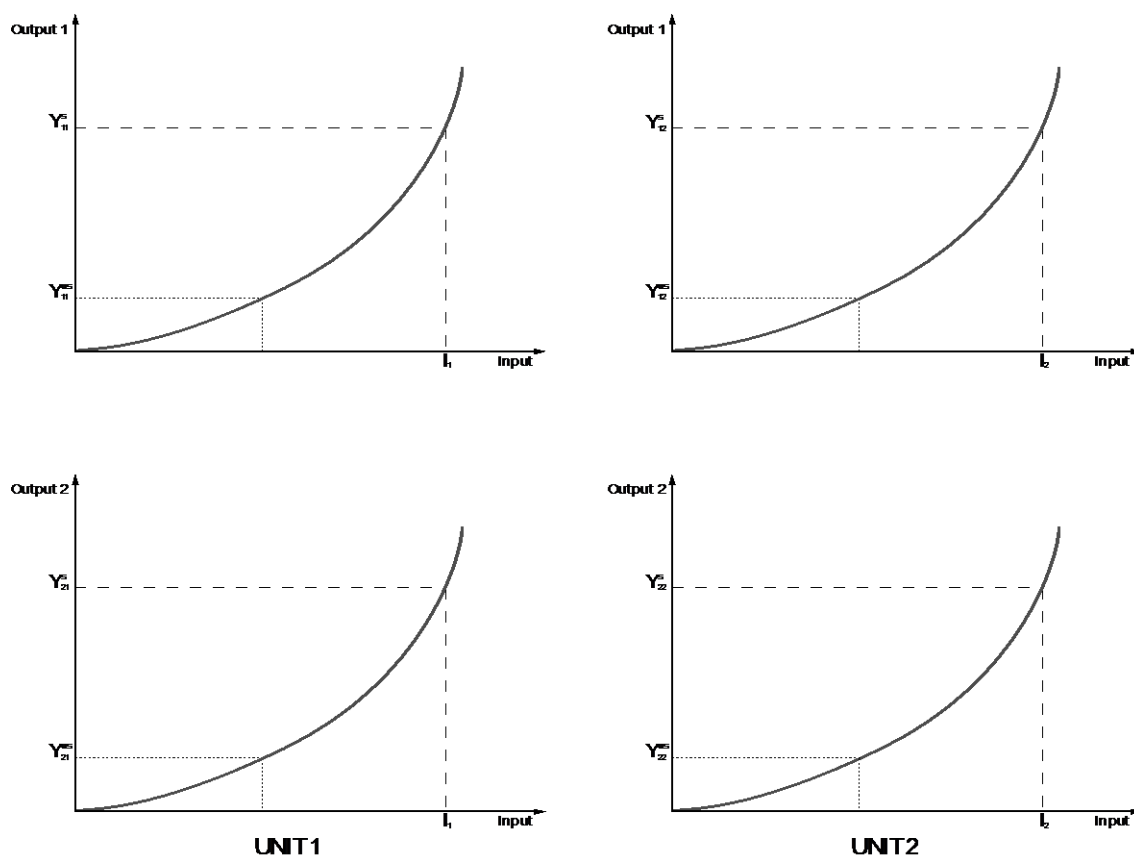
Many authors highlight that scientific production is a process in which manifold inputs (e.g., capital equipment, trained scientists, etc.) are transformed into various outputs

(e.g., publications, patents, knowledge transfer, etc.) (Rousseau and Rousseau 1997; Nagpaul and Roy 2003; Warning 2004; Johnes 2006). This is corroborated for example in Jansen et al. (2007), Schmoch et al. (2010) and Schubert (2009) who show that distinct profiles of production are present in scientific production. In particular, the authors find that research groups fall into characteristic clusters which focus on publication, graduate teaching and transfer activities, which closely resembles the three missions of the universities. Thus, from a descriptive point of view scientific research groups specialise in certain activities.

From a normative point of view of course the question is whether specialisation is a desirable feature in the sense that it makes best possible use of the available resources. It will turn out that the question of optimality is closely linked to the characteristics of the returns to scale.

To gain intuition for this question of specialisation we present a very simple illustrative framework, which consists of just two scientific units which may produce either of two outputs with the help of one input. Each output is produced according to the same production function. The units are identical, both in their technology and in the input equipment. The question then is the following: Should the first unit, say UNIT 1, produce Output 1 but not Output 2 (and UNIT 2 vice versa) or should any of the units produce a bit of both, i.e. should they specialise or not?

Figure 1: Specialisation benefits and curvature of the production function



Looking at Figure 2, the answer on the question of optimality of specialisation crucially depends on the shape of the production function. Here it is assumed that the production functions display strictly increasing returns to scale (they are bowed towards the y-axis). In this case, increasing the input by some constant factor increases output by an amount strictly higher than that factor (see also Section 3.1). For example, assume that if UNIT 1 and UNIT 2 decide to spend 50% of their input on producing Output 1 and 50% on producing Output 2, then the aggregate production of Output 1 and 2 are $Y_{11}^{US} + Y_{12}^{US}$ and $Y_{21}^{US} + Y_{22}^{US}$. If instead UNIT 1 specialised in Output 1 and UNIT 2 in Output 2, then the aggregate production would be Y_{11}^S for Output 1 and Y_{22}^S for Output 2. Since the production functions are convex, we have $Y_{11}^{US} + Y_{12}^{US} < Y_{11}^S$ and $Y_{21}^{US} + Y_{22}^{US} < Y_{22}^S$. Therefore, total specialisation would increase aggregate output. Obviously, if the production functions were concave, then specialisation would be uniformly detrimental calling for a "general-

ist" strategy. If the production functions were linear, then the specialisation strategy would not have any effects.

This illustrative model is obviously quite simplified of course. But it should be clear that the results are fairly robust with respect to several generalisations. First, if the production technologies used by each unit are not identical, then this affects only the question of who should specialise in which dimension (but not if at all). Second, it is unimportant if only two or a multitude of units exists. Third, it is unimportant if we have two or more outputs. Fourth, it does not matter if the units have identical input endowment.¹

Summarising, units should specialise in those outputs where their production functions exhibit increasing returns to scale. With constant returns to scale any specialisation strategy is equally efficient, while decreasing returns argue against specialisation. In Section 4 we will show that the production functions are characterised either by increasing returns or constant returns to scale. Before we do so, we will, however, present some theory about returns to scale in a neoclassical and a nonparametric understanding. The benchmark must be in any case the neoclassical wording, because of the theoretical results of this subsection.

2.3 The hypotheses

The guiding question of this paper is whether scientific production of research groups is characterised by increasing returns to scale in the neoclassical sense. This is the case when the production possibility set is non-convex². Because it was concluded in Section 2.2 that scientific units should pursue a specialisation strategy, if there are IRS in at least some regions and at least constant returns to scale (CRS) everywhere else, a suffi-

¹ Note however, if more than the maximum a certain unit can produce of Output 1, Y_{11}^S , were required by the society, then one unit would specialise and the other would also produce at least some of Output 1, i.e. the second unit would not specialise (unless it produces none of Output 2). Still, in a certain sense, we should speak of a specialisation strategy, because at least one unit is required to specialise and the other does so to the feasible extent, while still meeting the societal demands.

² Convexity of a set means that the linear combination of any two points on the boundary are again part of the set. Increasing returns functions do not possess this property as can easily be checked based on Figure 1.

cient condition is that the production possibility set is strictly non-convex at some points and weakly non-convex anywhere else. We thus hypothesise that:

H0 (decreasing or constant returns to scale everywhere): The production set is weakly convex everywhere.

H1 (increasing returns somewhere and constant returns anywhere else): The production set is strictly non-convex at least somewhere and weakly-non-convex anywhere else.

3. Methodology

If H1 is true IRS exist. We will test this feature by the use of nonparametric efficiency estimators. These will be explained in the following.

3.1 Non-parametric efficiency estimators

The frontier model is explained in the following using the case with one input and one output. This is for expositional reasons only, since the general frontier model can handle technology frontiers with arbitrary dimensionality. One of the most prominent estimators is the data envelopment estimator, which was originally proposed by Charnes et al. (1978). The major limitation of this estimator is however, that it can only deal with situations where the production function is either characterised by non-increasing (DRS) or constant returns to scale (CRS). A production function that is either CRS or DRS can also be said to exhibit non-increasing returns to scale (NIRS). If the production function exhibits increasing returns to scale as in Figure 1, the DEA estimator is biased. Then the free disposal hull estimator (FDH) must be used. The latter estimator is less efficient but unbiased irrespective of the returns to scale. This property will allow for a testing procedure of our hypotheses H0 vs. H1.

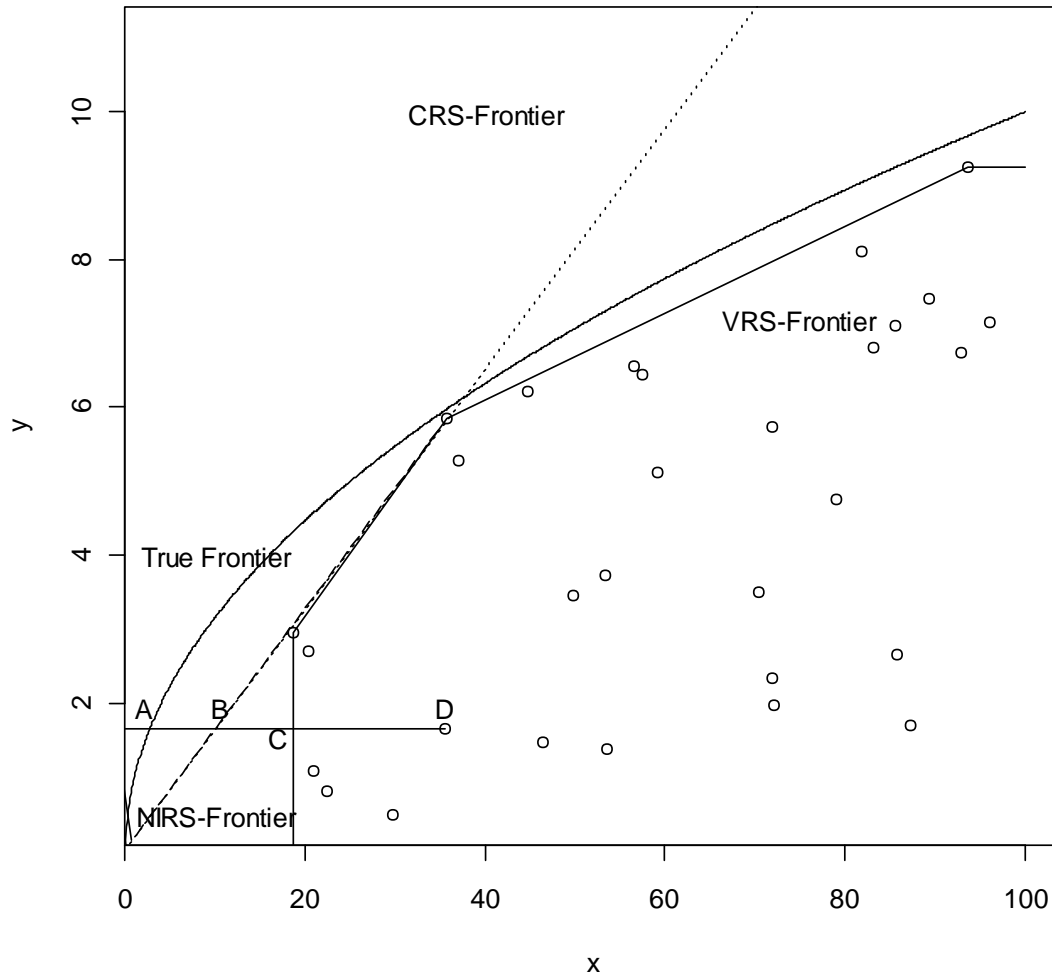
We will now shortly introduce the DEA estimator alongside the general frontier model. We leave aside as much as possible mathematical formulations. DEA is the most commonly used estimator of the procedures in non-parametric efficiency estimation. The idea of efficiency analysis is that a so-called decision-making unit (firms, persons, regions, or research groups) commands a set of inputs to produce certain outputs. The efficiency model assumes that, given a certain amount of input, there is a technological limit to the production of outputs, i.e. a unit cannot produce more than this output. The union of all these maximum points that correspond to a specific input amount is called the theoretical frontier. Units falling short of this theoretical

frontier are inefficient. Inefficiency is usually quantified by radial measures. If the theoretical frontier is observed, the estimation of inefficiency would be a trivial task. However, this is usually not the case. Instead only a given number of units are observed for which we have sample values of inputs and outputs. Using these data points, DEA is one way of estimating the theoretical frontier from the observed data.

In particular, DEA constructs the estimated frontier as the smallest convex hull "enveloping" all data points in the sample. In fact, there are several variants which will be explained later on, but expositional reasons let's start with the so-called VRS DEA frontier (bold line). Consider the one-input-one-output-case depicted in the figure below, where the true frontier is given as $y = \sqrt{x}$. x is taken to be non-random, and inefficiency is generated by $y^{obs} = y \cdot \exp(-|u|)$ with $u \sim N(0,1)$. The small circles give observed sample coordinates for the units. The smallest convex hull that envelops all the data points is the DEA frontier. Obviously, the DEA frontier does not coincide with the theoretical frontier, but if more and more units are observed, the DEA frontier will converge to it (Kneip et. al, 1998). Using this estimated frontier, it is easy to define a measure of inefficiency. In the input-direction this is simply the used input divided by the input needed to provide this output level. Looking at the figure given below, and focussing on the inefficient point D , this is given by the ratio of the length of line segment AD divided by the length of the line segment AC , whereas the true but unobserved inefficiency measure is $|AD|/|AC|$.

Two things are important to note. First, this input-inefficiency measure may take values of 1 and above, because $|AC| \leq |AD|$. Second, a value of 1 indicates that the unit is efficient, because then the unit is on the frontier. Any value greater than 1 indicates inefficiency. Thus low values are desirable. Also note that DEA also works for the multiple-input-multiple-output case. The interpretation of the measures remains the same.

Figure 2: Efficiency and productivity estimation in frontier models



Indeed there is not only one DEA estimator but several variants. The estimator represented by the bold line in Figure 2 is called the VRS estimator. The VRS estimator is the most general estimator. The most restricted estimator is the CRS estimator, which is just a straight line which passes through the origin and the observation that smallest convex which can be defined by a line. In Figure 2 it is characterised by the dotted line. It is also clear that it largely overestimates the true frontier, if this exhibits increasing returns to scale as in our case.

The NIRS estimator (dash-dotted line) additionally to the VRS frontier includes the origin. This makes it effectively a compromise between the VRS and the CRS estimator.

tor. It is identical to the latter up to the first observation on the frontier defining the CRS frontier and beyond that is identical to the VRS frontier.

3.2 Testing for IRS using frontier estimators

All DEA estimators are too restricted to deal with increasing returns to scale production functions, because for DEA to be applicable the production possibility set must be convex. The more general estimator, which can also handle non-convex production possibility sets, is the FDH estimator. Now, under H_0 both the FDH and the NIRS-DEA estimator are consistent, while under H_1 only the FDH estimator is. This implies that under H_0 both estimators should not differ markedly, while, if H_1 were true, they would. Thus our hypothesis from Section 2.3 implies:

H_{0a} : Both the FDH estimator and the NIRS-DEA estimator are consistent.

H_{1a} : Only the FDH estimator is consistent.

If H_{1a} is corroborated, then the production function displays increasing returns to scale at least somewhere. However, this does not mean that it has IRS everywhere. For example, it may also be characterised by DRS in other regions. This would be the case, if for example the production function had IRS for low output levels and DRS for high output levels as for example Johnston (1994) claimed. Under these circumstances, the finding that there are regions with IRS will practically imply nothing with respect to optimal size of the research groups or their specialisation. We thus have to rule out that there are regions with DRS. If this was true as well, then the production function would either be characterised by IRS or at least CRS.

Therefore, in the parlance of frontier estimation we have to show additionally that the production possibility is weakly non-convex (CRS) in the regions where it is not strictly non-convex (IRS), i.e. the efficient boundary is described by constant returns to scales, in which case specialisation is at least not detrimental:

H_{0b} : Both the CRS-DEA estimator and the NIRS-DEA estimator are consistent.

H1b: Only the NIRS-DEA estimator is consistent.

In summary based on H0a-H1b we can distinguish four case. First, when H0a is rejected and H0b is not rejected, then the production possibility set is strictly non-convex at least somewhere, but there is no evidence that it as strictly convex in other places. Consequently, specialisation will increase overall output. Second, when neither H0a nor H0b are rejected, there is no evidence against the hypothesis that the efficient boundary exhibits constant returns to scale everywhere. Both the degree of specialisation as well as group size are irrelevant. Third, when H0a is not rejected but H0b is, then there is evidence that the production possibility set is strictly convex at least somewhere and exhibits constant returns to scale elsewhere. In that case specialisation strategies lower overall output. Fourth, if H0a is rejected and H0b is rejected, then there is evidence that production possibility sets have regions where they are strictly concave and others where they are strictly convex. In that case now generally recommendations with respect to size and specialisation can be drawn.

Table 1: Case definition and recommendations

Case	H0a rejected, but H0b is not rejected	H0a is not rejected and H0b is not rejected	H0a is not rejected but H0b is rejected	H0a is rejected and H0b is rejected
Shape of the production possibility set	Strictly non-convex somewhere and weakly non-convex elsewhere	Both weakly non-convex and weakly convex	Strictly convex somewhere and weakly convex elsewhere	Both regions where it is strictly convex and strictly non-convex
Recommendation	Specialisation	Specialisation	Specialisation	Not generally

	beneficial, larger groups preferable	and size irrel- evant	detrimental, smaller groups prefer- able	valid recom- mendation pos- sible
--	--	--------------------------	---	---

3.3. The Data

In this analysis we use original data from a large online survey (data from 2007) within a research project funded by the German Research Association (DFG). The sample consists of 473 research units from the disciplinary fields of astrophysics, nanotechnology, biotechnology, and economics. This corresponds to a return rate of approximately 25%, as 1908 research units received a questionnaire. With this selection of fields we could guarantee that basic research fields from natural sciences (astrophysics), applied disciplines from natural sciences (biotechnology and nanotechnology), and a field which has both applied and basic research characteristics from the social sciences (economics) is included. The astrophysics makes up about 7% of the sample, nanotechnology about 42%, biotechnology 22% and economics 29%, which approximately reflect the shares in the population. The main objective of the survey was to determine the effects of different university governance models on research efficiency. Because the low sample share of astrophysics units does not allow reliable estimation. We excluded this group from our analysis.

The survey includes information on inputs and outputs of a research group as well as its organizational respectively its governance setting. The data on input and output is clearly at the focus of this paper.

Against the background of multidimensionality of outputs, we collected a variety of different activity indicators, which we regarded as scientific outputs. All in all, we collected the following 11 measures: SCI publications per scientist, citations per publication, conference articles per scientist, fraction of international co-publications, professorial job offers per scientist, expert reports for companies per scientist, cooperation with companies per scientist, membership in

advisory boards per scientist, number of doctoral theses per scientist, number of state doctoral theses per scientist, editorships per scientist.³

3.4 Definition of the Production Possibility Set

As argued by Jansen et al. (2007), Schmoch et al. (2010) and Schmoch and Schubert (2009), scientific outputs should at least consider the dimension of knowledge generation, graduate teaching, and knowledge transfer.

We assume that each of the dimensions, knowledge generation, graduate teaching, and knowledge transfer, can be appropriately represented by a single indicator. To be specific, the indicators number of SCI (bio and nanotech) SCOPUS (economics) publications (knowledge generation), number of finished doctoral/PhD theses (graduate teaching), and number of advisory services for companies plus cooperation with companies (knowledge transfer) were chosen. It was also assumed that the only relevant input is number of scientists. Thus our production possibility set is four dimensional (1 input and 3 outputs).

Yet, while the major tenet of nonparametric efficiency estimation is that these estimators can easily deal with this multidimensionality because the assumed production possibility set is not restricted with respect to the included inputs and outputs. In any case, there are at least two problems of the class of estimators with these estimators relevant for us.

Firstly, because non-parametric estimators make few assumptions which could help identification, they suffer from the curse of dimensionality. This means that convergence becomes very slow as the dimensionality of the production possibility set increases. Thus, if we include many outputs, we will also lower the precision of estimation drastically. To deal with this, apart from an analysis, where all outputs (see next Section) are considered simultaneously, we also run the analyses for each dimension

³ The bibliometric data was taken from the ISI Web of Knowledge for biotechnology and nanotechnology and from Scopus for economics due to better coverage.

separately. Additionally, this gives a more detailed impression of the production possibility set.

Secondly, nonparametric efficiency estimators are very sensible to measurement error, outliers, and model specification. Since the general classification of outputs into knowledge generation, graduate teaching, and knowledge transfer (see Section 2.2) provides only some insight into a sensible choice of indicators, but is far from providing a clear-cut definition we used two alternate sets of indicators.

In specific, we used number of citations of the SCI/SSCI publications, the number of finished habilitation theses, and the number of memberships in scientific advisory boards, where knowledge transfer here more commonly refers to politics than to the industry.

Even though the first set provides a better choice, at least in the author's opinion, if the results stemming from the second should be not too different, then the results can be deemed quite robust.

In order to increase the readability of this paper we discuss the exact testing procedures in the Appendix. The methodologies build on complex bootstrap algorithms and have partly been developed and described in Simar and Wilson (2001, 2002). We will now turn to the results.

4 Results

Some summary statistics for the core variables are presented in Table 2.

Table 2: Core input and output variables

Variable	self-reported	Time Period	Scale/Unit	Mean	S.D.	Min	Max
Research Outputs							
<i>Reputation-linked</i>							
WoS/Scopus Publications	NO	1998-2003	Count	81.67	150.00	0	900
Conference Papers	YES	2001-2002	Count	29.57	47.31	0	300
Received Citations per WoS/Scopus Publications	NO	1999-2001	Rational	4.52	5.80	0	34
Fraction of Internationally Co-authored WoS/Scopus Papers	NO	1998-2003	Percent	42.22	34.19	0	100
Received Professorial Job Offers	YES	2001-2002	Count	0.51	0.81	0	3
<i>Transfer-Linked</i>							
Advisory Services for Companies	YES	2001-2002	Count	0.25	0.86	0	5
Cooperations with Companies	YES	2001-2002	Count	1.94	3.47	0	17
Memberships in Scientific Advisory Boards	YES	2001-2002	Count	3.84	4.86	0	20
<i>Maintenance-Linked</i>							
Doctoral Theses	YES	2001-2002	Count	3.10	3.36	0	15
State Doctoral Theses	YES	2001-2002	Count	0.49	0.79	0	3
Editorships	YES	2001-2002	Count	0.58	1.44	0	10
Conferred Scholarships for Research Group Members	YES	2001-2002	Count	2.29	3.24	0	16
Research Inputs and Structural Variables							
Scientists in Fulltime Equivalents	YES	2001-2002	Real	13.00	20.17	2	150
Research Time Spent on Third Party Projects	YES	2001-2002	Percent	50.76	34.75	0	100

The estimation results are present in Table 3. We see that for any discipline that H0a is rejected and H0b is not rejected. This conclusion hold also holds with the alternative operationalisation of the input output set. Thus, we can be somewhat more confident that our results are not only due to model specification and should hold in a wider context.

According to Table 1 this means that the production possibility set is strictly non-convex somewhere and weakly non-convex elsewhere; or in the language of economics, the efficient boundary exhibits increasing returns to scale somewhere and at least constant returns to scale elsewhere. Therefore, specialisation will increase overall output. It also means that larger research groups can make better use of their resources.

Table 3: Test results of the shape of the production set (full output list)

		H0a against H1a	H0b against H1b
Primary	Nanotechnology	-0.6740**	-0.3962
	Economics	-0.6224**	-0.0016
Definition	Biotechnology	-2.1727**	-1.6516
Secondary	Nanotechnology	-1.3013**	-0.2307
	Economics	-0.8314**	-0.0423
(Robustness)	Biotechnology	-0.0140*	-1.2252

* significant at 5%-level

** significant at 1%-level

When separating this analysis by dimension (Table 4 and Table 5), and having first a closer look on our primary definition, we find a comparable structure, especially in the case of transfer (H0a is rejected, but H0b is not). This holds also for graduate teaching, except for the case of biotechnology, where H0a is not rejected. In knowledge generation H0a is only rejected for biotechnology but not for the other disciplines. The latter is quite in line with the results obtained in Brandt and Schubert (2013), who showed that in the case of knowledge generation as measured by publications IRS on the group level cannot be detected.

Turning to these interesting cases, where H0a is not rejected (remember that this implies that there is no evidence for regions of the efficient boundary which exhibit increasing returns to scale), Table 4 indicates that there is also no evidence for decreasing returns to scale. This means that although specialisation will not increase output, it will at least not decrease it.

However, using the second definition to test for robustness, we see also some differences in Table 4. For example in the case of economics, we cannot detect increasing returns to scale anymore, even though we found them in the full model. Partly, this likely to be due to the fact, that Table 3 provides a joint test on all output dimensions, while Table 4 provides three separated tests and therefore will have lower power. Further, the chosen indicators on graduate teaching and knowledge transfer are unlikely to be a very good choice, because they are comparably rare events. Looking also to the other two

research fields, we find that although IRS are found for the other disciplines, the dimensions along which they occur are not always the same as in the tests based on the primary definitions.

Yet, and this should be stressed, the general result, that there are locally increasing returns to scale but no evidence is found for local decreasing returns, stays the same. So again, we can observe robustness of this general result.

Table 4: Results of the shape of the production set for H0a against H1a (by output dimension)

		Knowledge Generation	Graduate Teaching	Knowledge Transfer
Primary	Nanotechnology	-0.5264	-0.5760**	-0.4413**
	Economics	-0.1776	-0.3665**	-0.4499**
Definition	Biotechnology	-2.9974**	-1.4367	-4.6979**
Secondary	Nanotechnology	-1.0108**	-0.0000	-0.3262
	Economics	-0.0332	-0.0000	-0.0341
(Robustness)	Biotechnology	-2.1200	-0.1563	-5.4461**

* significant at 5%-level

** significant at 1%-level

Table 5: Results of the shape of the production set for H0b against H1b (by output dimension)

		Knowledge Generation	Graduate Teaching	Knowledge Transfer
Primary	Nanotechnology	-0.0565	-0.3471	-0.1286
	Economics	-0.0000	-0.0000	-0.0000
Definition	Biotechnology	-0.9079	-1.2775	-0.1491
Secondary	Nanotechnology	-0.0009	-0.0181	-0.0203
	Economics	-0.0000	-0.0072	-0.0283
(Robustness)	Biotechnology	-1.8933	0.8123	-0.0000

* significant at 5%-level

** significant at 1%-level

Summarising the results from Table 4 and Table 5, we have cases where the efficient boundary exhibits increasing returns to scale at least somewhere and constant returns to scale elsewhere, which calls for specialisation. We also have cases where the efficient boundary is characterised by constant returns to scale everywhere, which implies that specialisation is at least not detrimental. Thus, for the analysed research fields we find strong evidence the existence of locally increasing returns to scale, while we cannot detect locally decreasing returns to scale. Taken together, this calls for an increased emphasis on specialisation of scientific research groups.

5. Conclusions

This paper deals with the optimality of specialisation in science. The empirical methodology used is much more robust than in older research in this field because no parametric production function was "forced upon" the data. Instead, flexible nonparametric estimation techniques were used. To make this approach feasible, it was necessary to integrate the parametric notion of returns to scale and that of nonparametric efficiency analysis. By doing that, it could be shown that the economic intuition which calls for a division of labour carries over from usual production of goods and services also to scientific goods. Specialisation in science will increase aggregate output. We also found that larg-

er research groups will make more efficient use of their resources. This result however only holds for the graduate teaching and transfer. It does not carry over to knowledge generation. This is in line with what has been found in earlier studies (cf. Brandt and Schubert 2013). In any case, it should be noted that great emphasis was placed on system performance, rather than individual rationality. The derived policy recommendation (Increase specialisation!) must still be implemented. Proper measures to achieve that were blanked out. So a *feasible* policy recommendation would be: Foster specialisation of scientific DMUs by setting appropriate financial and non-financial incentives (e.g. specialisation incentives in indicator-oriented budgeting).

Acknowledgement

The research underlying this paper was supported by the German Research Foundation (Deutsche Forschungsgemeinschaft, DFG) in a project on "performance indicators for research groups" (SCHM 1719/1-2) which is part of a larger research group on "international competitiveness and innovation capacity of universities and research institutions – new forms of governance" (FOR 517). I also acknowledge the comments of the reviewers who have helped us to improve this paper.

References

- Brandt, T., Schubert, T. (2013): Is the university model an organizational necessity? Scale and agglomeration effects in science, *Scientometrics*, 94, 541-565
- Bonaccorsi, A., Daraio, C. 2005, Exploring Size and Agglomeration Effects on Public Research Productivity, *Scientometrics*, 63, 87-120
- Carayol, N., & Matt, M. (2004). Does research organization influence academic production? Laboratory level evidence from a large European university. *Research Policy*, 33, 1081–1102.
- Charnes, A., Cooper, W. W., & Rhodes, E. (1978). Measuring the efficiency of decision

making units. *European Journal of Operational Research*, 2, 429-444.

Adams, J. D., & Griliches, Z. (2000). Research productivity in a system of universities. In D. Encaoua et al. (Eds.), *The economics and econometrics of innovation* (pp. 105–140). Dordrecht: Kluwer.

Cohn, E., Rhine, S., Santos, M. 1989, Institutions of higher education as multiproduct firms, economies of scale and scope, *Review of Economics and Statistics*, 71, 284-290

de Groot, H., McMahon, W.W. and Volkwein, J.F., 1991, The cost structure of American research universities. *Review of Economics and Statistics* 73, 424–431

Dundar, H. and Lewis, D.R., 1995, Departmental productivity in American universities, economies of scale and scope. *Economics of Education Review* 14, 119–144

Durbin, J. 1954. Errors in variables, *Review of the International Statistical Institute*, 22, 23–32

Färe, R., Grosskopf, S. 1985, A nonparametric cost approach to scale efficiency, *Journal of Economics*, 87, 594-604

Forsund, F.R., Hjalmarsson, L 2004, Calculating scale elasticities in DEA models, *Journal of the Operational Research Society*, 55, 1023-1038

Glass, J. C., McKillop, D. G., & Hyndman, (1995a). Efficiency in the provision of university teaching and research: an empirical analysis of UK universities. *Journal of Applied Econometrics*, 10, 61–72.

Glass, J. C., McKillop, D. G., & Hyndman, (1995b). The achievement of scale efficiency in UK universities: a multiple-input multiple-output analysis. *Education Economics*, 3, 249–263.

Grosskopf, S. 1986, The role of the reference technology in measuring productive efficiency, *The Economic Journal*, 96, 499-513

- Hanoch, G. 1970, Homotheticity in joint production, *Journal of Economic Theory*, 2, 432-436
- Hausman, J. A. 1978. Specification tests in econometrics, *Econometrica*, 46, 1251-1271
- Izadi, H., Johnes, G., Oskrochi, R., Crouchley, R. 2002, Stochastic frontier estimation of a CES cost function, the case of higher education in Britain, *Economics of Education Review*, 21, 63-71
- Jansen, D., Wald, A., Franke, K., Schmoch, U., Schubert, T. 2007. Third party research funding and performance in research. On the effects of institutional conditions on research performance of teams. *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, 59, 125–149
- Johnes, G. 1997, Costs and industrial structure in contemporary British higher education, *The Economic Journal*, 107, 727-737
- Johnes, G. (1999). The management of universities: Scottish Economic Society/Royal Bank of Scotland Annual Lecture. *Scottish Journal of Political Economy*, 46, 505–522.
- Johnes, G., Johnes, J., & Thanassoulis, E. (2008). An analysis of costs in institutions of higher education in England. *Studies in Higher Education*, 33, 527–549.
- Johnes, J. 2006. Data Envelopment Analysis and its Application to the Measure of Efficiency in Higher Education. *Economics of Education Review*, 25, 273–288
- Lloyd, P.J., Morgan, M.H. and Williams, R.A., 1993, Amalgamation of universities, are there economies of scale or scope?. *Applied Economics* 25, 1081–1092
- Koshal, R. K., & Koshal, M. (1995). Quality and Economics of Scale in Higher Education. *Applied Economics*, 27, 773–778.
- Nagpaul, P., Roy, S. 2003. Constructing a Multi-objective Measure of Research Per-

- formance. *Scientometrics*, 56, 383–402
- Narin, F., & Hamilton, K. S. (1996). Bibliometric performance measures. *Scientometrics*, 36, 293–310.
- Panzar, J.C., Willig, R.D. 1977, Economis of scale in multidimensional output production, *Quarterly Journal of Economics*, 41, 481-493
- Rousseau, S., Rousseau, R. 1997. Data Envelopment Analysis as a Tool for Constructing Scientometric Indicators. *Scientometrics*, 401, 45–56
- Sav, G. T. (2004). Higher education costs and scale and scope economies. *Applied Economics*, 36, 607–614.
- Laband, D. N., & Lentz, B. F. (2003). New estimates of economies of scale and scope in higher education. *Southern Economic Journal*, 70, 172–183.
- Simar, L. Wilson, P.W. 1998, Sensitivity analysis of efficiency scores, How to bootstrap in nonparametric frontier models, *Management Science*, 44, 49-61
- Simar, L., Wilson, P.W. 2001, Testing restrictions in nonparametric frontier models, *Communications in Statistics, Simulation, and Computation*, 30, 161-186
- Simar, L., Wilson, P.W. 2002, Nonparametric test of return to scale, *European Journal of Operational Research*, 139, 115-132
- Schmoch, U., Schubert, T., Jansen, D., Heidler, R., van Go" rtz, R. (2010). How to use Indicators to measure scientific performance? A Balanced Approach. *Research Evaluation*, 19, 2–18.
- Schmoch, U., & Schubert, T. (2009). Sustainability of incentives for excellent research—The German case. *Scientometrics*, 81, 195–218.
- Schubert, T. forthcoming, Empirical Observations on New Public Management to Increase Efficiency in Public Research – Boon or Bane? *Research Policy*.

- Schubert, T., Simar, L. 2009, Testing Restrictions in Production Analysis, Some new-Tests based on Subsampling and Application to the German Mechanical Engineering Sector, *Journal of Productivity Analysis*, 36, 55-69.
- van Tunzelman, N., Ranga, M., Martin, B., & Geuna, A. (2003). The effects of size on research performance: A SPRU review. UK: SPRU, Science and Technology Policy Research Unit, University of Sussex.
- Warning, J. 2004. Performance Differences in German Higher Education, *Empirical Analysis of Strategic Groups. Review of Industrial Organization*, 24, 393–408
- Wu, D.M. 1973. Alternative tests of independence between stochastic regressors and disturbance, *Econometrica*, 42, 529–546
- Worthington, A. C., & Higgs, H. (2011). Economies of scale and scope in Australian higher education. *Higher Education*, 61, 387–414.

Appendix: The Testing Procedure

The testing procedure for H_0b against H_1b has been developed by Simar and Wilson (2001, 2002). However, for the case of H_0a against H_1a a new test has to be proposed. But once the procedure defined by Simar and Wilson (2001, 2002) for the case of restrictions in the DEA model has been explained, it can readily be seen that an immediate extension is possible for the case where the alternative includes a production possibility set that is (possibly) non-convex.

Because under H_0b both the NIRS-DEA and the CRS-DEA estimator are consistent they should be approximately be the same on average, and the test statistic

$$T^{CRS} = \frac{1}{N} \sum_{i=1}^N \left(D_i^{NIRS} - D_i^{CRS} \right) \leq 0$$

where D_i^{VRS} and D_i^{CRS} are the respective efficiency estimates (estimated Shepard's distance function), should tend to zero as the sample size grows large.

In order to assert that T^{CRS} differs significantly from zero, it is necessary to determine its distribution under H0b. One way of doing that is based on the homogeneous bootstrap of Simar and Wilson (1998), where the homogeneity assumes that the distribution of the inefficiency does not depend on the input-output levels. In any case, this assumption can be dropped using subsampling bootstrap, which was done in Schubert and Simar (2009). Since usually differences are small, we leave that topic aside and stick to the homogeneity.

The testing procedure works as follows:

1. For each DMU compute D_i^{VRS} and D_i^{CRS} . Compute T^{CRS} .
2. Use a kernel estimator to obtain an estimate of the density f^{CRS} of D_i^{CRS} .⁴
3. Draw random deviates γ_i^* from f^{CRS} .
4. Calculate a pseudo-sample as follows $(\gamma_i^* x_i / D_i^{CRS}, y_i)$, where x_i / D_i^{CRS} is the input-oriented projection on the efficient frontier, when H0b is true.⁵
5. From this pseudo-sample calculate $D_i^{VRS^*}$ and $D_i^{CRS^*}$. Also compute T^{CRS^*} .
6. Repeat steps 3-5 B times.

⁴ This will be done by the usual way proposed by Simar and Wilson (1998) to calculate confidence intervals for efficiency measures in the DEA-model. The exact procedure is complicated but is implemented in the FEAR-package for R by P. W. Wilson. The function is called `dea.resample`.

⁵ Projecting the observed point onto the efficient frontier when H0b is true ensures that T^{CRS^*} is sampled as if H0b were true.

7. From the resulting simulated distribution of T^{CRS*} calculate the approximate p-value of the test-statistic T^{CRS} as follows: $pval \approx \sum_{j=1}^B 1[T_j^{*CRS} \leq T^{CRS}] / B$.

It is readily seen that the test procedure does not make any assumptions about the shape of the production possibility set when the null-hypothesis is not true. Therefore it is straightforward to obtain a test for H0a against H1a. Simply replace the test statistic by $T^{VRS} = N^{-1} \sum_{i=1}^N (D_i^{FDH} - D_i^{VRS}) \leq 0$, calculate f^{VRS} in step 2 instead, form the pseudo-sample by $(\gamma_i^* x_i / D_i^{VRS}, y_i)$, and calculate the p-value as $pval \approx \sum_{j=1}^B 1[T_j^{*VRS} \leq T^{VRS}] / B$. Everything else remains unchanged.