Is the university model an organizational necessity? Scale and agglomeration effects in science

Tasso Brandt · Torben Schubert

Abstract   In this paper we argue that the emergence of the dominant model of university organization, which is characterized by a large agglomeration of many (often loosely affiliated) small research groups, might have an economic explanation that relates to the features of the scientific production process. In particular, we argue that there are decreasing returns to scale on the level of the individual research groups, which prevent them from becoming to large, while we argue for positive agglomeration effects on the supra-research-group-level inside the university. As a consequence an efficient university organization would precisely consist of tying together many small individual research groups without merging them. Basing our empirical analysis on a multilevel dataset for German research institutes from four disciplines we are able to find strong support for the presence of these effects. This suggests that the emergence of the dominant model of university organization may also be the result of these particular features of the production process, where the least we can say is that this model is under the given circumstances highly efficient.

Keywords   Agglomeration effects · Scientific production · Returns to scale · University organization · Efficiency

Introduction

In almost every country, universities are the major places where basic research is conducted. A dominant organizational design has emerged here, according to which universities can be thought of as (usually) large agglomerations of relatively small individual research groups which are linked by an overarching, yet relatively weak, management level. To facilitate our argumentation we will call this the "standard university model".

T. Brandt · T. Schubert (✉)
Fraunhofer ISI CIRCLE, University of Lund, Lund, Sweden
e-mail: torben.schubert@isi.fraunhofer.de

T. Brandt
e-mail: brandt@ifsberlin.de

Acknowledging that this organizational structure is the result of historical and path-dependent processes, the question remains whether its development can also be explained in terms of economic pressures for efficiency. Or, in other words, is the university (set up as a large co-location of small research groups) an efficient organizational design?

This is an important question for several reasons. First of all, the adequacy of the organizational structures is probably a strong determinant of the "fitness" of an organization in an evolutionary process of survival (Nelson and Winter 1982; Hodgson 1993). In this respect, this dominant organization design may have emerged as the outcome of a continuous evolutionary process that eventually selected the best organizational structure for science, in line with Fisher's Fundamental Theorem of Natural Selection (Mulder et al. 2001). Therefore, analyzing the adequacy of the organizational structures of universities could pave the way for understanding (some of) the driving forces behind their organizational evolution.

Second, understanding the characteristics of the "scientific production process"[1] not only allows us to identify explanations for the observed organizational structure, but can also provide normative insights into how universities should be organized. In particular, we will be able to draw conclusions about the "big science" or "critical mass" hypotheses which have, with varying degrees of subtlety, affected the discussion about the allocation of resources in science by promoting the increasing concentration of resources.

To investigate these issues, we analyze the production-related returns to scale on the level of the research group and university-wide agglomeration effects, where the latter reflect the degree of co-location economies (Marshall 1890). We focus on the task of knowledge production which is measured using bibliometric indicators leaving aside other scientific goods, such as education or knowledge transfer.

Acknowledging that the latter restriction does not allow us to analyze the economies of scope in the production of distinct products, there is an innovation in this paper, which we regard as important. The literature dealing with returns to scale usually tries to derive statements about the optimal size of universities (see amongst others Cohen 1991; Seglen and Aksnes 2000; Cohn et al. 1989). We instead intend to draw conclusions about the optimality of the very organizational model of universities at different levels of analysis (in particular the university level vs. the research group level).

We base our analysis on three combined data sets for the years 2007–2009. The first data set is from a micro-level survey of inputs, outputs, and governance issues for German research groups in the fields of astrophysics, biotechnology, nanotechnology, and economics. This data was gathered in a research project funded by the Deutsche Forschungsgemeinschaft (FOR 517). The second source provides macro-level information about the structural characteristics of the corresponding universities, such as size. This data was collected during the course of the EU-funded EUMIDA project. The third set consists of bibliometric data for the research groups and the universities compiled from Thomson Reuter's Web of Knowledge.

Based on this data, we find decreasing returns to scale (DRS) at the production-related level of the research group. At the same time, we find positive agglomeration effects on their productivity if they are located in a large university. The latter can be interpreted as an indication of increasing returns to scale (IRS) with respect to the university level that might be due to shared infrastructure.

---

[1] We will drop the quotation marks from now on.

We argue that this constellation favors a holding-type organization, because this allows individual research groups to enjoy the university-level IRS while remaining small and thereby evading the production-related decreasing returns to scale. As a corollary, the concentration of resources on a few research groups may be detrimental, while the concentration on a few universities may be beneficial.

The remainder of this paper is organized as follows. Section "Theory and empirical evidence from the literature" first explains the concept of returns to scale and its impact on the organizational model. Then we give a short account of the literature on returns to scale in scientific production. Based on this, we highlight the necessity to fully model the production process in order to credibly analyze returns to scale. By applying a multilevel Input-Process-Output-model (IPO-model), we construct a theoretical framework of a multilevel production process in science, on which we base our empirical strategy. In "Methodology" we describe the methodology and the variable selection and present data and results in "Results". Section "Conclusions" concludes.

Theory and empirical evidence from the literature

The standard university model: conditions of efficiency

In the following we will derive conditions of efficiency for the standard university model. We will define these in terms of returns to scale at the research group level and in agglomeration effects at the university level. Before we do so, however, we will briefly explain the concepts of returns to scale and agglomeration effects.

Suppose there are many units of production—e.g. a firm producing chairs, or a university producing papers—where each unit provides one homogenous good using one or more homogenous inputs. If we intend to minimize the overall production costs for any desired level of output, we should ask whether it would be better to have just one/a few large production units supplying this output, or many small ones. If there are size advantages in production we will prefer large production units, while, in the opposite case, smaller would be better. Production-related size advantages are called economies of scale or increasing returns to scale (IRS). In the same way, production-related size disadvantages are called diseconomies of scale or decreasing returns to scale (DRS). The concept of agglomeration effects is closely connected with this and determines the degree to which it is advisable to *spatially and/or organizationally co-locate* individual research groups. Both concepts have in common that they relate size to the ability to effectively provide output. However, the first defines size in relation to the size of the producing unit (here the research group), while the latter refers to the "horizontal size" of the overarching organization framework—here the university (compare also Church and Ware 1999).

Economies of scale are often the result of learning or specialization advantages, meaning that workers become more experienced and more productive the more often they have participated in the production process. IRS commonly also result from so-called indivisibilities in the inputs, which means that some inputs (e.g. large machinery) cannot be scaled: a unit must buy the whole machine (rather than have half a machine, which is completely useless) and the unit should be large enough to make full use of its capacity.

Agglomeration effects in science, on the contrary, are not directly linked to the production taking place within the research group. Rather they can be interpreted as the production-relevant effects of co-location or clustering in a spatial meaning. Sometimes they are also called external returns to scale (Marshall 1890). In this respect a university is

understood as the co-location or agglomeration of multiple research groups.[2] Once again agglomeration advantages can result from shared physical or management infrastructure. Alternatively, there may be knowledge spillovers among the individual research groups (Anderson et al. 2004; Krugman 1991).

Now turning to our topic of the efficiency of university organization, suppose, for example, that we observe DRS at the production level (i.e. production should take place in smaller units) but agglomeration effects at the university level. Then this constellation would favor large agglomerations of research groups that are themselves relatively small, because this evades the production-related diseconomies of scale while achieving agglomeration advantages by increasing "horizontal size". It is interesting to see that related arguments have been made by Hannan and Mavinga (1980), Lewis and Webb (2007) as well as Lo and Lu (2006) in the context of the banking sector, contending that a major benefit of holding companies is their profiting from economies of scale and scope. These companies are organized much like universities: they are the owners of a large portfolio of smaller companies that are themselves neither interrelated nor actively steered by the holding.

Accordingly, the observable university model in science can be considered economically efficient if we detect diseconomies of scale associated with the actual production in the research groups accompanied by agglomeration effects at the level of the university. At the same time, this would also provide an explanation for the prevailing university model in economic terms.

Returns of scale and agglomeration effects in science

In the past, the existence of size advantages in science has been disputed. The main rationale for assuming that larger research groups are more productive (i.e. assuming IRS) is that the members of larger groups profit from the pooling of intellectual, budgetary and time resources, receive greater stimulus for research ideas through interactions with colleagues with similar research interests, and also have a higher chance of generating 'in-house' the critical mass required for successful applications for research grants; also, larger research groups are more likely to attract high-quality researchers (see, e.g. Tunzelmann et al. 2003; Kyvik 1995).

As we will see, the empirical results on this score are mixed which is why we develop a more systematic perspective of this issue.

We have already highlighted that whether or not size effects exist depends on the importance of production-related specialization (resulting from the division of labor), learning advantages, communication and administration costs, indivisibilities (Bonaccorsi and Daraio 2005), and the diversity of the accessible knowledge base. In the following, we will discuss some analytical considerations that guided our modeling approach.

Since returns to scale from specialization and learning are directly related to the production process, they primarily affect the research group level. Their importance as an argument for IRS depends on the degree to which learning and specialization is possible. This will be higher when acquired knowledge can be re-used, particularly when production routines are established that are repeated over and over again. In this context, the major productivity increase accompanying Fordist production lines in 1920s was only possible, because the activities consisted of easily repeatable manual tasks. Theoretically this has

---

[2] In fact, this is more than a mere co-location because it provides a common organizational and legal frame for all the affiliated research groups. However, this shall not concern us here.

already been implied by what Marx (1867, p. 369) called the "virtuosity of detail worker" (original wording "Virtuosität des Detailarbeiters").

However, this is obviously unlikely to be the case for scientific research. Scientists rarely specialize in a routine task that they continuously repeat and thereby become ever more productive. Furthermore, because science is a field that is subject to perennial change in the knowledge base, learning effects are limited. Acquired knowledge usually cannot be re-used without being modified in some way. A further argument by Bonaccorsi and Daraio (2005) claims that the specialization argument does not easily extend to science, because publication activities allow a cognitive division of labor which functions irrespective of whether scientists are located in the same organizational unit or not. Nonetheless, there of course exists some degree of specialization advantages. As an example, in social science research more quantitatively and more qualitatively oriented researchers may fruitfully work together, each of them excelling in its own core competences. Yet we believe that the potential for IRS on the research group level resulting from specialization and learning is existent but due to the lower degree of routine works still limited than in more traditional domains of economic activity.

We now turn to the potential costs and benefits associated with communication and administration. In this regard, small research groups—rather than large ones—may benefit from the greater individuality and creativity of their researchers, being less hampered by administrative burdens and coordination processes (see, e.g. Tunzelmann et al. 2003; Kyvik 1995). These costs are probably quite low at the level of the university, because this represents merely an overarching organizational umbrella with only limited interaction with the research groups, implying that the communication and organizational requirements will only increase slightly when an additional research group is established. On the other side, this also argues against considerable IRS stemming from shared communication and administration.

A further source of IRS and agglomeration advantages is related to indivisibilities, which means that inputs are not arbitrarily scalable. This argument is usually made with respect to physical infrastructure: Usually, it does not make sense to purchase half a radio telescope, because this is not operational. Consequently, if the units only have the option to buy one indivisible radio telescope, then the group should be large enough to make full use of it. This suggests the existence of a minimum efficient size.

Whether this argument carries, depends first of all on the discipline involved. While it might constitute a viable source of IRS for more capital-intensive disciplines, this may be negligible for disciplines that are less dependent on physical infrastructure.

Furthermore, even with capital-intensive disciplines, dealing with indivisibilities can also be done by pooling resources at the level of the university or at least of the faculty. There is no necessity to establish large research groups. In this respect, we observe that large-scale infrastructure is commonly used by many institutes or even many universities (instead of making one research group big enough to own it by itself). An illustrative example can be found in the large lunar observatories (e.g. the SALT in South Africa) that are frequented by scientists from all over the world. Another example is bibliometric databases that are usually subscribed to on the level of the university. In line with the argument of shared infrastructure, size advantages in the form of returns to scale or agglomeration effects are probably more important on the level of the university than on that of the research group.

We summarize these arguments in Table 1, where "−" indicates strong diseconomies of scale while "+++" refers to strong economies of scale.

Source	Research group level IRS	University level agglomeration advantages
Specialization advantages/ learning effects	+	0
Costs	–	0/–
Indivisibilities	+	++
Knowledge diversity/ weak ties	0	+++
Hypothesized overall effect	Decreasing returns to scale	Positive agglomeration effects

Table 1 Size advantages in science by level

A final source of size advantages is the existence of diversity in the knowledge base. It is well known from organizational learning theories (March 1991) that learning potentials arise from diversity in the knowledge base. The idea is remarkably simple. One cannot learn from someone who knows exactly the same. Therefore, more distant knowledge sources are highly important, even though they may be often irrelevant. This line of argument is already reflected in Granovetter's (1973) strength of weak ties in personal networks. However, because distant knowledge bases are unlikely to occur in close units such ties are much more important on the level of the university than they are on the level of the research group.

Looking at Table 1, we expect that, at the research group level, the costs of being large probably outweigh the positive size effects in the form of indivisibilities. The opposite is likely to be true for the university level. Overall, we expect DRS at the research group level and agglomeration advantages at the university level.

Some results from the literature

The results with respect to returns to scale in science are, as already mentioned, somewhat mixed. Furthermore, the literature so far has primarily focused on the university level, where the usual approach is to collect cost and output data at the *university level* and then estimate cost functions.[3] A few exceptions to this can be found in Tunzelmann et al. (2003), who reviewed the existing literature for size effects on research group productivity: 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

---

[3] Conditions for the IRS/DRS and agglomeration advantages/disadvantages in terms of cost and production functions are given in Sect. 3.

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. Yet 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. It should be noted that the latter work is based on the production rather than the cost function approach (see "Methodology"). While this is less conventional, it comes much closer to what we will present in this paper.

Others like 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. The latter two papers use a stochastic frontier approach instead of the usual cost function approach.

To sum up, there is little agreement in the literature. Furthermore, the results are also obscured by differences in the levels of analysis: Most authors focus on the university while some conduct their analysis at the level of the research group. Taking universities as the observation unit implies that the university is the locus of production which differs from our approach where the research group is the production unit and the university only provides an administrative frame. We believe our approach to be better suited because it gives a more accurate description of how university research is organized. This also complies with Laredo and Mustar (2001), who argue in the same direction.

Furthermore, much of the existing literature neglects confounding variables. These should be controlled for, irrespective of whether we analyze the university or the research group level. Important potential confounders will be identified in the next section, where our discussion is based on an augmented Input-Process-Output-Model (IPO), drawing on elements of organizational theory and an economic production perspective.

(Other) determinants of research productivity: a multilevel, multifactor model

IPO models have been extensively used to describe the relevant influences of team effectiveness of production units (for an overview see, e.g. West et al. 1998; Salas et al. 2007). What makes these models so valuable is the particular attention paid to the context that working groups are subject to, where the models categorize the variables into input/structures, processes, and outputs.

However, as useful as the IPO models may be for variable selection, they usually lack an explicit recognition of the production process. Thus, we augment the IPO model paying special attention to what economists call the characteristics of the "production technology", which determine the process by which inputs are transformed into outputs. This is important for our analysis, because returns to scale are also characteristics of the "production technology".

Looking at the actual production process. we follow the concept in Laredo and Mustar (2001) and stress that it is the research group level at which production takes place in reality. Looking at the individual level would neglect collaboration among team members or the dependence on equipment provided by the research group. The organizational level of

universities, on the other hand, seems too remote, since at this level we are forced to sum inputs and outputs across the boundaries of disciplines, which provide diverse production logics.

In summary, the model that we describe in the following is an IPO model augmented by an economic production perspective. Figure 1 provides a schematic overview of an IPO model of the productivity of research groups, based on the more generic discussion of team effectiveness in firms (see, e.g. Salas et al. 2007; West et al. 1998). It is important to note that our aim here is not to provide a comprehensive review of the literature but rather to outline a heuristic for identifying central influential factors on research groups' perfor-mance. The model depicts inputs, processes and outputs as three interrelated variable complexes, while differentiating between three major analytical levels which influence research groups' scientific productivity: (a) individual level characteristics and related cognition processes, (b) group level characteristics and processes and (c) organization level structures and processes.

As can be seen, we deliberately allow this model to be very data demanding, in the sense that it draws on information at various institutional levels. In this respect, the model may well exceed the limits of many existing data sets, including our own. These empirical limitations notwithstanding, the depicted framework should be useful in drawing attention to many of the important determinants of research productivity inside *and* outside the realm of the research group. Based on this model, we will, in a second step, empirically analyze the influence of economies of scale and agglomeration effects as well as other potential factors on research group bibliometric productivity.

What, then, are the potential determinants of research outputs? As previous research has revealed, scientific productivity is influenced by a variety of factors at different analytical levels, including the individual characteristics of researchers, research teams and groups, and also broader contextual and organizational level factors (for an overview of previous
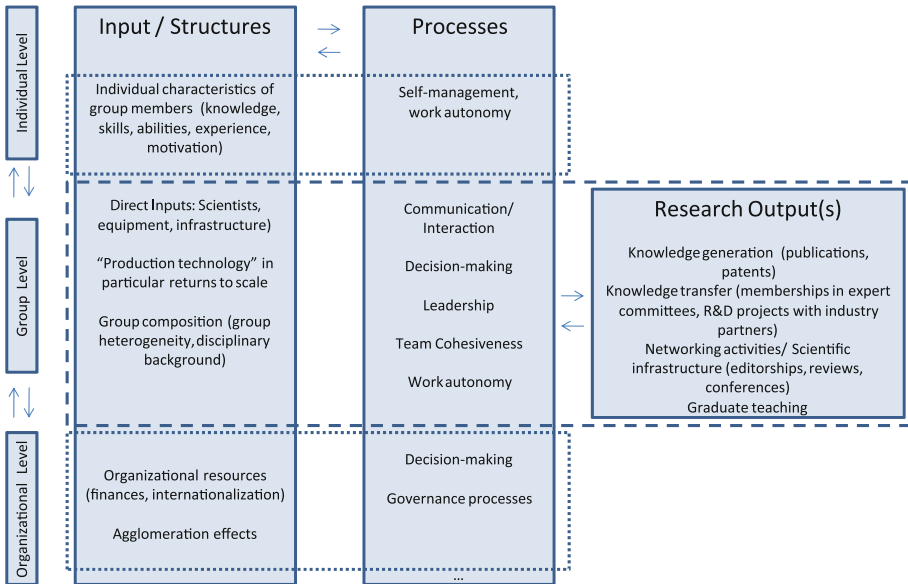
Fig. 1 Ideal IPO model for analyzing the research output of research groups. Source: extended and adapted based on West et al. (1998, p. 157) and Salas et al. (2007, p. 216)

studies on research productivity see, e.g. Libaers 2006; Dundar and Lewis 1998; Carayol and Matt 2004; Hemlin and Gustafsson 1996). Despite the reasonable progress that has been made in understanding the importance of these factors for productivity in science, the majority of studies in the field have either focused on individual researcher's performance, or the performance of research institutions or universities—the latter mainly reflecting the discussion of university rankings (Dundar and Lewis 1998). Only a few have examined the productivity of research groups (see e.g. Andrews 1979; Carayol and Matt 2004; Seglen and Aksnes 2000; Schmoch and Schubert 2009; Schubert 2009).

At the various levels depicted in Fig. 1 several influencing factors have been taken into consideration as evidenced in the literature on organizational theory and on research productivity (see, e.g. Salas et al. 2007; West et al. 1998; Libaers 2006; Dundar and Lewis 1998).

Looking at characteristics at the level of individuals, researchers' knowledge, skills, abilities, experience and motivations are probably the most important *input factors* in explaining publication output.

At the level of the research group, we have to make a distinction between direct inputs and variables that describe structures. The former relate, e.g. to the number of scientists in the research project, as well as to the available equipment and infrastructure. The others encompass mediating characteristics such as disciplinary heterogeneity in the research group (group composition). From a production perspective, the difference between direct inputs and mediating characteristics is that the former should per definition positively influence output, while the latter may have positive, negative, or neutral impacts. We introduce this seemingly artificial distinction here already, because it will be crucially important when explaining our methodology.

In any case, while there is a general consensus that the mediating variables stemming from group composition are also relevant for predicting team effectiveness, research in this area is limited (West 2004; an exception is e.g. Rey-Rocha et al. 2006). In this context, e.g. the question still has to be analyzed whether the heterogeneity or homogeneity of team members with regard to age, publication experience and disciplinary background positively fosters publication output, and the right mix of team members may vary in different contexts.

In addition to these variables, we introduce the "production technology", which defines the "technological law" by which inputs are transformed into outputs. As already mentioned, this includes the returns to scale as a prominent factor which determines the degree of size advantages at the production level.

As a third relevant variable complex, the overall organizational context (i.e. the university or research institution in which a group is embedded) can provide a supportive environment for publication activities. In this context, resources play a role such as the infrastructure as well as motivation-related factors such as reward systems and the overall organization culture regarding publication support, including feedback and training (see, e.g. Salas et al. 2007). Also, as a central factor, agglomeration effects reflecting university-level resources come into play here such as shared infrastructure or management capabilities.

Another important aspect is also the university-wide governance structure because this strongly influences the way the research group may operate. In the context of the New Public Management reforms, these governance structures determine, for example, to what degree the research group may shift its financial resources between personnel and capital expenditures (Schubert 2009). This also determines the degree to which a research group is subjected to competitive pressure due to reliance on third party funds (Jansen et al. 2007; Schmoch et al. 2010). Therefore, this variable complex is a decisive factor with respect to our main question about size effects, since the governance framework also determines how much universities or research groups may grow via financial incentive schemes.

Governance structures are closely related to the relevant *process variables,* which refer to the interaction processes which influence teams while trying to achieve their goals. Here, again, processes at different levels come into play. These include individual aspects such as self-management and the work autonomy of individual researchers as well as factors at group level such as the quality of communication/interaction between team members and decision-making processes (or the lack thereof). Team cohesiveness (i.e. the identification of team members with the group) and team leadership also play an important role in fostering team output (see, e.g. Salas et al. 2007; West et al. 1998). Also, work autonomy of the team members should be a significant factor, presuming that teams are motivated towards publishing and that teams with a high level of autonomy are able to put more effort into tasks which they consider highly motivating (West et al. 1998, 2004). Individual and group-level processes, in turn, may be influenced by overall organizational decision-making and governance processes.

With regard to research outputs, our focus here is on analyzing bibliometric productivity as a central indicator of knowledge generation. From a theoretical perspective, it is important to note that scientific performance in an overall perspective is a multidimensional concept and can hardly be capture by only one single metric. Drawing on previous research (Jansen et al. 2007; Schmoch and Schubert 2009), we have identified the most relevant dimensions to be generation of new knowledge, graduate training, transfer of knowledge to society, and activities that help to maintain the scientific community. In general, *research group outputs* are not one-dimensional but multivariate. In this paper, however, we will focus on knowledge generation as measured by bibliometric measures.

The important question raised by this paper is how the returns to scale (as a feature of the production technology) and the agglomeration effects on the university level (which define the "horizontal size" of the research organization) affect bibliometric performance, where we control for various influencing factors that can be delineated from Fig. 1.

We proceed by defining agglomeration effects and returns to scale more formally. This will lay the basis for the estimation in "Results".

Methodology

Our methodology of measuring returns to scale is based on the notion of a scientific production function similar to that proposed by Griliches (1979) or Mairesse and Mohnen (2002). According to this concept, scientific goods (in our case new knowledge) are produced by using inputs, most prominently researchers. Furthermore, the implied input–output transformation may be mediated by other control variables, for example, the governance framework.

Using the concept of the production function is very appealing because it allows us to directly define returns to scale at the level of the university and the level of the research group. This can be done based on the production function itself, or on the cost function it implies. In this section, we first define returns to scale more formally. Then we derive a suitable production function and describe the estimation strategy. We then go on to select variables based on the preceding discussion.

Defining returns to scale and agglomeration effects

There is a direct analogy between the cost and the production function: Assuming the production function is characterized by diseconomies of scale, additional units of inputs will induce increases in output that then taper off if the inputs continue to increase. This

Table 2 Returns to scale and the analogy between the production and the cost function

	Production function	Cost function
Decreasing returns to scale	$f(\lambda x_i) < \lambda f(x_i)$	$c(\lambda y_i) > \lambda c(y_i)$
Increasing returns to scale	$f(\lambda x_i) > \lambda f(x_i)$	$c(\lambda y_i) < \lambda c(y_i)$
Constant returns to scale	$f(\lambda x_i) = \lambda f(x_i)$	$c(\lambda y_i) = \lambda c(y_i)$

implies that the production function is concave. Likewise, the corresponding cost function must be convex, because with any desired increase in output, the costs grow disproportionately to the already achieved volume of production. The reverse is true for increasing returns to scale.

If $y_i = f(x_i)$ is the production function of research group i for a given level of inputs $x_i = (x_{i1}, \ldots, x_{ij})$, $c(y_i)$ the associated cost function for a given level of output y_i, and $\lambda > 1$ a scaling constant, we can summarize the above arguments as shown in the following table:

Taking the IRS case, for example, the mathematical relationships in Table 2 simply state that increasing all inputs by a constant factor (say to a level of 120 % of the original input, implying $\lambda = 1.2$) will result in a disproportionately large output (we might have an output level that is 130 %). With respect to costs, a similar argument implies that increasing the desired level of outputs to 120 % will induce costs that are lower than 120 % of the costs associated with the original output level.

With respect to agglomeration effects, the focus is somewhat different. Agglomeration advantages mean that individual research groups become more productive by increasing the "horizontal size" of the university. Define $x_{-i} = \sum_{k=1, k \neq i}^{n} x_k$ as the input of all other research groups in the university. Furthermore, let x^*_{-i} and x^{**}_{-i} be two specific input levels such that $x^*_{-i} > x^{**}_{-i}$, agglomeration effects can be defined as in Table 3.

Again this is just a formal definition of a simple relationship. With respect to positive agglomeration effects, the formula states that, with a given level of input, a research group will produce more if it is part of a large university. Likewise, with any given level of output, the associated costs will be lower.

In order to estimate the returns to scale and the agglomeration effects, we have to analyze the properties of either the cost or the production function. In the literature, the usual approach has been based on cost functions, although two exceptions are Bonaccorsi and Daraio (2005) and Jansen et al. (2007) who focused on the production function. In any case, because the methods are theoretically equivalent, the choice usually follows arguments of practicability. In our case, we do not have cost information in the data set which is why we base our analysis directly on the production function.

Constructing a production function

Relying on the production function approach to simultaneously test returns to scale and agglomeration effects requires us to specify a functional form that can capture the salient

Table 3 Agglomeration effects and the analogy between the production and the cost function

	Production function	Cost function
Negative agglomeration effects	$f(x_i, x^*_{-i}) < f(x_i, x^{**}_{-i})$	$c(y_i; x^*_{-i}) > c(y_i; x^{**}_{-i})$
Positive agglomeration effects	$f(x_i, x^*_{-i}) > f(x_i, x^{**}_{-i})$	$c(y_i; x^*_{-i}) < c(y_i; x^{**}_{-i})$
Neutral agglomeration effects	$f(x_i, x^*_{-i}) = f(x_i, x^{**}_{-i})$	$c(y_i; x^*_{-i}) = c(y_i; x^{**}_{-i})$

features of the production process. This process will be affected by the inputs and additional control variables, where we take the research group size in terms of researchers in FTE as the direct input. It is worth noting that university size in our way of modelling is actually a control variable and not an input, because it does not reflect resources that are directly at the disposal of the research group.

We also need to specify the functional form of production, which should be flexible enough to allow increasing as well as decreasing returns to scale. A common choice is the Cobb-Douglas function[4] and "Appropriateness of the Cobb-Douglas specification" presents evidence for its validity here. This function is given as follows:

$$y = (\text{research group size})^{\beta} \cdot (\text{university size})^{\gamma_1} \cdot \prod_{j=2}^{k} (\text{control variable}_j)^{\gamma_j} \qquad (1)$$

where we have to estimate the coefficients β and $\gamma = (\gamma_1, \ldots, \gamma_k)$.

Based on Table 2, it is easy to show that this production function exhibits decreasing returns to scale with respect to the size of the research group if and only if $\beta < 1$.[5] There are increasing returns to scale if $\beta > 1$.

Looking at Table 3, agglomeration effects occur whenever the research group benefits from being part of a large university so that we have to test whether $\gamma_1 = 0$. If this is the case, the size of the university no longer plays any role after controlling for research group size and other factors. If $\gamma_1 > 0$, then being located in a large university is beneficial; if $\gamma_1 < 0$ it is better to be located in a small university which implies agglomeration disadvantages. Yet there is also a further meaning: If $0 < \gamma_1 < 1$ then university size effects exist but taper off as the university gets larger and larger. If $\gamma_1 > 0$, the university size advantages are self-intensifying.

In summary, our expectations of increasing agglomeration advantages on the university level and decreasing returns to scale on the research group level directly translate into $\gamma_1 > 0$ and $(0 <) \beta < 1$.

In principal, estimating Eq. (1) calls for a regression approach. Since the model is non-linear, one way to implement this would be non-linear least squares. However, this may be cumbersome because of numerical convergence problems. In our case it is easier to log-transform Eq. (1). In error form, we end up with a model which has linear parameters that can be estimated using regular techniques:

$$\ln y = \beta \ln(\text{research group size}) + \ln(\text{university size}) + \sum_{j=2}^{k} \gamma_j \ln(\text{control variable}) + u$$

$$(2)$$

While this could principally be estimated using ordinary least squares, we have to account for observations with zeros in the output variables. These would be omitted. Thus we add 1 to the output before taking the log, which leads to mass-points at zero for the independent variables. With this data structure, Tobit regression models are most

---

[4] We have also experimented with even more flexible production functions. In particular we used a production function that follows the law of diminishing returns. This allows returns to scale to be modeled that depend on the level of input. This production function in practice, however, just imitated the Cobb-Douglas function in the relevant region.

[5] Of course we expect that $\beta > 0$ because only then leads a higher number of researchers to more output.

appropriate. As a point of reference it should be mentioned that corresponding ordinary regression calculations delivered very similar results.

Selection of variables

Our analysis is framed by our data, which combine two unique data sets supplemented by bibliometric information from Thomson Reuter's Web of Knowledge. The first source consists of information from a micro-level survey of German research units conducted during February and March 2007. In a lengthy process, we were able to identify 1908 university chairs and corresponding extra-university units from the disciplinary fields of astrophysics, nanotechnology, biotechnology, and economics. Choosing this particular set of disciplines was inspired by trying to achieve the most dissimilar case design (for further discussion see Schmoch and Schubert 2008). The first divide is along the dimension of natural sciences (astrophysics, bio- and nanotechnology) vs. social sciences (economics). However, the second divide is along the dimension of basic (astrophysics, economics) versus applied research (bio- and nanotechnology). By March we had received 473 valid answers (astrophysics: 34, nanotechnology: 201, biotechnology: 136, economics: 102), which implies a participation rate of almost 25 %. 140 of the respondents were from (public) extra-university units belonging to institutions such as the Max Planck or the Fraunhofer Society. The remaining 333 were from universities that were included in this analysis. This survey provides information about inputs (in particular the number of scientists in FTE and the share of time spent in research activities), outputs (publications in the Web of Knowledge as well as received citations), governance structures as well as some team descriptors. Data on individual universities are based on a comprehensive data set collected within the framework of the EUMIDA project, which was financed by the European Commission (European Commission 2010; see http://www.ec.europa.eu/research/era/docs/en/eumida-final-report.pdf [August 2011]). The overall EUMIDA data set includes about 2 457 institutions in the EU-27 countries (minus Denmark and France) plus Norway and Switzerland, covering approximately 90 % of the total number of students in tertiary education (ISCED levels 5 and 6). In the case of Germany, 409 institutions (universities as well as universities of applied science) were included. For these institutions, a core set of indicators has been developed, which includes basic institutional descriptors (country, legal status, foundation year, total staff etc.), indicators of educational activities (e.g. students at ISCED 5 and ISCED 6 levels, subject specialization, highest degree delivered), research activities (number of doctorates) and international attractiveness (number of international students at ISCED 5 and ISCED 6 levels), among others. Most importantly for our analysis, this dataset contains the second size variable of interest: the number of university staff in FTE.

Definition of research group

The definition of the research group deserves some attention. While, in principal, a production-oriented definition would identify a research group as individuals who collaborate on a day-to-day basis in order produce research results, in the survey we had to choose a definition that was operational without knowing the detailed workflows of the units. We therefore identified research groups as the individual chairs that are usually led by one professor. Although of course it is possible that the chair itself consists of thematic groups, which might in some cases come closer to a production-oriented definition of a research

group, we believe that in the majority of cases, the chair as an observational unit will serve to delimit the research group quite well.

How do our data fit within the IPO model?

We stress that the model depicted above outlines something like a wish-list of variables that would be ideally incorporated in a comprehensive analysis of the productivity of research groups, while recognizing that existing data sets often fall short of providing information for all the relevant variables. This is also true for our data set. Going back to Fig. 1, ideally we should have information on all dimensions and analytical levels regarding (a) inputs/structures, (b) processes, and (c) outputs. Table 4 provides an over-view of the variables included in our analyses.

   With regard to research *outputs*, our empirical approach focuses on research pro-ductivity in the form of publications as a central measure of knowledge generation. In particular, we will use the (?) metrics *publications* and *publications per scientists* in order to account for quantity aspects of research as well as *citations* in order to account for impact aspects. We recognize that there is a large body of literature dealing with the appropriateness of these indicators. However, we will put this aside here, because these indicators are very well established and widely accepted as measures of performance—at least for the disciplines surveyed here. This may cause a problem because the scientists (our central input variable) are not fully available for research tasks but may, for example, be engaged in teaching activities. However, we can solve this problem because data is available on the average share of time that is spent on research in each research group.

   Looking at the potential determinants of bibliometric output, our data provide ample information in the *input/structures domain* with regard to group-related resources as well as organizational resources and governance structures (compare Table 4). In contrast, however, our data set is virtually silent about other variables including *individual researchers' characteristics* such as ability, stamina, personality, gender and experience, which are all known determinants of research output (Dundar and Lewis 1998, 614). The only variable we have here is the age of the research group leader. The same holds for factors characterizing group composition.

   Regarding *process variables*, we have some data on individual-level processes and group-internal processes. In this context, there is evidence that the 'working or research culture' of a group, as internalized by its members (Creswell 1986), and collaboration with other 'high quality' researchers are important factors with regard to research performance because of knowledge spillovers as well as synergies and complementarities concerning expertise and research interests (Carayol and Matt 2004; Allison and Long 1990). Here, our main variables relate to the research impulses that are generated by the leader from his/her own ideas in contrast to the research group. We believe this is a good approximate for various salient features of internal governance. For example, we can suspect that a group leader who gains few ideas from the group most likely does not interact with the group very much in research tasks and may be more focused on him/herself or other external cooperation options.

   In summary, with the exception of individual research group members' personal traits, important information is available for most other items. This will be explained in the following where we examine each of the variables used in the light of the results found in the literature.

Table 4 Summary statistics

Variable	Scale (source)	n	Mean	SD	Min	Max
Knowledge Output						
# Citations	Count (WoS)	316	116.68	175.55	0.00	1359.00
# Publications	Count (WoS)	321	30.70	39.17	0.00	320.00
Individual level (individual characteristics and processes)						
Impulse: own idea	1-5 Likert scale (DFG-survey)	317	4.44	0.57	1.00	5.00
Impulse: own research group	1-5 Likert scale (DFG-survey)	317	4.00	0.76	1.00	5.00
Year of PhD thesis	Metric (DFG-survey)	313	1984.01	9.05	1948.00	2003.00
Group level (inputs and their characteritiscs)						
Researchers (weighted by time share research)	Metric (DFG-survey)	321	4.14	3.91	0.50	22.40
Astro	Dummy (DFG-survey)	321	0.07	0.25	0.00	1.00
Nano	Dummy (DFG-survey)	321	0.33	0.47	0.00	1.00
Econ	Dummy (DFG-survey)	321	0.26	0.44	0.00	1.00
Satisfaction equipment	1-5 Likert scale (DFG-survey)	320	3.66	0.87	1.00	5.00
Organizational level (resources)						
University staff in FTE	Metric (EUMIDA)	321	6117.11	3639.75	174.00	13955.00
Student density	Metric (EUMIDA)	321	4.65	3.28	1.17	47.13
PhD density	Metric (EUMIDA)	321	0.09	0.04	0.03	0.46
# Subjects covered	Count (EUMIDA)	321	6.76	1.31	2.00	9.00
Distance education	Dummy (EUMIDA)	321	1.25	0.61	1.00	2.72
PhD internationalization	Metric (EUMIDA)	319	24.31	7.03	6.55	48.91
Year of foundation	Metric (EUMIDA)	321	1716.06	215.19	1386.00	2003.00
University publications staff	Metric (EUMIDA)	320	0.42	0.23	0.02	2.88
Hospital present	Dummy (EUMIDA)	321	2.07	0.84	1.00	2.72
Organizational level (processes)						
Influence of university presidents	1-5 Likert scale (DFG-survey)	296	3.84	0.91	1.00	5.00
Influence of deans	1-5 Likert scale (DFG-survey)	300	3.09	0.92	1.00	5.00
Bibliometric indicators used	Dummy (DFG-survey)	309	1.82	0.86	1.00	2.72
Regular evaluations	Dummy (DFG-survey)	313	1.65	0.83	1.00	2.72
Strict personnel quota	Dummy (DFG-survey)	312	2.38	0.68	1.00	2.72
Research council present	Dummy (DFG-survey)	304	2.20	0.79	1.00	2.72

At *research group level,* our data provide information on several relevant indicators which have been identified in the literature as influential determinants of research productivity.

A major variable—which lies at the center of our analysis—is the *number of scientists* making up a research group, *i.e. group size* measured by researchers in full-time equivalents.

Two additional group characteristic variables which refer to the available resources for research activities are the researcher's *time share* for *research* activities (in comparison to teaching and administrative tasks) and the researcher's self-stated *satisfaction with* their *IT-infrastructure and equipment*. The first variable serves as a proxy for time resources, since research and teaching can be considered as conflicting areas of activity, with more productive researchers usually spending fewer hours teaching (see, e.g. Fox 1983). Since we only deal with knowledge output and neglect other scientific products such as teaching, we can use this variable to adjust the time resource input according to what was effectively available for research. Therefore, the variable 'researchers (weighted by time share research)' is calculated as researchers in FTE multiplied by the average amount of time that is invested in research tasks. The second variable is used as a proxy for the quality of the infrastructure resources.

Another input variable is *year of PhD* of the respective leaders of research groups, i.e. chairs. Strictly speaking, this is an individual level characteristics. Nonetheless, since, in Germany, research groups at universities are dominantly shaped by the respective chairs/ professors, the year he/she obtained his/her PhD can be considered as an input indicator for the overall experience and social status of the group, which should also result in potentially higher research productivity because of the accumulated experience regarding research and publication activities (Libaers 2006; Dundar and Lewis 1998).

Our analysis also includes the *scientific discipline* as a control variable and proxy for the concrete task being fulfilled, since scientific fields differ significantly with regard to their publication cultures and channels and thus publication performance (see, e.g. Carayol and Matt 2004).

With respect to *group-level processes*, we recognize that university research groups in Germany are very focused on their group leader, i.e. the professor. This does not exclude the possibility that the group members work very independently, but it means that whether they do so often depends on a deliberate decision, the management or leadership style of the professor. In this respect, many of the internal processes are characterized by the way the research group leader sees him/herself in relation to the team members. For this purpose, we use two variables relating to the locus of where the research impulses originate. The first rates the *importance of the researcher's own research group as a source of knowledge and ideas*. This should be informative in the sense that a group leader giving this variable a high rating regards the group as a valuable source of ideas. At the same time there should be strong interdependencies between the leader and his/her co-workers and it is probable that team members are treated alike. The research group in this case is likely to be a multipolar network with intense communication and interdependency and a flatter hierarchy. The second variable measures the *degree to which the research group leader regards him/herself as the major source of ideas*. A high rating here ceteris paribus indicates a strong focus on the research group leader. In an extreme case, the research group in this case might resemble a highly centralized network with the other members circling the center as more or less important satellites. Here, communication and interdependency is assumed to be lower, and the whole hierarchy is steeper.

At the overarching *organizational level, i.e. the university*, we differentiate between *two major subsets of variables*, where the *first* represents *characteristics of the broader structural institutional context of the universities*, in which research groups are embedded: Here, again, we are most interested in factors influencing economies of scale and scope.

With regard to potential scale effects, our analysis includes the number of *university staff* (total staff enrolled per institution in full-time equivalents). With regard to potential economies of scope, the *number of subjects* taught at the respective university is included. Two additional variables which reflect the resources needed for conducting research and thus influencing productivity are *students per member of university staff (student density)*, as an indicator of the teaching load and *PhD students per member of university staff (PhD density)* as an indicator of research intensity.

Another variable which serves as an indicator for departmental prestige is *internationalization in PhD courses*. The assumption here is that excellent research universities are more successful in recruiting excellent (young) researchers from abroad, and that these researchers in turn contribute significantly to the university's research output. In addition, there is evidence for the US that foreign-born scientists are more productive than their native counterparts in terms of scientific productivity (see, e.g. Libaers 2006). With regard to universities' prestige, the analysis also includes the *year of founding* of the respective university as a control variable, since older universities on average tend to be larger than new ones, benefitting from potential economies of scale effects. Older universities on average also tend to have greater prestige than newer ones, thus attracting higher-quality researchers.

Our analysis also includes a variable to allow control for the structural particularities of universities specialized in *distance education*: This variable is introduced to identify those institutions whose main mission is to deliver distance education and where almost all programs are offered as distance learning courses to students who do not regularly attend the campus, except for workshops or short seminars.

Another control variable refers to the presence of a *university hospital*, since this often has a strong effect on the nature and orientation of the university. In particular, a hospital represents a service to society that is very cost-intensive. People employed in the hospital, though generally classified as researchers, usually only have a limited research mission due to their involvement in medical treatment. This is particularly important and should be controlled for in our data set, because our sample also consists of biotechnology research groups, which are often involved in clinical services. Finally, we include the variable *university publications per staff*, calculated as the overall number of publications of the university divided by overall staff to control for status effects that could provide an alternative explanation for why larger universities are more productive (Münch 2008).

The second subset of variables relates to the wider *university level governance* that the research groups are subject to. This 'group' is dominated by the variables that set out governance structure given by the New Public Management (NPM) paradigm. In order to understand NPM, it is helpful to contrast the pre-NPM system in Germany and New Public Management because, in a certain sense, the two are inverted mirror images. This can be seen most easily using the 'equalizer' model of de Boer et al. (2007). He argues that, in essence, there are 5 governance dimensions, whose specific importance can describe each university system. Therefore, each constellation of the 5 so-called 'sliding controllers' describes a different university system. Any of these constellations can be thought of as a specific setting on an equalizer, giving a particular tune. The first of the 'sliding controllers' is the level of state regulation and operative flexibility (measured here by the existence of a state-set rigid personnel quota). This mechanism determines the strength of government influence. The second controller concerns external control through government or societal stakeholders, often in the form of research councils, which was used here, and similar to the board of directors in a company (Mayntz 2002). The third is academic self-management, which measures the degree to which a chair holder can make

autonomous decisions. Its antitype is an internal hierarchical control exerted by the deans and chancellors within the university. For these two dimensions we use the influence of deans and university presidents. The fifth is the degree of competition. Major university-internal mechanisms are operated via the use of evaluations and bibliometric indicators to redistribute basic funds. We use these two variables. We acknowledge that this discussion of NPM is not extensive, but it should give an impression of some of the subtleties which ought to be discussed in this context.

Results

In "Methodology", we proposed the use of the Cobb-Douglas production function to estimate the returns to scale. This parametric specification allows for increasing, constant, and decreasing returns to scale, which makes it relatively flexible. At the same time, it does not allow for varying returns to scale, for example, when we observe IRS for low levels of input followed by DRS above certain thresholds.

Appropriateness of the Cobb-Douglas specification

It is important to examine whether our parametric Cobb-Douglas specification is indeed flexible enough to correctly model the returns to scale in our data set. Therefore, we used a semi-parametric partially linear regression model in a pre-test that delivers a visual impression of the likely behavior of the returns to scale. The characteristic feature of the partially linear regression model is that the linear specification is imposed on some variables as in a regular regression, while others are allowed to have more flexible impacts on the response variable. For the non-parametrically treated variables, we can then extract a non-parametric estimate of their impact on the response variable by running a local linear regression (so-called lowess smoother) of the predicted response variable on the non-parametrically treated covariates.

In order to get an impression of the returns to scale on the research group and university levels, we allow both the number of researchers in the research group as well as the number of staff in the university to have an arbitrary impact on the number of publications and citations, while we impose the linear restrictions on all other variables. The partial linear regression then allows a lowess smoother to be constructed which plots the relationship between outputs and research group or university size on a graph.

The results can be found in Fig. 2, where the relationships between publications, citations and research group size are shown in the two upper panels and the corresponding relationships with respect to university size in the lower panels.[6]

It is obvious that all the estimated mean functions exhibit a positive slope. Therefore, both university size and a greater number of researchers in the research group increase knowledge output, irrespective of whether we consider publications or citations. Furthermore, the function seems to be concave (upper left, upper right graphs) or concave to linear (lower left, lower right). Thus we can ascertain that the proposed Cobb-Douglas specification is flexible enough, because there is no indication of returns to scale that vary with the level of output.

We will not enter a detailed discussion about returns to scale based on Fig. 2, because the results may also be due to disciplinary differences that are hard to control for in the

---

[6] It should be noted that the predicted values are plotted on the y-axes which is why they can be negative.
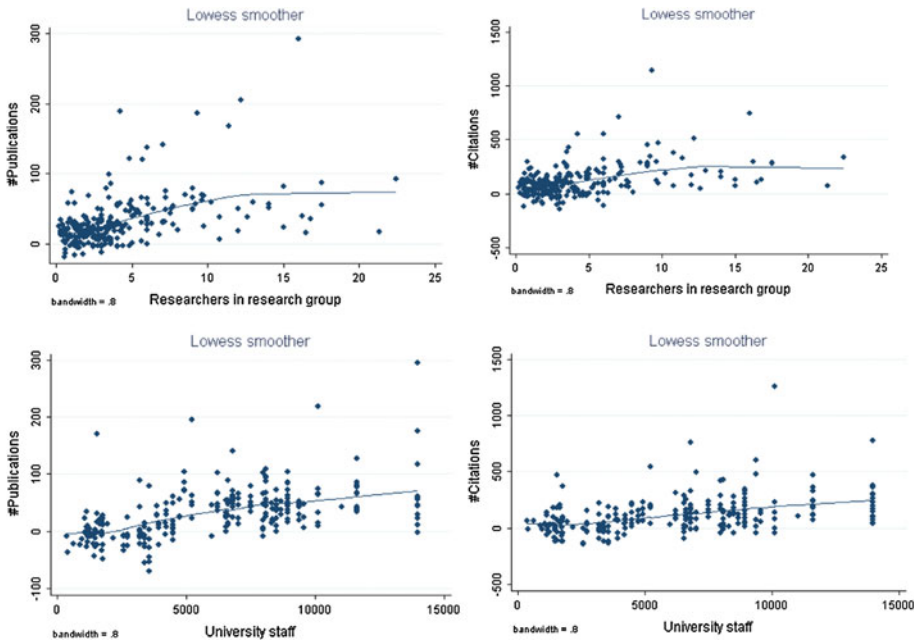
Fig. 2 Non-parametric relationship of predicted knowledge output and size (lowess after partial linear regression)

employed semi-parametric approach. Thus we turn now to the parametric specification based on Eq. (2), where we allow the coefficients of size variables 'Researchers (weighted by time share research)' and 'University staff in FTE' to vary by discipline.

Returns to scale in science: main results

Table 5 presents the main output regressions based on the Cobb-Douglas function, where all variables have been taken in logs. The regression on the left shows the effects for the number of publications, the right-hand one for the number of citations. There are remarkable similarities between the two.

In particular, with respect to the main size variables, we observe that for the fields of biotechnology, nanotechnology, and economics in the publication regression both the direct input variable 'Researchers (weighted by time share research)' as well as the organizational size variable 'University staff in FTE' are positively significant. The same holds true for Nanotechnology and Biotechnology with respect to citations. The insignificance with respect to astrophysics is likely to be due to the fact that there are only 25 research groups in total.

As explained in "Defining returns to scale and agglomeration effects", positive size effects on the university level are indicated by positively significant coefficients of university staff. Therefore, we can already conclude that the size of the university positively affects the performance of the research groups. This is clearly an indication of IRS at the level of the university.

Because the number of researchers in the group is a direct input variable, the interpretation concerning returns to scale differs here. As explained previously, decreasing

Table 5 Knowledge output regression (all variables in logs)

	# Publications		# Citations	
	Coefficient	SE	Coefficient	SE
Individual level (individual characteristics and process)				
Impulse: own idea	0.1980	0.4055	0.3339	0.4208
Impulse: own research group	−0.1176	0.2566	−0.0576	0.3479
Group level (direct inputs and their characteristics)				
Astro*Researchers (weighted by time share research)	0.3000	0.3094	0.2541	0.3623
Nano*Researchers (weighted by time share research)	0.5501***	0.1180	0.6929***	0.1481
Econ*Researchers (weighted by time share research)	0.3340*	0.1898	0.2231	0.3394
Bio*Researchers (weighted by time share research)	0.4494***	0.1025	0.6428***	0.1526
Astro	5.5924**	2.6173	5.3936	4.2469
Nano	0.9440	2.2693	−3.0648	3.7145
Econ	−2.0491	1.8321	−7.8130**	3.6305
Satisfaction equipment	−0.0256	0.1675	−0.3234	0.2831
Year of PhD thesis	−5.8741	15.3291	−16.7966	18.9796
Organisational level (resources)				
Astro*University staff	−0.0776	0.2594	−0.0498	0.4215
Nano*University staff	0.3870*	0.2211	0.7648**	0.3612
Econ*University staff	0.5848***	0.2100	1.0096***	0.3629
Bio*University staff	0.5476***	0.1560	0.4916*	0.2978
Student density	0.0935	0.2752	−0.3560	0.3879
PhD density	0.0920	0.2729	−0.0732	0.4455
# Subjects covered	−0.3156	0.3876	−0.0287	0.5829
Distance education	0.2307	0.2409	0.0732	0.3654
PhD internalization	0.3986	0.3546	0.7139	0.4455
Year of foundation	−0.7186	0.5080	−0.1631	0.7248
University publications per staff	0.3674**	0.1611	0.7424***	0.2567
Hospital present	−0.2432	0.2646	−0.3935	0.3882
Organisational level (processes)				
Influence of university leaders	0.1311	0.1997	0.2992	0.2783
Influence of faculty leaders	−0.0005	0.1474	−0.1794	0.2631
Bibliometric indicators used	0.0625	0.1362	0.1169	0.2637
Regular evaluations	0.2981**	0.1183	0.4240**	0.2093
Strict personnel quota	−0.1145	0.1436	−0.0586	0.2255
Research council present	−0.2706**	0.1356	−0.5366**	0.1981
Constant	47.8616	117.2771	127.3377	146.1086
Sigma	0.8857***	0.0661	1.2256***	0.0900
n	258		258	
# Clusters	58		58	
Pseudo-R^2	0.2834		0.3043	

***1 % significance, ** 5 % significance, *10 % significance

returns to scale are indicated by coefficients smaller than 1, constant returns are associated with a value of exactly 1, while increasing returns imply a coefficient of greater than 1. Since all the estimated coefficients are below one, we believe there are decreasing returns to scale.

Thus, the results obtained so far suggest DRS on the research group level, while we observe IRS at the level of the university. We argued that this constellation favors an organizational model which is characterized by relatively small research groups concentrated in large universities. This is exactly the organizational model that we can observe all over the world. Thus, from an efficiency perspective, this dominant design for organizing public research does indeed seem justified.

To corroborate this conclusion, we now present additional tests. First, since the significance levels in Table 5 only indicate that the coefficients are different from zero, we also provide tests that indicate whether they are indeed significantly below one. In other words, we test whether the effect of additional researchers in the group really tapers off resulting in DRS. Second, we test whether the effects differ by discipline. Both tests are conducted for 'Researchers (weighted by time share research)' and 'University staff in FTE'. The results are summarized in Table 6.

We note that both with respect to citations and publications the coefficient associated with the number of researchers in the group is significantly below one for all disciplines. This is also corroborated by a test that all coefficients are jointly below one. Thus, in fact, we are able to corroborate that DRS can be observed at the level of the research group. Additionally, we can observe that the test for equality of all coefficients is not significant, which implies that it was not necessary to control for the discipline.

A similar picture emerges for the effects of university size. The coefficients are usually significantly below unity, implying that the effect of university size is positive but tapers off. This implies that the additional advantages of ever-increasing universities will gradually diminish. The only exceptions to this can be observed for economics and nanotechnology with respect to the number of citations. For both disciplines the coefficients are not significantly different from one. As a consequence, the advantages of increasing universities remain constant irrespective of the size of the university. In any case, the fact that the coefficients for university size are significantly below one in the majority of cases does not impede the existence of IRS, because we only required this coefficient to be positive for increasing returns to scale. That this is the case can be seen in Table 6.

Table 6 Constant returns to scale test for 'Researchers (weighted by time share research)

	# Publication	# Citation
Astro	5.12***	4.24**
Econ	12.31***	5.24**
Bio	28.84***	5.47**
Nano	14.55***	4.30**
Joint test (constant returns)	78.82***	20.5***
Joint test (all equal)	1.51	2.66

Tables 6 and 7 should be read as follows: a significant test result means that the coefficient in Table 5 is significantly smaller than 1.

***1 % significance, ** 5 % significance

Table 7 Constant returns to scale test for "University staff in FTE"

	# Publications	# Citations
Astro	17.26***	6.2**
Econ	3.91**	0
Bio	8.41***	2.91*
Nano	7.69***	0.42
Joint test (constant returns)	25.49***	9.08*
Joint test (all equal)	6.07	5.78

***1 % significance, ** 5 % significance, *10 % significance

Further results

Some secondary effects are worth mentioning. First of all, we observe that the research intensity of the university (measured by university publications per staff member) is positively significant with respect to the number of both publications and citations. This means that the research intensity of the university has a positive effect on the knowledge output of the research groups. There may be several reasons for this. One is that research intensity captures the effect of knowledge spillovers, as we could argue that their value is increased by the quality and amount of research conducted in the university as a whole. On the other hand, it is also possible that already research-intensive universities find it easier to attract able and productive researchers, which facilitates the formation of high-performing research groups. Also status-effects as suggested by Münch (2008) may be important. Thus, this positive association could also be due to selection effects. Additionally this observation is related to a whole series of papers having been written by van Raan (2007a, b) who deals with the question of scaling rules in science, that determine inasmuch university or research group size is related to performance in terms of citations. One of the major findings is that in particular at the lower end of the productivity distribution increasing size causes increases in citation performance. While there are differences in respect to our methodology and research design, that relate to the measurement of size using publications (we use scientific staff) as well as explanation patterns (we focus on characteristics of the production function while van Raan (2007a, b) highlights network patterns in conjunction with the topology of the "network space"), these results are in particular with respect to the level of the university quite similar. They may reflect two different perspectives on a common and well-known phenomenon: the Matthew effect that postulates the existence of cumulative advantages of size and reputation.

Furthermore, we observe that regular evaluations have a positive effect on the research groups' output, while the existence of research councils is negative. This is comparable to the results in Schubert (2009) and Schmoch and Schubert (2010), who point out that, while evaluations are science-driven strategic governance mechanisms which specifically aim at improving publication performance, research councils are more strongly influenced by state and society representatives, who are usually more interested in pushing the applicability of knowledge, knowledge transfer, and teaching.

Conclusions

This paper examined the organizational model of universities. We argued that there is a dominant organizational design, where a large number of relatively small individual research groups are concentrated within a single organizational frame. This frame is what

we call a university. In a certain sense, universities are organized as holdings, where the steering level holds a large number of individual research groups that are relatively free to go about their production-related tasks. While this organizational shape is certainly also the result of evolutionary and historical processes, we asked whether it is an optimal one from an efficiency perspective. We argued that decreasing returns to scale associated with production at the research group level (for example due to the increasing costs of coordination and communication) and agglomeration effects at the university level (for example as a result of shared infrastructure) should give rise to an optimal organizational form akin to a holding.

Based on an extraordinarily rich data set for German research groups, we then tested the hypothesis of positive agglomeration effects and DRS at the production-related research group level. This constellation of multilevel size effects could be corroborated, which suggests this form of organization is also optimal from an efficiency perspective. Thus this paper contributes to the understanding of the patterns of university organization from an economic point of view. Not wanting to belittle the power of historical processes that have helped to shape this form, we still think that the efficiency argument is also a strong force which has played a significant role in shaping the long-term development of this organizational design.

As a direct corollary of our results, it is also possible to draw some conclusions with respect to the increasing concentration of financial resources in science. In particular, whether the "big science" hypothesis is reasonable or not depends on what we mean by it. If this means that we should concentrate resources on only a few research groups in order to induce growth on the production level, then the hypothesis is probably false. However, if we mean that we should concentrate the resources on a few universities, while leaving the size of the research groups unaffected, then there is probably some truth to this.

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.

Allison, P., & Long, S. (1990). Departmental effects on scientific productivity. *American Sociological Review, 55*, 469–478.

Anderson, R., Quigley, J. M., & Wilhelmson, M. (2004). University decentralization as regional policy: the Swedish experiment. *Journal of Economic Geography, 4*, 371–388.

Andrews, F. M. (Ed.). (1979). *Scientific Productivity: The Effectiveness of Research Groups in Six Countries*. Cambridge: Cambridge University Press; UNESCO Paris.

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.

Church, J., & Ware, R. (1999). *Industrial Organization: A strategic Approach*. USA: Irwin Mc-Graw-Hill.

Cohen, J. E. (1991). Size, age and productivity of scientific and technical research groups. *Scientometrics, 20*, 395–416.

Cohn, E., Rhine, S., & Santos, M. (1989). Institutions of higher education as multi-product firms: economies of scale and scope. *Review of Economics and Statistics, 71*, 284–290.

Creswell, J. W. (1986). *Measuring Faculty Research Performance. New Directions for Institutional Research, 50*. San Francisco: Jossey-Bass.

de Boer, H., Enders, J., Schimank U. (2007). On the way towards new public management? The governance of University Systems in England, the Netherlands, Austria, and Germany. In D. Jansen (Ed.), *New forms of governance in research organizations. Disciplinary approaches, Interfaces and Integration* (pp. 137–152). Dordrecht: Springer

De Groot, H., Mc Mahon, 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. (1998). Determinants of research productivity in higher education. *Research in Higher Education, 39*, 607–631.

European Commission, Research Directorate-General Directorate C (Ed.), (2010). Feasibility Study for Creating a European University Data Collection. Final Study Report. Download: ec.europa.eu/research/era/docs/en/eumida-final-report.pdf [August 2011].

Fox, M. F. (1983). Publication productivity among scientists: A critical review. *Social Studies of Science, 13*, 285–305.

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.

Granovetter, M. S. (1973). The Strength of Weak Ties. *The American Journal of Sociology, 78*, 1360–1380.

Griliches, Z. (1979). Issues in assessing the contribution of research and development to productivity growth. *Bell Journal of Economics 10*(1), 92–116.

Hannan, T. H., & Mavinga, F. (1980). Expense preference and managerial control—The case of banking firms. *Bell Journal of Economics, 11*, 671–682.

Hemlin, S., & Gustaffson, M. (1996). Research production in the arts and humanities- a questionnaire study of factors influencing research performance. *Scientometrics, 37*(3), 417–432.

Hodgson, G. M. (1993). *Economics and Evolution*. Cambridge: Polity Press.

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. (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.

Johnston, R. (1994). Effects of resource concentration on research performance. *Higher Education, 28*(1), 25–37.

Koshal, R. K., & Koshal, M. (1995). Quality and Economics of Scale in Higher Education. *Applied Economics, 27*, 773–778.

Krugman, P. (1991). Increasing returns and economic geography. *Journal of Political Economy, 99*, 483–499.

Kyvik, S. (1995). Are big university departments better than small ones? *Higher Education, 30*, 295–304.

Laband, D. N., & Lentz, B. F. (2003). New estimates of economies of scale and scope in higher education. *Southern Economic Journal, 70*, 172–183.

Laredo, P., & Mustar, P. (Eds.). (2001). *Research and innovation policies in the new global economy*. Edward Elgar: An International Comparative Analysis

Lewis, D., & Webb, J. R. (2007). Potential cost synergies from banks acquiring real estate brokerage services. *Journal of Banking & Finance, 31*, 2347–2363.

Libaers, D. (2006) Organizing for scientific performance: The impact of organizational affiliation on scientific productivity in Nano Science and Technology Portland International Conference on Management of Engineering and Technology, 1, art. no. 4077377, pp. 115–126.

Lo, S. F., & Lu, W. M. (2006). Does size matter? *Finding the profitability and marketability benchmark of financial holding companies, Asia-Pacific Journal of Operational Research, 23*, 229–246.

Mairesse, J., & Mohnen, P. (2002). Accounting for innovation and measuring innovativeness: an illustrative framework and application. *American Economic Review, 92*(2), 226–230.

March, J. G. (1991). Exploration and Exploitation in Organizational Learning. *Organization Science, 2*, 71–87.

Marshall, A. (1890). *Principles in Economics*. London: Macmillan Press Ltd.

Marx, K. (1867). Das Kapital: Buch 1 Der Produktionsprozess des Kapitals. download: http://mises.ch/library/Marx_Das_Kapital.pdf.

Mulder, Peter, De Groot, Henri L. F., & Hofkes, Marjan W. (2001). Economic growth and technological change: A comparison of insights from a neo-classical and an evolutionary perspective. *Technological Forecasting and Social Change, 68*, 151–171.

Münch, R. (2008). Stratifikation durch Evaluation. Mechanismen der Konstruktion von Statushierarchien in der Forschung. *Zeitschrift für Soziologie, 37*, 60–80.

Narin, F., & Hamilton, K. S. (1996). Bibliometric performance measures. *Scientometrics, 36*, 293–310.

Nelson, R. R. S., & Winter (1982). *An Evolutionary Theory of Economic Change*. Cambridge: Harvard University Press.

Rey-Rocha, J. A. B., Garzón-García, B. A., & Martín-Sempere, M. J. A. (2006). Scientists' performance and consolidation of research teams in biology and biomedicine at the Spanish Council for Scientific Research. *Scientometrics, 69*(2), 183–212.

Salas, E.A., Stagl, K.C.B, Burke, C.S.B, Goodwin, G.F.C. (2007): Fostering team effectiveness in organizations: Toward an integrative theoretical framework (Nebraska Symposium on Motivation, 52, pp. 185–243.

Sav, G. T. (2004). Higher education costs and scale and scope economies. *Applied Economics, 36*, 607–614.

Schmoch, U., & Schubert, T. (2008). Are international co-publications an indicator for quality of scientific research? *Scientometrics, 74*(3), S361–S377

Schmoch, U., & Schubert, T. (2009). Sustainability of incentives for excellent research—The German case. *Scientometrics, 81*, 195–218.

Schmoch, U., & Schubert, T. (2010). Strategic steering of research by new public management in German universities: a looming state-science conflict? *Research Evaluation, 19*, S209–S216

Schmoch, U., Schubert, T., Jansen, D., Heidler, R., van Görtz, R. (2010). How to use Indicators to measure scientific performance? A Balanced Approach. *Research Evaluation*, 19, S2–S18.

Schubert, T. (2009). Empirical observations on new public management to increase efficiency in public research—Boon or Bane? *Research Policy, 38*, 1225–1234.

Seglen, P. O., & Aksnes, D. W. (2000). Scientific productivity and group size. A bibliometric analysis of Norwegian microbiological research. *Scientometrics, 49*, 125–143.

van Raan, A. F. J. (2007a). Scaling rules in the Science System: Influence of field-specific citation characteristics on the impact of Research Groups. *Journal of the American Society for Information Science and Technology, 59*, 565–576.

van Raan, A. F. J. (2007b). Bibliometric Statistical Properties of the 100 Largest European Research Universities: Prevalent Scaling Rules in the Science System. *Journal of the American Society for Information Science and Technology, 59*, 461–475.

Vin 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.

West, M. A. (2004). *Effective teamwork: Practical lessons from organizational research* (2nd edn). Oxford: Blackwell Publishing.

West MA, Borrill CS, Unsworth KL (1998) Team effectiveness in organizations. In Cooper CL,Robetson IT (eds.) International review of industrial and organizational psychology pp. 1–48. 13.

Worthington, A. C., & Higgs, H. (2011). Economies of scale and scope in Australian higher education. *Higher Education, 61*, 387–414.