

Are there scale economies in scientific production? On the topic of locally increasing returns to scale

Torben Schubert

Received: 24 May 2013 / Published online: 18 December 2013
© Akadémiai Kiadó, Budapest, Hungary 2013

Abstract In this paper the question of returns to scale in scientific production is analysed using non-parametric techniques of 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 suggests that the production functions for the groups in the sample are characterised by increasing returns to scale in some regions and at least constant returns to scale otherwise. This has two implications for the organisation of scientific research: first, the size of at least some groups in our sample is suboptimal and they would benefit from growth. Second, greater specialisation in certain tasks in science (e.g. transfer-oriented groups vs. research-oriented groups) would increase the output of the overall system.

Keywords Research units · Specialisation · Production · Efficiency · Returns to scale · DEA

JEL Classification C14 · O30

Introduction

It is widely believed that efficient regular production systems (e.g. firms or economies) are characterised by the division of labour. Indeed, this principle is easily communicated when

T. Schubert (✉)
CIRCLE, Lund University, Sölvegatan 16, 22100 Lund, Sweden
e-mail: torben.schubert@isi.fhg.de; torben.schubert@circle.lu.se

T. Schubert
Fraunhofer Institute for Systems and Innovation Research, Breslauer Str. 48, 76139 Karlsruhe, Germany

T. Schubert
Technological University Berlin, Berlin, Germany

looking at the alternative: a subsistence economy, where each person produces the goods they need without trading. Although the absence of trade reduces transaction costs, these gains are certainly outweighed by the losses from not reaping the benefits of specialisation; in particular, learning effects and accumulation.

Although this principle seems natural for the production of regular goods and services, there is little agreement whether this insight can be simply transferred to the production of “scientific goods”. There are many arguments to support the 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 nineteenth century in Germany), while some are at least implicitly based on efficiency arguments. For example, it is sometimes assumed that economies of scope exist (Johnes 1997; Cohn et al. 1989) which would render specialisation inefficient.

Many authors have since then analysed the question of increasing returns to scale (IRS) and scope in science (among many others Lloyd et al. 1993; Johnes 1997; de Groot et al. 1991; Dundar and Lewis 1995). Although the results with respect to economies of scope are mixed, the majority of analyses, at least at the level of universities, demonstrate the existence of IRS. However, these analyses suffer from two limitations. First, most of these studies use aggregated university-level data. Second, the methodologies employed 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 estimation biases.

So far only a few papers have analysed the production returns at lower levels of aggregation. One example is the paper by Bonaccorsi and Daraio (2005), who use local regression techniques and do not find clear-cut evidence of IRS. Brandt and Schubert (2013) use parametric regression techniques and find increasing returns on the university level and decreasing to constant returns on the level of the research group with respect to publication output. This analysis is particularly interesting because it tries to disentangle the production returns on a micro and a macro level of analysis. However, their approach is limited not only by the parametric specification, but also by the assumption that scientific output is unidimensional taking publications as an adequate proxy.

The latter two problems can be avoided by the use of non-parametric 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 integrate the neoclassical or parametric understanding of scale economies into a non-parametric framework of efficiency estimation. We analyse whether IRS exist or not.

If IRS are present, two implications for efficient organisation follow. First, (at least some) research groups would benefit from an increase in size. Second, the specialisation of research groups in 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 “[The nature of scientific production](#)” section we review the literature on returns to scale in science. We then show under which conditions specialisation is an optimal strategy. “[Methodology](#)” section describes our methodology. “[Results](#)” section presents the estimation results concerning the returns to scale of scientific production functions. “[Conclusions and limitations](#)” section concludes.

The nature of scientific production

Returns to scale in scientific production

The literature on returns to scale is largely based on aggregated university level data, where most authors find (at least for some outputs) IRS. This means that groups with IRS can produce disproportionately 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 IRS exist at the level of the university.

However, the latter also analysed the returns to scale at 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 analyses are necessarily based on survey data or case studies.

van Tunzelman et al. (2003), who reviewed the existing literature on the level of research groups for size effects on research group productivity, conclude 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 (van Tunzelman 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 decreasing returns to scale (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 (CRS), 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 were 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 parametric assumptions (with the exception of Bonaccorsi and Daraio (2005), who use local regression techniques) and eliminate any potential inefficiency that the groups might exhibit (with the exception of Johnes (1999) and Izadi et al. (2002), who use parametric frontier estimation). Both assumptions can lead to severe estimation bias if they are not true. We therefore reinvestigate the topic of IRS at 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 obviously, groups with IRS can produce disproportionately 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 $\delta > \lambda$ (cf. Brandt and Schubert 2013 for details). Thus, such groups would benefit from growth. More subtly, if IRS exist among research groups (and total inputs over all groups are fixed), the aggregate outputs of all research groups would increase if they specialised in 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 using an illustrative example.

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; Johns 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, these authors find that research groups fall into characteristic clusters which focus on publication, graduate teaching or transfer activities, which closely resemble the three missions of universities. Thus, from a descriptive point of view, scientific research groups tend to 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 the best possible use of the available resources. It turns out that the question of optimality is closely linked to the characteristics of the returns to scale.

To enhance the understanding of this question of specialisation, we present a very simple illustrative framework consisting of just two scientific units which have one input and can use it to produce either of two outputs. Each output is produced according to the same production function. The units are identical in their technology and input equipment. The question asked is the following: Should the first unit, UNIT 1, produce Output 1 but not Output 2 (and UNIT 2 vice versa), or should each unit produce a bit of both, i.e. should they specialise or not?

Looking at Fig. 2, the answer to the question of the optimality of specialisation crucially depends on the shape of the production function. Here, it is assumed that the production functions display IRS (they are bowed towards the y-axis). In this case, increasing the input by a constant factor increases the output by an amount strictly higher than that factor (see also “[Non-parametric efficiency estimators](#)” section). For example, if it is assumed that 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 specialises 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, specialisation would increase aggregate output. Obviously, if the production functions were concave, then specialisation would be uniformly detrimental calling for a “generalist” strategy. If the production functions were linear, then the specialisation strategy would not have any effect.

This illustrative model is obviously quite simplified. 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 whether they should specialise 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

To summarise, units should specialise in those outputs where their production functions exhibit IRS. With CRS, any specialisation strategy is equally efficient, while decreasing returns argue against specialisation. In “[Results](#)” section we will show that the production functions are characterised either by increasing returns or CRS. Before we do so, we outline the theory about returns to scale in a neoclassical and non-parametric understanding, where the focus will be on the neoclassical one because of the results of this subsection.

The hypotheses

The guiding question of this paper is whether the scientific production of research groups is characterised by IRS in the neoclassical sense. This is the case when the production possibility set is non-convex.² Because it was concluded in “[Specialisation in science](#)” section that scientific units should pursue a specialisation strategy if there are IRS in at least some regions and at least CRS everywhere else, a sufficient 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 everywhere else*) The production set is strictly non-convex at least at some points and weakly non-convex everywhere else.

Methodology

If H1 is true, IRS exist. We test this hypothesis using non-parametric efficiency estimators. These are explained below.

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 (DEA), which was originally proposed by Charnes et al.

¹ Note, however, if more than the maximum a certain unit can produce of Output 1, were required by 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). In a certain sense, we should still speak of a specialisation strategy in this case, because at least one unit is required to specialise and the other does so to the feasible extent, while still meeting 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 be easily verified based on Fig. 1.

(1978). The major limitation of this estimator is that it can only deal with situations where the production function is either characterised by 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 IRS as in Fig. 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 allows us to use it to test our hypotheses H_0 versus H_1 .

We now introduce the DEA estimator alongside the general frontier model. We leave out mathematical formulations as much as possible. DEA is the most commonly used estimator in non-parametric efficiency estimation. The idea behind efficiency analysis is that a so-called decision-making unit (firms, persons, regions, or research groups) commands a certain 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 was always 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 the data points in a sample. In fact, there are several variants which will be explained later on but, for expositional reasons, we start with the so-called VRS DEA frontier (bold line). Consider the one-input one-output case depicted in the Fig. 2, where

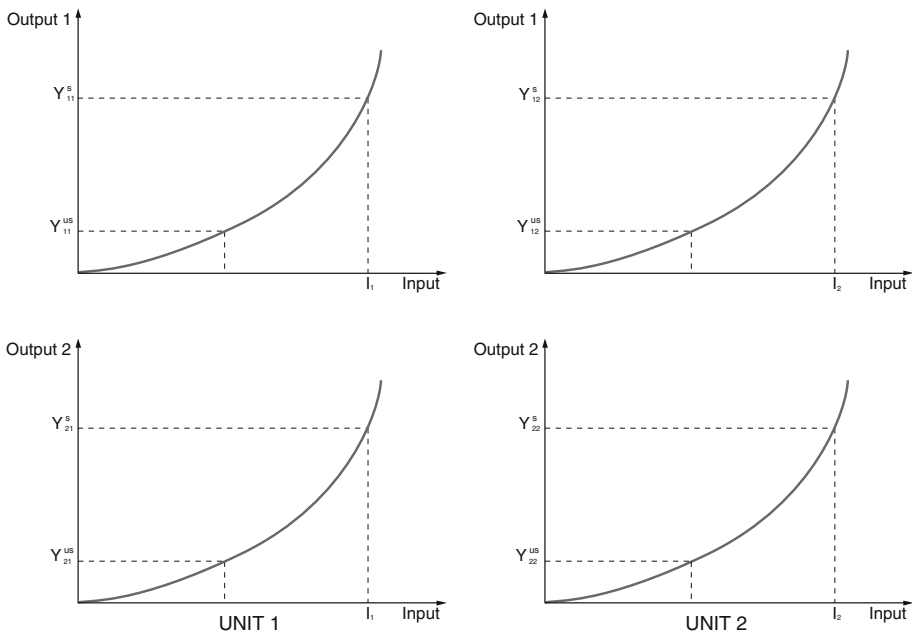


Fig. 1 Specialisation benefits and curvature of the production function

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 mark 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. With regard to input, this is simply the used input divided by the input needed to provide this output level. Looking at the figure given below, and focusing 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 $|AD|/|AC|$.

Three things are important to note. First, this input-inefficiency measure may take values of 1 and above. Second, a value of 1 indicates that the unit is efficient, because then the unit is on the frontier. Any value >1 indicates inefficiency. Thus low values are desirable. Third, note that DEA also works for the multiple-input, multiple-output case. The interpretation of the measures remains the same.

Indeed there is not only one DEA estimator but several variants. The estimator represented by the bold line in Fig. 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 passing through the origin and most outward observation. In Fig. 2 it is indicated by the dotted line. It is also clear that it largely overestimates the true frontier if this exhibits IRS as in our case.

The NDRS estimator (dash-dotted line) is like the VRS estimator but additionally includes the origin. Effectively this makes it a compromise between the VRS and the CRS estimators. 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.

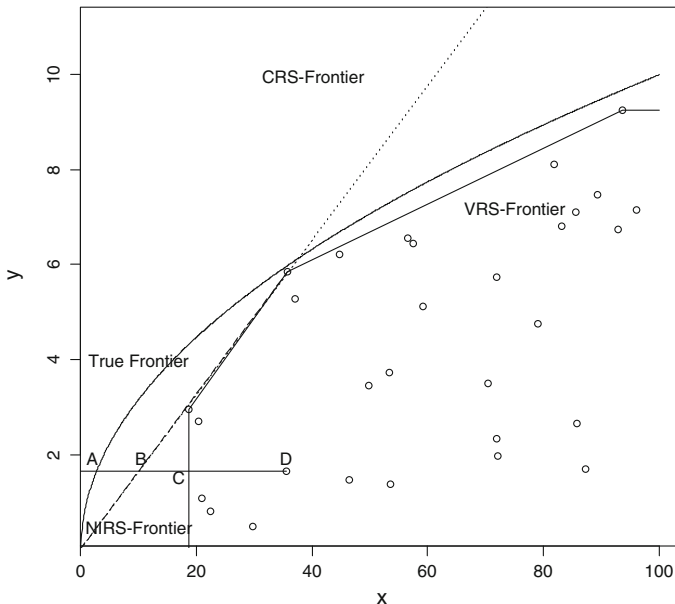


Fig. 2 Efficiency and productivity estimation in frontier models

Testing for IRS using frontier estimators

All DEA estimators are too restricted to deal with IRS production functions, because the production possibility set must be convex for DEA to be applicable. 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 only the FDH estimator is consistent under H_1 . This implies that, under H_0 , the two estimators should not differ markedly, while, if H_1 were true, they would. Thus our hypothesis from “[The hypotheses](#)” section implies:

H0a Both the FDH estimator and the NIRS-DEA estimator are consistent.

H1a Only the FDH estimator is consistent.

If H_{1a} is corroborated, then the production function displays IRS 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 the production function had IRS for low output levels and DRS for high output levels as claimed, for example, by Johnston (1994). Under these circumstances, the finding that there are regions with IRS would imply practically nothing with respect to the 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 additionally show 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 CRSs, in which case specialisation is at least not detrimental:

H0b Both the CRS–DEA estimator and the NIRS–DEA estimator are consistent.

H1b Only the NIRS–DEA estimator is consistent.

In summary, based on H_{0a} – H_{1b} , we can distinguish four cases. First, when H_{0a} is rejected and H_{0b} is not rejected, then the production possibility set is strictly non-convex at least somewhere, but there is no evidence that it is strictly convex in other places. Consequently, specialisation will increase overall output. Second, when neither H_{0a} nor H_{0b} are rejected, there is no evidence to refute the hypothesis that the efficient boundary exhibits CRS everywhere. Both the degree of specialisation and group size are irrelevant. Third, when H_{0a} is not rejected but H_{0b} is, then there is evidence that the production possibility set is strictly convex at least somewhere and exhibits CRS elsewhere. In this case, specialisation strategies lower overall output. Fourth, if H_{0a} and H_{0b} are rejected, there is evidence that production possibility sets have regions where they are strictly concave and others where they are strictly convex. In this case, general recommendations cannot be derived with respect to size and specialisation.

The data

In this analysis we use original data from a large online survey (data from 2007) conducted as part of 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 $\sim 25\%$, as 1908 research units received a questionnaire. This selection of fields guaranteed the

inclusion of basic research fields from the natural sciences (astrophysics), applied disciplines from the natural sciences (biotechnology and nanotechnology), and a field which has both applied and basic research characteristics from the social sciences (economics). Astrophysics makes up about 7 % of the sample, nanotechnology about 42 %, biotechnology 22 % and economics 29 %, which correspond roughly to 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 the inputs and outputs of a research group, as well as its organisational or governance setting. This paper focuses on the input and output data.

Against the background of the multidimensionality of outputs, we collected a variety of different activity indicators, which we regarded as scientific outputs. In total, 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

Definition of the production possibility set

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

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

The major tenet of non-parametric efficiency estimation is that these estimators can easily deal with multidimensionality, because the assumed production possibility set is not restricted with respect to the included inputs and outputs. However, for us there are at least two relevant problems concerning the class of estimators.

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 multiple outputs, we also lower the precision of estimation drastically. To deal with this, apart from an analysis where all the outputs (see next section) are considered simultaneously, we also run the analyses for each dimension separately. This also provides a more detailed picture of the production possibility set.

Secondly, non-parametric efficiency estimators are very sensitive to measurement error, outliers, and model specification. Since the general classification of outputs into knowledge generation, graduate teaching, and knowledge transfer (see “[Specialisation in science](#)” section) provides only some insights into a sensible choice of indicators, but is far from being a clear-cut definition, we used two alternative sets of indicators.

³ The bibliometric data were taken from the ISI Web of Knowledge for biotechnology and nano-technology and from Scopus for economics due to its better coverage.

This alternative includes the number of citations of the SCOPUS/SCI/SSCI publications, the number of completed habilitation theses, and the number of memberships in scientific advisory boards, where knowledge transfer more commonly refers to policy making than to industry.

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

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

Results

Some summary statistics for the core variables are presented in Table 1.

The estimation results are presented in Table 2. We see that H0a is rejected and H0b is not rejected for every discipline. This conclusion also holds with the alternative operationalisation of the input–output set. Thus, we can be more confident that our results are not only due to the specific model specification and should also hold in a wider context.

According to Table 3, 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 IRS somewhere and at least CRS elsewhere. Therefore, specialisation will increase overall output. It also means that larger research groups can make better use of their resources.

When separating this analysis by dimension (Tables 4, 5), and taking a closer look at our primary definition, we find a comparable structure, especially in the case of transfer (H0a is rejected, but H0b is not). This also holds for graduate teaching, except for the case of biotechnology, where H0a is not rejected. In knowledge generation, H0a is rejected only for biotechnology but not for the other disciplines. The latter finding is 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 cannot be detected on the group level.

Looking at the interesting cases where H0a is not rejected (remember that this implies there is no evidence for regions along the efficient boundary which exhibit IRS), Table 4 indicates that there is also no evidence for DRS. 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 also see some differences in Table 4. For example, in the case of economics, we can no longer detect IRS, even though we found them in the full model. This may be partly due to the fact that Table 2 presents a joint test on all the output dimensions, while Table 4 shows three separated tests and therefore has lower power. Further, the chosen indicators on graduate teaching and knowledge transfer are probably not a very good choice, because they are comparatively rare events. Looking at the other two research fields as well, 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 IRS but no evidence for locally decreasing returns remains the same. So, again, we observe the robustness of this general result.

Table 1 Core input and output variables

Variable	Self-reported	Time period	Scale/unit	Mean	SD	Min	Max
Research outputs							
Reputation-linked							
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
Transfer-linked							
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
Maintenance-linked							
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

Table 2 Test results of the shape of the production set (full output list)

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

* Significant at 5 % level; ** significant at 1 % level

Table 3 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	Regions where it is strictly convex and others where it is strictly non-convex
Recommendation	Specialisation beneficial, larger groups preferable	Specialisation and size irrelevant	Specialisation detrimental, smaller groups preferable	No generally valid recommendation possible

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

	Knowledge generation	Graduate teaching	Knowledge transfer
Primary definition			
Nanotechnology	−0.5264	−0.5760**	−0.4413**
Economics	−0.1776	−0.3665**	−0.4499**
Biotechnology	−2.9974**	−1.4367	−4.6979**
Secondary definition (robustness)			
Nanotechnology	−1.0108**	−0.0000	−0.3262
Economics	−0.0332	−0.0000	−0.0341
Biotechnology	−2.1200	−0.1563	−5.4461**

* Significant at 5 % level; ** significant at 1 % level

Summarising the results from Tables 4 and 5, we find cases where the efficient boundary exhibits IRS at least somewhere and CRS elsewhere, which calls for specialisation. We also have cases where the efficient boundary is characterised by CRS everywhere, which implies that specialisation is at least not detrimental. Thus, for the analysed research fields, we find strong evidence for the existence of locally IRS, while we cannot

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

	Knowledge generation	Graduate teaching	Knowledge transfer
Primary definition			
Nanotechnology	-0.0565	-0.3471	-0.1286
Economics	-0.0000	-0.0000	-0.0000
Biotechnology	-0.9079	-1.2775	-0.1491
Secondary definition (robustness)			
Nanotechnology	-0.0009	-0.0181	-0.0203
Economics	-0.0000	-0.0072	-0.0283
Biotechnology	-1.8933	0.8123	-0.0000

* Significant at 5 % level; ** significant at 1 % level

detect locally DRS. Taken together, this calls for the increased specialisation of scientific research groups.

Conclusions and limitations

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 non-parametric estimation techniques were used. To make this approach feasible, it was necessary to integrate the parametric notion of returns to scale and that of non-parametric efficiency analysis. By doing so, it could be shown that the economic intuition which calls for a division of labour can be transferred from the traditional production of goods and services to scientific goods as well. Specialisation in science will increase aggregate output. We also found that larger research groups will make more efficient use of their resources. In terms of policy recommendations, this paper suggests that specialisation should be increased and that resources should be concentrated in larger research groups. This could be achieved, for example, by creating appropriate incentives in the resource allocation system (e.g. indicator-based funding schemes).

Two qualifiers are necessary with respect to the major implications. First, this result only holds for graduate teaching and transfer. It does not carry over to knowledge generation, which is in line with what has been found in earlier studies (cf. Brandt and Schubert 2013). Thus the general claim of the increased efficiency of larger units only holds for a subset of the activities. Second, the observation that there are IRS might be due to reverse causality for the following reason: typically better units are more able to attract new funding and, inherently, more efficient firms might be more likely to grow. Thus, unobserved heterogeneity could be the driver of this result, rather than actual IRS. While this is definitely a limitation of this study, it should be a problem of small samples where there are groups which are efficient and small. This will be true for example, when the groups are still young, implying that they have not yet grown strongly. Since there are few relatively young units in the sample, it seems reasonable to assume that reverse causality issues have limited impact. Nonetheless, it would only be possible to fully control for this effect if panel data are used to capture the full dynamics. Since the dataset used here is a purely cross-sectional one, this remains an issue for future research.

Acknowledgments 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 the “international competitiveness and innovation capacity of universities and research institutions—new forms of governance” (FOR 517). I am also grateful to the reviewers whose comments have helped to improve this paper.

Appendix: the testing procedure

The testing procedure for H0b against H1b was developed by Simar and Wilson (2001, 2002). However, for the case of H0a against H1a, 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 H0b, both the NIRS-DEA and the CRS-DEA estimator are consistent, they should be approximately the same on average, and the test statistic

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

where $D_i^{\widehat{\text{VRS}}}$ and $D_i^{\widehat{\text{CRS}}}$ are the respective efficiency estimates (estimated Shepard’s distance function) should tend to zero as the sample size increases.

In order to assert that T^{CRS} differs significantly from zero, it is necessary to determine its distribution under H0b. One way of doing this 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 differences are usually small, we omit this topic and focus on the homogeneity.

The testing procedure works as follows:

1. For each DMU compute $D_i^{\widehat{\text{NIRS}}}$ and $D_i^{\widehat{\text{CRS}}}$. Compute T^{CRS} .
2. Use a kernel estimator to obtain an estimate of the density $f^{\widehat{\text{CRS}}}$ of $D_i^{\widehat{\text{CRS}}}$.⁴
3. Draw random deviates γ_i^* from $f^{\widehat{\text{CRS}}}$.
4. Calculate a pseudo-sample as follows $(\gamma_i^* x_i / D_i^{\widehat{\text{CRS}}}, y_i)$, where $x_i / D_i^{\widehat{\text{CRS}}}$ is the input-oriented projection on the efficient frontier when H0b is true.⁵
5. From this pseudo-sample, calculate $D_i^{\widehat{\text{NIRS}^*}}$ and $D_i^{\widehat{\text{CRS}^*}}$. Also compute T^{CRS^*} .
6. Repeat steps 3–5 B times.
7. From the resulting simulated distribution of T^{CRS^*} calculate the approximate p value of the test-statistic T^{CRS} as follows: $pval \approx \sum_{j=1}^B 1 [T_j^{\text{CRS}^*} \leq T^{\text{CRS}}] / B$.

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

⁴ This will be done in 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.

$T^{\text{VRS}} = N^{-1} \sum_{i=1}^N (\widehat{D}_i^{\text{FDH}} - \widehat{D}_i^{\text{VRS}}) \leq 0$, calculate f^{VRS} in step 2 instead, form the pseudo-sample by $(\gamma_i^* x_i / D_i^{\text{VRS}}, y_i)$, and calculate the p value as $pval \approx \sum_{j=1}^B 1 [T_j^{*\text{VRS}} \leq T^{\text{VRS}}] / B$. Everything else remains unchanged.

References

- 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.
- Bonaccorsi, A., & Daraio, C. (2005). Exploring size and agglomeration effects on public research productivity. *Scientometrics*, 63, 87–120.
- Brandt, T., & Schubert, T. (2013). Is the university model an organizational necessity? Scale and agglomeration effects in science. *Scientometrics*, 94, 541–565.
- 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.
- 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., & Volkwein, J. F. (1991). The cost structure of American research universities. *Review of Economics and Statistics*, 73, 424–431.
- Dundar, H., & Lewis, D. R. (1995). Departmental productivity in American universities, economies of scale and scope. *Economics of Education Review*, 14, 119–144.
- 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.
- 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, J. (2006). Data envelopment analysis and its application to the measure of efficiency in higher education. *Economics of Education Review*, 25, 273–288.
- 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.
- Johnston, R. (1994). Effects of resource concentration on research performance. *Higher Education*, 28, 25–37.
- Kneip, A., Park, B., & Simar, L. (1998) A note on the convergence of nonparametric DEA efficiency measures. *Econometric Theory*, 14, 783–793.
- Koshal, R. K., & Koshal, M. (1995). Quality and economics of scale in higher education. *Applied Economics*, 27, 773–778.
- Laband, D. N., & Lentz, B. F. (2003). New estimates of economies of scale and scope in higher education. *Southern Economic Journal*, 70, 172–183.
- Lloyd, P. J., Morgan, M. H., & Williams, R. A. (1993). Amalgamation of universities, are there economies of scale or scope? *Applied Economics*, 25, 1081–1092.
- Nagpaul, P., & Roy, S. (2003). Constructing a multi-objective measure of research performance. *Scientometrics*, 56, 383–402.
- Narin, F., & Hamilton, K. S. (1996). Bibliometric performance measures. *Scientometrics*, 36, 293–310.
- 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.

- Schmoch, U., & Schubert, T. (2009). Sustainability of incentives for excellent research—The German case. *Scientometrics*, *81*, 195–218.
- Schmoch, U., Schubert, T., Jansen, D., Heidler, R., & van Gö rtz, R. (2010). How to use indicators to measure scientific performance? A balanced approach. *Research Evaluation*, *19*, 2–18.
- Schubert, T. (2009). Empirical observations on new public management to increase efficiency in public research—Boon or bane? *Research Policy*, *38*, 1225–1234.
- Schubert, T., & Simar, L. (2009). Testing restrictions in production analysis, some newtests based on subsampling and application to the German Mechanical Engineering Sector. *Journal of Productivity Analysis*, *36*, 55–69.
- 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.
- van Tunzelman, N., Ranga, M., Martin, B., & Geuna, A. (2003). *The effects of size on research performance: A SPRU review*. Brighton, 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.
- Worthington, A. C., & Higgs, H. (2011). Economies of scale and scope in Australian higher education. *Higher Education*, *61*, 387–414.