Journal of Informetrics 11 (2017) 613–628

Contents lists available at ScienceDirect

Journal of Informetrics

journal homepage: www.elsevier.com/locate/joi

Regular article

Are there any frontiers of research performance? Efficiency measurement of funded research projects with the Bayesian stochastic frontier analysis for count data

Rüdiger Mutz[a,*], Lutz Bornmann[b], Hans-Dieter Daniel[a,c]

[a] Professorship for Social Psychology and Research on Higher Education, ETH Zurich, Zurich, Switzerland
[b] Division for Science and Innovation Studies, Administrative Headquarters of the Max Planck Society, Munich, Germany
[c] Evaluation Office, University of Zurich, Zurich, Switzerland

ARTICLE INFO

Article history:
Received 20 December 2016
Received in revised form 26 April 2017
Accepted 26 April 2017
Available online 14 June 2017

Keywords:
Stochastic frontier analysis
Productivity and efficiency analysis
Data envelopment analysis
Research funding

ABSTRACT

In recent years, scientometrics has devoted increasing attention to the question of measurement of productivity and efficiency in research. In econometrics, the question is usually examined using data envelopment analysis. Alternatively, in this paper we propose using a statistical approach, Bayesian stochastic frontier analysis (B-SFA), that explicitly considers the stochastic nature of (count) data. The Austrian Science Fund (FWF) made data available to us from their peer review process (ex-ante peer evaluation of proposals, final research product reports) and bibliometric data. The data analysis was done for a subsample of N = 1,046 FWF-funded projects (in Life Science and Medicine, Formal and Physical Sciences). For two outcome variables, a general latent research product dimension (CFACTOR) and the total number of publications (P), technical efficiency values (TE) were estimated for each project using the SFA production functions. The TE values for CFACTOR and P were on average 0.86 and 0.27, as compared with a maximum TE value of 1.0. With regard to CFACTOR, female PIs, younger PIs, and projects with longer durations have slightly higher TE than male PIs, older PIs, and projects with shorter durations. A simulation study showed the statistical behavior of the procedure under different sampling conditions.

1. Introduction

In recent years, scientometrics has devoted increasing attention to the question of measurement of productivity and efficiency in research, as shown for example by a special section of the *Journal of Informetrics* (Volume 10, Issue 2) published in 2016. The use of the concepts of productivity and efficiency favors economic approaches, which relate research output to input (e.g., the amount of funding). For instance, Abramo and D'Angelo (2014), who are proponents of this perspective, try to define and measure research productivity within a microeconomic theory of production framework and utilize a numerical nonparametric approach, data envelopment analysis (DEA). In econometrics, DEA is considered to be a standard method of efficiency analysis (Liu, Lu, Lu, & Lin, 2013a; Liu, Lu, Lu, & Lin, 2013b), and it also has numerous applications in research on higher education (e.g., Bessent, Bessent, Charnes, Cooper, & Thorogood, 1983; Sinuanystern, Mehrez, & Barboy, 1994). In the 2016 special section of *Journal of Informetrics*, several contributions discuss Abramo and D'Angelo's (2016, p. 646)

* Corresponding author at: Professorship for Social Psychology and Research on Higher Education, ETH Zurich, Muehlegasse 21, 8001, Zurich, Switzerland.
E-mail address: mutz@gess.ethz.ch (R. Mutz).

http://dx.doi.org/10.1016/j.joi.2017.04.009

proposition "to switch" from size-independent indicators based on the ratio to publications, such as crown indicators (e.g., mean normalized citation score), "to rankings by research efficiency." The issues of productivity and efficiency are discussed mainly in the context of performance-based university research funding. That approach demands, in addition to the ex-ante evaluation of research projects, also their ex-post evaluation. This raises explicitly the questions of the effectiveness and efficiency of research funding (Hicks, 2012; Rabovsky, 2014a, 2014b).

The use of nonparametric methods of productivity and efficiency analysis like DEA requires deterministic indicators (Glänzel, 2010, p. 314). Any kind of random noise or stochastic component is not considered. Alternatively, in this paper we propose a statistical approach for analysis of productivity and efficiency, stochastic frontier analysis (SFA), and apply it to the input and output data of projects funded by the Austrian Science Fund (FWF) as an example.

The next section below looks at the foundations of SFA as opposed to ordinary regression analysis and DEA. The following sections then outline the research questions and formulate the hypotheses, describe the data, and provide methodological-statistical details. The paper concludes with a presentation of the results and a discussion of the findings.

2. Regression, DEA, and stochastic frontier

Economic productivity and efficiency analyses are based largely on four elements: Decision making units (DMUs), outputs, inputs und a function that describes the transformation of input into output formally (mathematically), the production function. For research funding organisations are the funded projects the DMUs, which owing to the financial input, especially the grant sum, and the intellectual capital of a proposal (Falzagic, 2005) produce research output (e.g., journal article, book, conference contribution, PhD). The intellectual capital of a project is rated by the reviewers in the ex-ante evaluation of the proposals.

From the perspective of data analysis, in a productivity and efficiency analysis, as compared with the classical approach of statistical regression, it is not primarily about prediction. The aim is not to predict the research output of a DMU on average but instead to determine the maximum expected research outputs of the DMU with the given input (Archibald & Feldman, 2008), for example the maximum possible number of publications given a certain grant sum. The basis is the transformation of a set of research inputs or production factors into a set of research outputs. The set of maximum expected research outputs for a set of given research inputs defines the production function, or the *frontier* of production. Technical efficiency (TE) characterizes "the relationship between observed production and some ideal, or potential production" (Greene, 2008, p. 100), which is the maximum feasible output for an input as it is represented by the estimated production function. Productivity is defined "as the ratio of output y (what we produce) over input x (the resource we use: Productivity = y/x)" (Faire, Grosskopf, & Margaritis, 2008, p. 522).

In econometrics, DEA is the standard method of analysis of TE; it is also already being used frequently in research on higher education and in scientometrics (e.g., Abbott & Doucouliagos, 2003; Abramo, Cicero, & D'Angelo, 2011; Abramo & D'Angelo, 2014; Athanassopoulos & Shale, 1997; Charnes, Cooper, & Rhodes, 1978; Johnes, 2004, p. 625; Johnes, 2006; Liu et al., 2013a): "DEA involves the use of linear programming methods to construct a non-parametric piece-wise surface (or frontier) over the data. Efficiency measures are then calculated relative to this surface" (Coelli, Rao, O'Donnell, & Battese, 2005, p. 162). Here, the level of analysis is mostly the entire institution or certain scientific disciplines of an institution, which, however, makes direct assignment of funding resources to research output difficult. In the end it is the individual research project in which the assignment of input to output takes place. In this connection, Abramo and D'Angelo (2014) point to the necessity for a two-step procedure "first measuring the productivity of the individual researchers in their field, and then appropriately aggregating the data" (p. 1132). Whereas in Abramo and D'Angelo (2016) the DMUs are researchers, in our case they are the research projects.

As compared with other methods of production efficiency, especially the statistical approach of stochastic frontier analysis (SFA) (Coelli et al., 2005; Greene, 2008; Johnes, 2004; Kumbhakar, Wang, & Horncastle, 2015; Parmeter, 2014), DEA, as a numerical nonparametric technique, is preferred for the following reasons (Abramo et al., 2011, pp. 231&232; Abramo & D'Angelo, 2014, p. 1133; Bonaccorsi & Daraio, 2004; Chen, Delmas, & Lieberman, 2015): First, with DEA, complex production functions with multiple outputs and inputs can be analyzed and described with a single efficiency indicator. Second, with DEA, no specific functional connection in the form of a production function has to be defined (e.g., the Cobb-Douglas production function). Third, benchmark best practices can be identified directly; "in other words, comparisons are to real production units that are used as references for best practices" (Abramo & D'Angelo, 2014, p. 1133). Last but not least, the numerical technique produces robust solutions for the desired production efficiency.

SFA has the following advantages over these undoubtedly strong arguments for DEA: First, SFA takes account of random measurement errors. The research output of DMUs, such as number of publications or citations, are not fixed values but are affected by many random factors (e.g., periodic updates of bibliographic data bases, random fluctuations of universities' performances). The research output and impact of DMUs fluctuates to a certain extent by chance. Second, for the measurement of productivity and efficiency it is crucial what factors actually determine the measured values. In DEA this is a two-stage procedure: The productivity efficiency values calculated in the first stage of DEA are regressed statistically on predictors by ordinary least-squares regression (OLS) in a second step (e.g., Johnson & Kuosmanen, 2012; Ramalho, Ramalho, & Henriques, 2010). This procedure has limitations, however, as ". . . second-stage OLS estimation is consistent only under very peculiar and unusual assumptions on the data-generating process that limit its applicability" (Simar & Wilson, 2011, p. 205). In SFA, the efficiency values and the regression parameters are estimated unbiased in one step (Wang & Schmidt, 2002). Third, by combining SFA with multi-level models it is possible to model hierarchically structured performance data (e.g., projects nested within faculties, faculties nested within universities, and so on). Fourth, since with SFA it is possible to test what production function fits the data best, the problem of explicit definition of a functional connection is mitigated (Ehlers, 2011). Also, considerably more complex models can be estimated with SFA than with DEA, such as panel data and multilevel and latent class models (Parmeter, 2014).

3. Research questions and hypotheses

Besides the introduction to the statistical approach in scientometrics, this empirical contribution focuses on the general issue as to whether there actually is a frontier of research performance with a given input of resources. This general question will be answered taking the example of research grant data (data on ex-ante and ex-post evaluations and bibliometric data) from the FWF, Austria's central funding agency for basic research. As the statistical method we favor SFA over DEA for the reasons mentioned above.

Specifically, this study will answer the following questions:

1. Can a production function, i.e., frontier of research performance, be identified for the FWF projects?
2. How high is the TE of FWF-funded research projects?
3. What are possible exogenous determinants of TE?
4. How does the statistical procedure behave under different sampling conditions, including sample size, scale level (continuous data, count data), and proportion of random noise by conducting a simulation study?

The following hypotheses on determinants of TE, which are discussed in the literature, guide the analysis:

1. *Age of PI*: The hypothesis that with increasing age of the principal investigators (PIs) their productivity and thus also TE decrease is a controversial issue (e.g., Abramo, D'Angelo, & Murgia, 2016; Gingras, Lariviere, Macaluso, & Robitaille, 2008; Over, 1988; Stroebe, 2010). We assume that older PIs are more strongly engaged in other tasks (e.g., teaching, management) in addition to research than younger PIs are.
2. *Gender*: In line with our empirical investigations of the peer review process at FWF on the topic of gender bias (Mutz, Bornmann, & Daniel, 2012), we do not expect to find any gender differences also in this study, although there is some empirical evidence for gender imbalance in productivity in favor of men (e.g., Lariviere, Ni, Gingras, Cronin, & Sugimoto, 2013).
3. *Approved grant sum*: It is common practice to choose factors that feed into the production function (such as grant sum, ex-ante peer evaluation) also as determinants of TE (Kumbhakar et al., 2015, p. 80f; Wang & Schmidt, 2002, p. 130). According to our empirical results (Mutz, Bornmann, & Daniel, 2016) we assume that the requested grant sum reflects the project costs. The approved grant sum, however, only comprises a more or less reduced proportion of the requested grant sum. This proportion depends on the merit of the proposal, rated in the ex-ante peer evaluation. It can thus be assumed that with a higher approved grant sum (financing capital), TE will also be higher.
4. *EXANTE*: For the ex-ante peer evaluation (EXANTE) as a further factor in the production function it is expected that a higher rating of research proposal in the EXANTE, or higher "intellectual capital," is also associated with higher TE.
5. *Scientific discipline*: It can be assumed that both the production function and TE vary across scientific disciplines (e.g., Abramo et al., 2011).
6. *Size of institution*: In connection with bibliometric analysis of the performance of universities, also the question of the effect of organization factors is frequently examined (e.g., Dundar & Lewis, 1998; Horta & Lacy, 2011). According to Dundar and Lewis (1998, p. 611), larger institutions can better facilitate collaboration among researchers than smaller ones and have greater access to funding resources and more degrees of freedom in their use. For this reason, it is assumed that large universities like the University of Vienna (which is significantly larger than other universities in Austria), promote TE more than smaller universities do (such as Alpen-Adria-Universität Klagenfurt).
7. *Project duration*: A further important factor that can affect TE is project duration. A survey of PIs commissioned by the National Science Foundation in the United States (Ballou, Mishkind, Mooney, & van Kammen, 2002) found key negative impacts on a research project that are connected with project duration cuts, as reported by the PIs (the percentage of PIs reporting the problem is shown in parentheses) (Ballou et al., 2002, p. 10): "ability to achieve their research objectives within the specified time" (67%), "ability to obtain high quality personnel" (55%), "ability to pursue high-risk ideas" (51%), "collaborate with researchers in the area of research" (50%). Upon this background, it is assumed that TE increases with increasing project duration.

4. Data and methods

4.1. Data

The statistical analysis is based on a reanalysis of research grant data from the peer review procedure at the FWF (for the funding program "Stand-Alone Projects") from 1999 to 2009. The FWF provided us with data on the review process (ex-ante evaluation of the submitted proposals and ex-post evaluation of the funded project) and additional bibliometric data on the scientific output (articles, letters, and reviews) of funded research projects (Wijk van & Costas-Comesana, 2012). The bibliometric data were the result of Van Wijk and Costas-Comesana's analysis conducted at the Centre of Science and Technology Studies (CWTS), Leiden, based on 13,733 unique FWF-funded publications, as they were reported to the FWF by the project leadership, from the period from 2001 to 2010 and on the base of the bibliographic multidisciplinary database Web of Science (Clarivate Analytics, formerly the IP & Science business of Thomson Reuters). We received the data in anonymized form, which means that the names of the PIs were not identifiable.

As data on the research output of the funded projects are included in the SFA (especially bibliometric data on the journal articles), projects from all scientific disciplines cannot be included in the analysis. Projects can be included only in those disciplines where journal articles make up a large share of the research output. For this study, therefore, we could choose only the scientific disciplines named in the ex-ante peer evaluations where journal articles play a large role in research output (such as Life Science). However, since the research reports of the ex-post evaluation of the funded projects contained data on the complete research output (e.g., books, journal articles, etc.), we chose the disciplines based on that data, i.e. evidence-based. Here we utilized a typology of disciplines that we had obtained statistically via a latent class analysis of research output (cluster analysis) (Mutz, Bornmann, & Daniel, 2013).

For this study we selected latent clusters (LCLUSTER) with a high proportion of journal articles as research output. They are 12 scientific disciplines from the following three LCLUSTER: Life Sciences and Medicine (biology, botany, zoology, geoscience, preclinical medicine, clinical medicine, agricultural science/forestry/veterinary sciences), Formal Sciences (mathematics, computer science, economic sciences), and Physical Sciences (physics/astronomy, mechanics).

Of the total N = 8,496 research proposals, which were ex-ante rated (on a scale from 1 to 100) by external referees (on average by 2–3 referees for each submitted proposal), and a total N = 2,204 approved proposals and funded projects, N = 1,046 approved proposals and funded projects with complete data were available (data on ex-ante evaluation, ex-post evaluation, and bibliometric data) for the selected disciplines.

Of the FWF grant applicants (complete data, N = 1,046), 13% were women, 87% were men. Their median age was 44 ($\bar{x}$ = 45.4, SD = 9.3, MIN = 26, MAX = 79).

4.2. Variables

Included in the statistical analysis were three groups of variables:

(a) *Peer review data*: Each research proposal was rated by 2–3 referees on a scale from 1 to 100 (ex-ante peer evaluation, EXANTE).

(b) *Project data*: For each of the FWF-funded projects, a report on the research outputs and data on the approved grant sum (GRANT) was available. According to Kumbhakar et al. (2015, p. 1) "more generally, SF analysis can be applied to any problem where the observed outcome is either less or more than the potential outcome". The research products were journal articles, books, proceedings, conference contributions, and scientific qualifications (diploma, doctoral thesis, and habilitation). This information was coded quantitatively and typologized statistically (types of products and types of scientific disciplines) using latent class analysis (LCA) (see Mutz et al., 2013). Comparable to a factor analysis in an LCA, actually, categorical latent dimensions (types, clusters) are extracted that overall summarize the correlations between the set of research products. In order to explain all correlations, also latent continuous factors can be extracted that represent a quantitative dimension of total production. In our analysis (Mutz et al., 2013), we extracted independently from the latent classes one latent continuous dimension. This dimension was standardized with mean value of 0. The higher the value of a research project on this dimension is, the higher the research project scores on overall research output in terms of number of articles, number of books, and number of proceedings. This variable was used as an indicator of the total production of a project and as an output variable (CFACTOR) in the SFA. Because in production theory an output cannot be negative, the standard normally distributed CFACTOR variable (N(0,1)) was mean-transformed by adding a constant to obtain the mean value 5.0, which does not have any effect on the results of the statistical analysis, however.
In addition, the following covariates were included in the analysis: age ($\bar{x}$ = 45.7, SD = 9.3 MIN = 27, MAX = 75) and gender (14.9% female, 85.1% male) of the PI, research project was conducted at the University of Vienna (31.7%) or not (68.3%), project duration ($\bar{x}$ = 40.9 months, SD = 7.9, MIN = 10, MAX = 60), and latent cluster of scientific disciplines (61.2% Life Sciences, 12.5% Formal Sciences, 26.3% Physical Sciences). These variables provided for the determinants of technical inefficiency (TIE).

(c) *Bibliometric data*: Included in the data analysis was the total number of publications (P), i.e., articles, reviews, letters, of a research project (Waltman, van Eck, van Leeuwen, Visser, & van Raan, 2011). The data were available in summarized and anonymized form for each project. Publication data were favored towards citations data for the reason that for citation data single publication citations should be available as well, in order to make certain corrections (e.g., fractional counting, field normalization), if necessary.

In the SFA the variables CFACTOR and P were used as output variables. Differences between the scientific disciplines were examined in the statistical analysis and not controlled beforehand through the use of field normalized citation scores (e.g.,

Table 1
Correlations among bibliometric and other variables (Spearman rank correlation) and sample description (N = 1,046 approved proposals).

No	Covariate	1	2	3	4
1	EXANTE	1.00			
2	CFACTOR	.03	1.00		
3	P	.11*	.35*	1.00	
4	GRANT	.17*	.29*	.22*	1.00
	Unit	100			[10K€]
	$\bar{x}$	89.2	5.12	5.5	19.7
	SD	4.3	0.75	5.5	8.3
	CV	4.85	14.8	101.3	42.1

Note: EXANTE = ex-ante peer evaluation of a proposal; P = number of publications; GRANT = approved grant sum.
* $p < .05$ (H_0: $\rho = 0$).

MNCS, TNCS). Serving as production factors or inputs in the production function were approved grant sum and EXANTE. Table 1 shows correlations and summary statistics on the included variables.

4.3. Bayesian stochastic frontier analysis (B-SFA)

In this study, for a sample of DMUs (i.e., research projects) there was a set of inputs X and outputs Y, S(X, Y), whereby the production function describes the frontier of the maximum output with given inputs. This function must fulfill the following conditions (Coelli et al., 2005, p. 12):

(1) *Nonnegativity*: The values of the production function are positive, real numbers.
(2) *Weak Essentiality*: The production of a positive output requires at least one input.
(3) *Monotonicity*: By addition of one unit of an input, the output may not decrease.
(4) *Concavity*: A linear combination of inputs produces an output that is no less than the sum of outputs produced by each input separately.

A special function that not only fulfills these conditions but also is suitable for questions in research evaluation and bibliometric analysis is the Cobb-Douglas production function (Greene, 2008, p. 107). It assumes that the (production) factors (e.g., labor, capital) contribute multiplicatively to output. This corresponds to empirical results in bibliometrics, according to which bibliometric indicators are lognormally distributed (Daniel, 1988; Limpert, Stahel, & Abbt, 2001; Stringer, Sales-Pardo, & Amaral, 2010). Whereas with normally distributed data a number of determining factors have an additive effect, with lognormally distributed data the factors have a multiplicative effect.

For a DMU *i* (i.e., a research project) the Cobb-Douglas production function of an output y can be defined as a function of two input factors X_1 (approved grant sum) and X_2 (EXANTE), as follows (Greene, 2008, p. 107):

$$y_i = f(x_{1i}, x_{2i}; A, \beta_1, \beta_2) = A x_{1i}^{\beta_1} x_{2i}^{\beta_2} \quad (1)$$

Regarding the economic characteristics of this production function, we refer the reader to Kumbhakar et al. (2015, p. 20f).

It should not be assumed that all DMUs are absolutely efficient at the frontier of the maximum output, as Eq. (1) implies. That is, the actual *TE* of a DMU, $TE = y_i/f(x_{1i}, x_{2i})$, is not always equal to 1.00. Further, there will be random shocks or random noise – that is, random effects that either prevent a DMU from achieving its maximum output or, vice versa, even lead the DMU to surpass the maximum expected output (e.g., a paper including a scientific breakthrough). Therefore, an error structure $f(u_i, v_i)$ must be added to the model (Eq. (1)), whereby u_i is the technical inefficiency of a DMU (TIE) and v_i the random shock:

$$y_i = f(x_{1i}, x_{2i}; A, \beta_1, \beta_2) f(u_i, v_i) = A x_{1i}^{\beta_1} x_{2i}^{\beta_2} e^{v_i - u_i}, \quad u_i \sim N_+(0, \sigma^2 u), \quad v_i \sim N(0, \sigma^2 v) \quad (2)$$

whereas u_i is independently and identically (i.i.d.) half-normal distributed, $N_+(0, \sigma^2_u)$, and v_i is i.i.d. normally distributed. In consequence, the total error structure, $\varepsilon_i = v_i - u_i$, should be negatively skewed, i.e., it should be skewed to the left (Kumbhakar et al., 2015, p. 56). This assumption can be tested empirically. The two error random variables *u* and *v* are stochastically independent of each other.

The term e^{-u_i} denotes the *TE* of the DMU *i*, varying from 0 (= non efficient) to 1 (=efficient). A DMU with *TE* = 1.0 is working at the frontier of maximum output for a given input, as specified by the production function (Eq. (2)).

It is convenient to use a logarithmic version of the production function (Fig. 1), which simplifies the model estimation:

$$\ln y_i = \beta_0 + \beta_1 \ln x_{1i} + \beta_2 \ln x_{2i} + v_i - u_i, \quad u_i \sim N_+(0, \sigma^2 u), \quad v_i \sim N(0, \sigma^2 v) \quad (3)$$

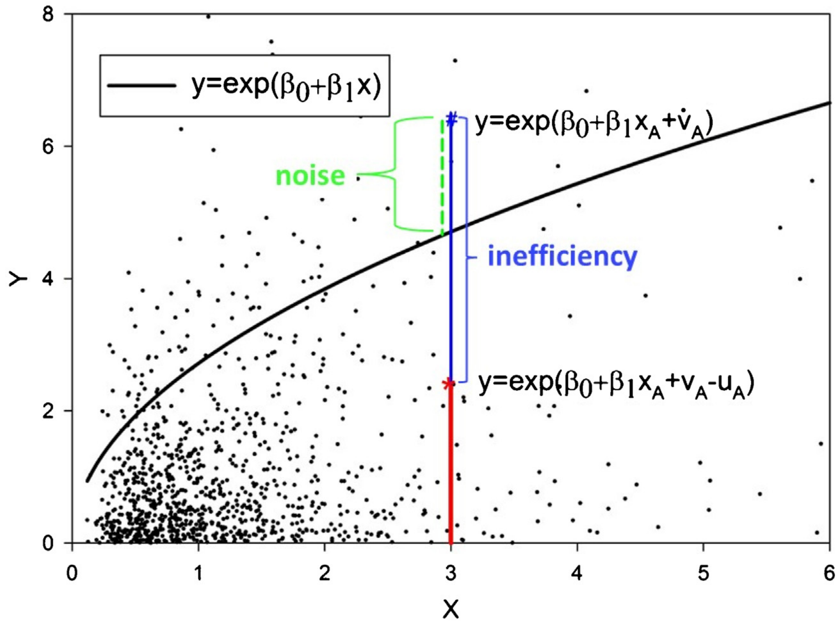


Fig. 1. Example of a stochastic production frontier (simulated data) with a funded project A, where the red star indicates the observed output value of A (cf. Coelli et al., 2005, p. 244). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

where $\beta_0 = \ln(A)$.

The basic model can be variously extended. In this contribution we focus on four model extensions:

(a) *Distributions of inefficiency u_i*: The standard distribution for u_i is the half-normal distribution. As alternatives, the truncated normal distribution with a location parameter $\bar{u}$ (mean technical inefficiency), the exponential distribution, and the gamma distribution were estimated.

(b) *Multilevel models*: We do not assume that the same production function holds for different scientific disciplines. For this reason, we used multilevel models to test whether the intercept and/or the slope of the production function varies across the scientific disciplines. The model (Eq. (3)) was extended as follows (Greene, 2005; Hox, 2010; Tsionas, 2002, p. 129):

$$\ln y_i = (\beta_0 + b_{0j}) + (\beta_1 + b_{1j})\ln x_{1i} + \beta_2 \ln x_{2i} + v_i - u_i, \tag{5}$$

where b_{0j} and b_{1j} are the random effects of the intercept (β_0) and the slopes (β_1) for each scientific discipline *j*, which are assumed to be normally distributed with variance components σ^2_{b0} and σ^2_{b1}.

In addition, we tested whether the scientific disciplines differed in the variances of technical inefficiency and random error, σ^2_u, and σ^2_v. A multilevel model was formulated using an exponential link function:

$$\sigma_u^2 \sim \exp(\alpha_0 + a_{0j}), \tag{6}$$

where α_0 is the fixed intercept and a_{0j} is the random intercept for scientific discipline *j*, which is assumed to be normally distributed ($a_{0j} \sim N(0,\sigma^2_a)$). Similarly, an explanation model for σ^2_v can be formulated. Random effects models usually require a large data set (Olivares & Wetzel, 2014, p. 662), as is the data set in our study.

(c) *Exogenous determinants of TIE*: In line with the hypotheses formulated in this study, different covariates Z were included to explain technical inefficiency, u_i. To include covariates, the statistical procedure suggested by Wang and Schmidt (2002) was applied (Kumbhakar et al., 2015, p. 85f). This approach was favored for the reason that the overall shape of the TIE distribution remains constant (*basic distribution*); only the scale of the distribution changes according to a multiplication factor as a function of Z, $h(\alpha, Z)$:

$$u_i^* = h(a, z)u_i, \tag{7}$$

where $h(\alpha, z) > 0$. The function $h(\alpha, z)$, the *scaling function*, could be expressed as an exponential function, $h(\alpha, z) = \exp(\alpha z)$ with parameters α. If, for instance, h = 0.8 for Life Science, the variance of inefficiency, σ^2_u, will be reduced by the factor 0.8^2 (=0.64). In other words, the amount of inefficiency of a funded project in Life Science will be reduced by 0.80 in comparison with the total sample, and vice versa, the TE $\exp(-u_i)$ will be higher than in the total sample. Before including the covariates in the model, the continuous variables were *z*-transformed ($\bar{x} = 0$, SD = 1), and the categorical variables were effect-coded. By this procedure the mean value of all variables is zero, and the h-function value equals 1.0 (h = exp(0) = 1.00), i.e., there is no scaling.

(d) *Count data*: Whereas the CFACTOR can be treated as normally distributed variable, especially the random error component, bibliometric data, such as number of publications, are by nature count data (positive integer values including 0) and therefore Poisson-distributed, implying a logarithmic link function (Hilbe, 2012; Lindsey, 1995). However, the literature on SFA for count data is scarce and mostly published in the form of reports (Drivas, Balafoutis, & Rozakis, 2014; Drivas, Economidou, & Tsionas, 2014; Fé & Hofler, 2013). Alternatively, the simple use of logarithmic transformed variables (log(x + 1)) could lead to biased estimates, as a simulation study showed (O'Hara & Kotze, 2010). Therefore, we applied the Poisson SFA model suggested by Fé and Hofler (2013), which we extended to a negative binomial SFA model. This might better meet the specific data conditions in bibliometrics (overdispersed data) than the rather restrictive Poisson distribution (with the assumption that the mean equals the variance of an empirical distribution). The basic equations can be formulated as follows:

$$\begin{aligned} y_i &\sim Poisson(\lambda_i) \\ E(y_i) = \lambda_i &= \exp(\beta_0 + \beta_1 \ln x_{1i} + \beta_2 \ln x_{2i} - u_i)\mathrm{v}_i, \\ \mathrm{u_i} &\sim \mathbf{N}_+(0, \sigma^2 \mathrm{u}) \\ \mathrm{v_i} &\sim \mathrm{GAM}(\mathrm{k}, \mathrm{scale} = 1/\mathrm{k}) \end{aligned} \quad (8),$$

where the expected value, $E(v_i)$, equals 1.00 and the variance of v_i, $VAR(v_i)$, equals 1/k.

As mentioned above, the residuals (v_i-u_i) should be negatively skewed. As bibliometric data are usually positively skewed, with a high proportion of papers with low or zero citations and a low proportion of papers with high citations, negatively skewed residuals in a SFA were not very likely. In this case the ordinary least squares SFA or the maximum likelihood SFA can no longer be applied, especially as their applications might lead to inefficiency variances, σ^2_u, of zero (Parmeter, 2014, p. 27f).

The estimation of the model parameters was obtained by applying a Bayesian approach (Bornmann, Stefaner, de Moya Anegón, & Mutz, 2016; Griffin & Steel, 2007; Hoff, 2009; Kaplan, 2014; Lunn, Jackson, Best, Thomas, & Spiegelhalter, 2013; O'Hagan, 2008; Olivares & Wetzel, 2014; van den Broek, Koop, Osiewalski & Steel, 1994; SAS Institute Inc., 2014). Models can be compared and selected by information criteria like the deviance information criterion (DIC) instead of null hypothesis statistical testing. The lower the DIC is, the better the model fits the data. For covariate selection the DIC was used as well.

More statistical details on the models and the estimation process can be found in the supplementary material (Appendix A, Bayesian approach to SFA).

5. Results

5.1. Simulation study

In view of the fact that there are few existing applications of Bayesian SFA and especially SFA for count data, in a first step we conducted a simulation study. We wanted to find out how the SFA procedure behaves under various sampling conditions. Following an experimental design, three conditions were varied (2 × 4 × 5 design): the scale level (continuous data versus count data), the sample size (100, 500, 1,000, 1,500), and the γ coefficient, which indicates the amount of inefficiency variance to the total residual variance (0.30, 0.50, 0.70, 0.90, 0.99). The size of the bias of the parameter estimation was the dependent variable, whereby bias was defined as the deviation of the estimated parameter from the true parameters in percent of the true parameters. The coefficients of the production function showed a small bias under most of the conditions, except for the conditions small sample size (N = 100) or low value of the γ coefficient (=0.30). In contrast, the bias with inefficiency variance, σ^2_u, and the γ coefficient were dependent on the value of the γ coefficient and sample size. Small sample size and/or small γ coefficient resulted in a considerable bias. What was surprising was that the Poisson SFA resulted in more or less unbiased estimates under all conditions. More information on the method and the results of the simulation study can be found in the Supplementary Material, Appendix B (see Table B1).

5.2. Bayesian stochastic frontier analysis – basic models

5.2.1. Model comparison

Before performing SFA, it is necessary to check whether SFA is at all appropriate with the given data or whether the better choice is (multiple) regression (assumption of projects with perfect TE). Another question needing clarification concerns the distribution of the technical inefficiencies. These questions will be answered in the following by comparing different models. The comparison criterion used is the deviance information criterion (DIC). The lower the DIC value is, the better the model fit. A weight additionally indicates how important the difference between one model and another is (Ehlers, 2011, p. 2438). The closer the weight is to 1.0, the better the model is compared with all others and the more strongly it stands out from the other models.

The model comparison (Table 2) was done separately for the two output variables (CFACTOR, number of publications) and for different distribution assumptions (regression, half-normal, truncated normal, gamma, exponential).

Table 2
Model comparison for basic Bayesian SFA models (Cobb Douglas production function) with the deviance information criterion (DIC) (N = 1,046 FWF funded projects).

No	Model	DIC	$\overline{D}$	p_D	R	Con	w
Log(CFACTOR)							
1	Regression	−1,085.1	−1,089.1	4.0	5	+++	0.00
2	Half-normal	**−4,288.4**	**−1,767.9**	**−2,520.5**	**1**	+++	1.00
3	Truncated normal	−1,417.1	−1,895.0	477.9	2	–	0.00
4	Gamma	−1,165.9	−1,399.3	233.4	4	-/+	0.00
5	Exponential	−1,195.3	−1,510.2	310.5	3	++	0.00
NUMBER OF PUBLICATIONS							
1	Poisson Regression	7,429.8	7,426.8	3.0	5	+++	0.00
2	Half-normal	**−7,875.8**	**4,540.6**	**−12,416.5**	**1**	+++	1.00
3	Truncated normal	5,149.6	4,450.5	699.1	3	-/+	0.00
4	Gamma	5,095.1	4.323.8	771.3	2	+++	0.00
5	Exponential	5,568.8	4,696.87	871.4	4	++	0.00

Note: $\overline{D}$ = posterior mean of deviance, p_D = effective number of parameters, DIC = $\overline{D}$+ p_D, R = model rank ("1" = best one, in boldface); Con = scale of convergence of the MCMC iteration process (— = no convergence, . . ., +++ = perfect convergence); w = weight of the model.

For the output variable CFACTOR an SFA model with half-normally distributed inefficiencies outdid all other models (weight = 1.0), especially also the regression model, which takes on the absolutely perfect TE of 1.0 (or vice versa TIE = 0) for each research project. Additionally, this model showed satisfying convergence properties. Thus, for CFACTOR, an SFA could be used successfully, and it was also better than a simple regression model. Additionally, if the error component, e_i, is normally distributed and the technical inefficiencies, u_i, are half-normal normally distributed, then the resulting residuals should be skewed to the left or should be negatively skewed (Kumbhakar et al., 2015, p. 56). Considering a skewness of $\sqrt{b_1} = -0.34$ and a statistically significant Coelli test (M3T = −4.53, $p < .05$), the residuals of SFA for CFACTOR were actually negatively skewed. Here, approximately 50% of the residual variance stemmed from the inefficiency variance, $\sigma^2{}_u$, (Table 4, M_1).

For the output variable number of publications, the SFA model with half-normal distribution of the technical inefficiencies was clearly better than all other models (w = 1), especially also better than the Poisson regression model. The convergence of the MCMC iteration process was perfect. The residuals, which were estimated with an ordinary SFA (log(P)), were, however, as expected positively skewed ($\sqrt{b_1} = .47$, M3T = 6.17, $p < .05$). Although an ordinary SFA model could not be used, the Bayesian approach resulted in unbiased estimates (see simulation study, Appendix B). In the case of count data, additionally the ratio of the Pearson χ^2 (expected values by the model and observed counts) and the degrees of freedom should be inspected, which should not strongly deviate from 1.0. In our case a value of 1.24 was obtained, which indicates some overdispersion (variance is higher than the mean). In the case of moderately-sized data a value of 1.25, and in the case of a huge sample size a value of 1.04, can be tolerated (Hilbe, 2014, p. 82). Considering the optimal properties of pure Poisson SFA models (see the simulation study in the Appendix), we abstained from modeling the data with a negative binomial SFA model and applied a Poisson SFA model.

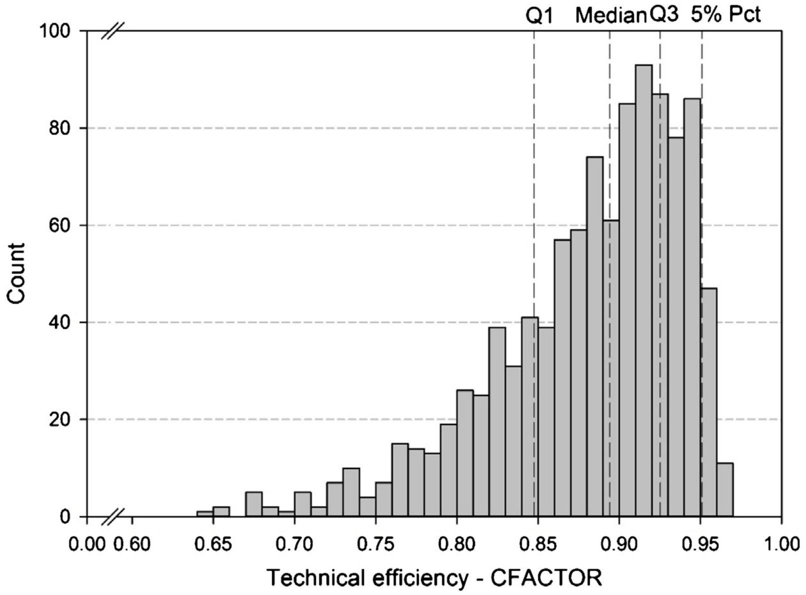
5.2.2. CFACTOR

Regarding model M_2 (Table 2, half-normal distribution) for each FWF-funded project the posterior distribution of TE, exp(-u_j), was estimated. The mean value of this distribution provided for the TE value of the respective research project. Fig. 2 (top) shows a histogram of TE for all projects. The distribution of TE was as expected negatively skewed (half-normal). A TE value of 1.0 indicates that a research project produced output with maximum efficiency considering the overall production function in a field – that is, right at the frontier that the production function specifies. The research projects showed a wide range in their TE, from minimal 0.69 to maximal 0.97, with a standard deviation of 0.16. But the mean TE was quite high at 0.89, which means that 50% of all funded research projects had TE values higher than 0.89 and 5% of the research projects even had TE values above 0.95. Thus, overall, the projects funded by the FWF were on average very efficient with regard to the output variable CFACTOR.

A good impression of the production frontier can be gained from a plot of the raw data of approved grant sum and CFACTOR with the associated production function (Fig. 3, top). Unlike in the statistical analysis, the output variables were not logarithmized for the figure, so that the dispersion of the raw values seemed especially high. Also the nonlinear character of the production function as the frontier of the maximum output with given input was visible. For example, with a financial input or grant sum of 10 k€, a maximal output of 5.5 could be reached, which is a value just slightly above the mean CFACTOR (=5.0). The second factor EXANTE is not important, as Fig. 4 (top) shows as well.

5.2.3. Number of publications

Regarding the output variable number of publications, the distribution of TE was as expected positively skewed (Fig. 2, bottom). Only a few research projects showed high TE. TE ranged from minimal 0.10 to maximal 0.97 with an estimated standard deviation, σ_e, of 0.17 (Table 4, M_1). The median TE was accordingly low at 0.27, which means that 50% of all funded



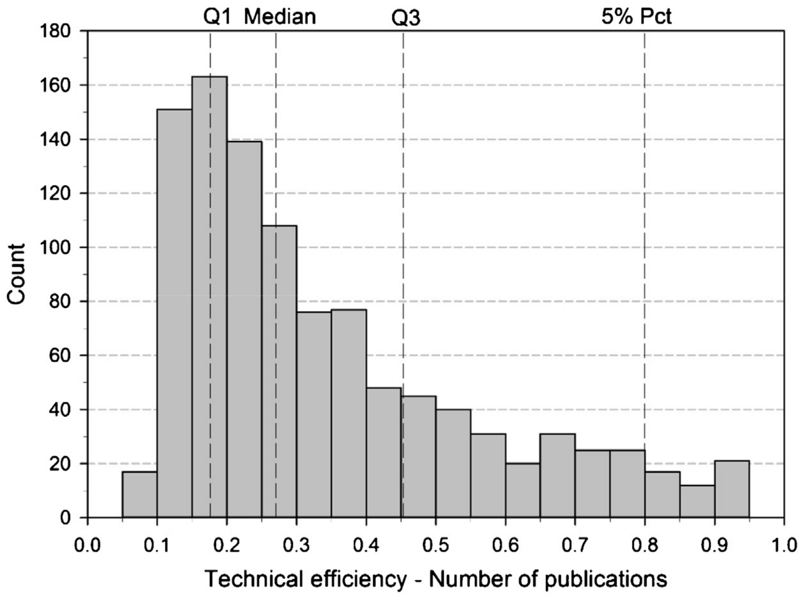


Fig. 2. Histogram of the technical efficiency for each of the 1046 FWF funded projects (means of the posterior mean of the posterior parameter distribution) for CFACTOR (top), and number of publications (bottom).

research projects had TE values above 0.27. Five percent of the research projects had TE values above 0.80. The TE values with respect to CFACTOR and number of publications were correlated with a Spearman rank correlation of $r_s = 0.31$. The production function (Fig. 3, bottom) showed the frontiers of production. For example, with a financial input of 10k€, a maximum of about 6 publications was expected and with input of 40k€, 12 publications. Furthermore, not only the grant sum, but also the EXANTE played a role (Fig. 4, bottom). In contrast to CFACTOR, the isoquants of the same number of publications were not horizontal but arc-shaped, which pointed to the substitution elasticity of the production factors (approved grant sum, EXANTE) in the Cobb-Douglas production function. To a certain extent, a higher approved grant sum was able to compensate for a lower EXANTE.

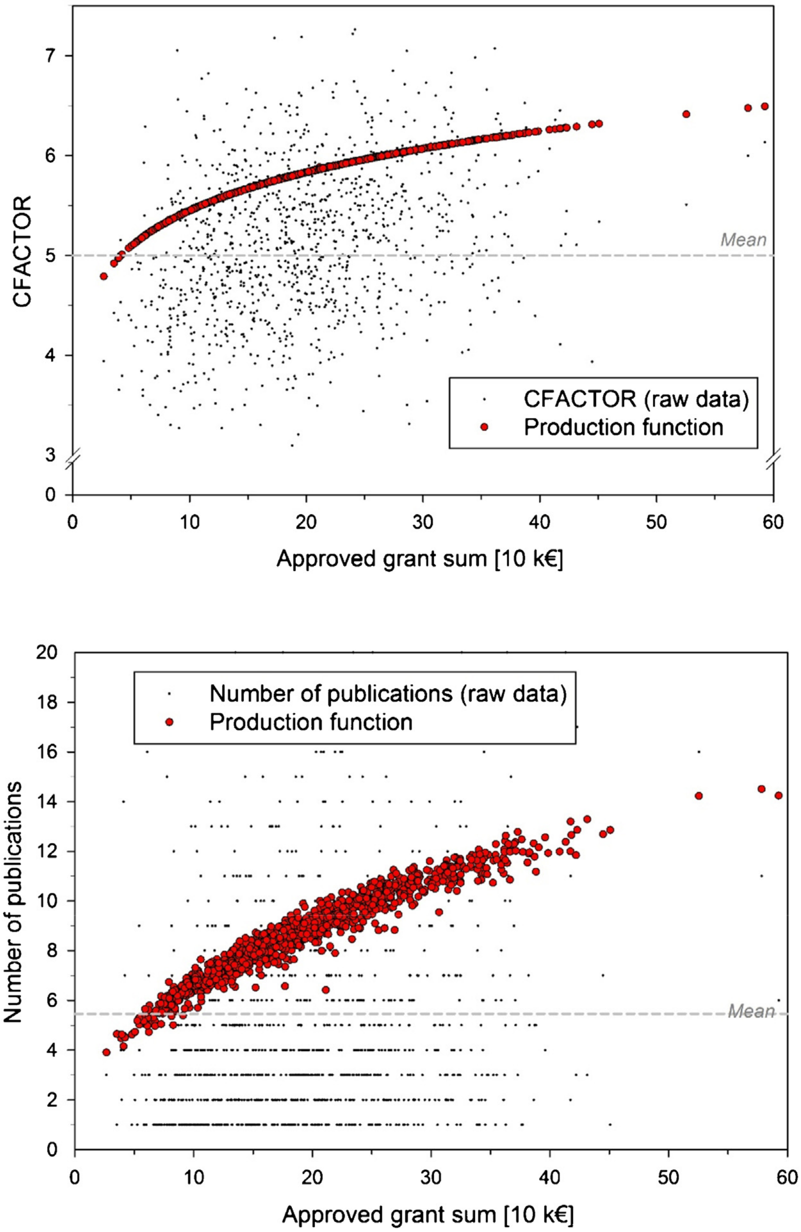


Fig. 3. Production functions for CFACTOR (top) and number of publications (bottom).

5.3. Bayesian stochastic frontier analysis – sophisticated models

5.3.1. Model overview

Above we clarified the principal possible applications of SFA at least for the variables CFACTOR and number of publications. In the following, various sophisticated models were tested (Table 3). To do this, we created various models that represent different assumptions and then, as we did with the basic models, compared them using the DIC.

As a starting model or null model, we chose for both variables the SFA model with half-normal distribution of inefficiencies (TIE), in which the mean or expected value of TIE can deviate slightly from 0. In a first category of models (multilevel models) we tested whether different production functions in the intercept and slope of requested grant sum can be assumed for different main scientific disciplines, or whether there was a single production function for all scientific disciplines included in this study. Alternatively, in a second category of models, we tested the residuals (variance heterogeneity). This was a test of the extent to which under the assumption of an overall production function there were discipline-specific random error variance components (M_4) or TIE variance components (M_5). In a third category of models (exogenous determinants

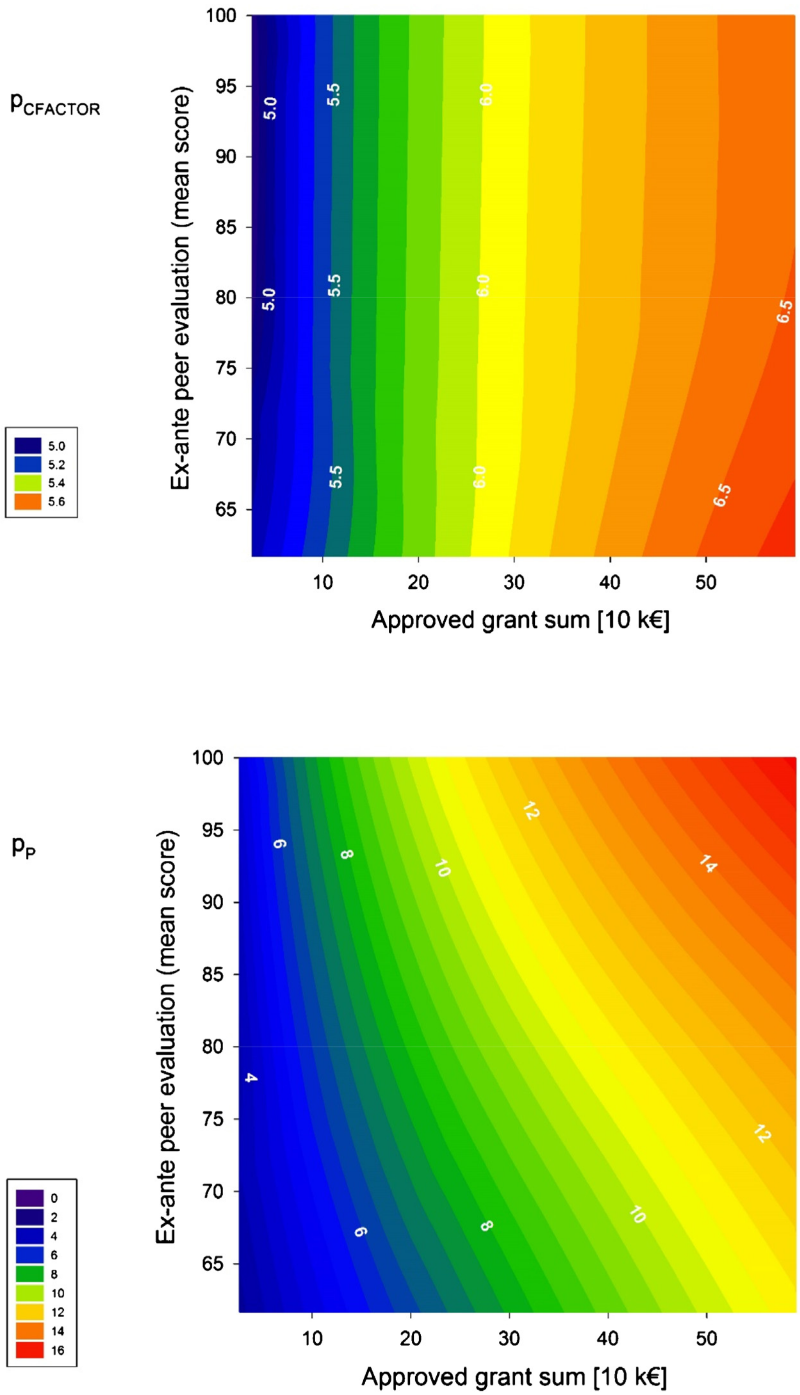


Fig. 4. Contour plots (isoquants) of EXANTE, approved grant sum, and the corresponding values of the production function for CFACTOR (top), and for number of publications (bottom).

of TIE), explanatory covariates for TIE were included (M_6-M_{12}). In a final category of models, all exogenous determinants of TIE were tested in a multiple regression (M_{13}). In a multiple regression the effects of the single covariates were statistically controlled by the other covariates in the model. Note that by including covariates with no explanatory power, the model could still improve in terms of a smaller DIC, because the h-weight becomes 1.0 (exp(0) = 1), and then the model fits at least as well as a model without covariates or even better.

Table 3
Model comparison for sophisticated models with the deviance information criterion (DIC) (N = 1,046 FWF funded projects).

N	Model description	log(CFACTOR)				P			
		DIC	R	Con	w	DIC	R	Con	w
1	Null model (half-normal)	−4,288.4	3	+++	0	−7,875.8	4	+++	0
Multilevel model (main disciplines)									
2	ML: Random intercept	−4,067.4	9	+	0	−4,259.2	11	-/+	0
3	ML: Random intercept/slope	−4,051.5	10	+	0	−4,150.3	12	+	0
Variance heterogeneity (main disciplines)									
4	Heterogeneity $\sigma^2{}_e$	**−4,508.5**	**1**	**+++**	**1**	-	-	-	-
5	Heterogeneity $\sigma^2{}_u$	−3,724.0	12	+	0	−7,501.4	8	+	0
Determinants of TIE									
6	Approved grant sum	−4,227.4	5	+++	0	**−7,928.6**	**1**	-/+	**1**
7	EXANTE	−4,214.8	6	+++	0	−7,846.7	6	+++	0
8	LCluster	−4,196.4	7	+++	0	−7,459.2	9	++	0
9	Gender of PI (female)	−4,147.6	8	+++	0	−7,880.7	2	+++	0
10	Age of PI	−4,273.9	4	+++	0	−7,879.1	3	+++	0
11	Univ. of Vienna or not	−4,300.0	2	+++	0	−7,869.2	5	++	0
12	Project duration	−3,784.6	11	+++	0	−7,760.4	7	+++	0
Additional models									
13	All determinants of TIE	−2,777.6	13	+++	0	−7,385.0	10	+	0

Note: Rank = model rank ("1" = best, in bold face); Con = scale of convergence of the MCMC iteration process (— = no convergence, . . ., +++ = perfect convergence); w = weight of the model.

Table 4
Parameter estimation for selected models for log(CFACTOR).

		M_0		M_1		M_2		M_3		M_4	
		(Table 2, No = 1)				(Table 2, No = 2)		(Table 3, No = 4)		(Table 4, No = 13)	
		Bayesian reg.		ML SFA half		Bayesian SFA half		Bayesian SFA-Het($\sigma^2{}_e$)		All covariates	
Variable	Par	$\bar{x}$	SD	Value	SE	$\bar{x}$	SD	$\bar{x}$	SD	$\bar{x}$	SD
Fixed effects											
Production function											
Intercept	β_0	1.19	0.61	1.30	0.60	1.30+	0.60	1.30+	0.61	1.02	1.12
LOG(GRANT)	β_1	0.10+	0.01	0.10*	0.01	0.10+	0.01	0.10+	0.01+	0.09+	0.02
LOG(EXANTE)	β_2	−0.01	0.09	−0.01	0.09	−0.01	0.09	−0.01	0.09	0.03	0.17
Determinants of u (TIE)											
Grant sum	α_1									0.08	0.08
EXANTE	α_2									0.01	0.09
Life Sciences & Med.	α_3									0.08	0.06
Physical Sciences	α_4									0.01	0.07
Gender PI (female)	α_5									−0.17+	0.08
Age PI	α_6									0.10+	0.04
Univ. ofVienna	α_7									0.02	0.04
Project duration	α_8									−0.21+	0.06
Random effects											
θ_j (discipline)	u_θ							−4.57	0.16		
	$\sigma^2{}_\theta$							0.03	0.03		
	γ			0.50	–	0.47	0.07	0.49	0.08	0.23	0.10
u_i (TIE)	σ_u			0.17*	0.01	0.16	0.01	0.17	0.01	0.10	0.02
e_{ji} (Error)	σ_e	0.14	0.003	0.10*	0.01	0.10	0.01	0.10	0.01	0.12	0.01

Note: M = mean of the posterior parameter distribution; SD = standard deviation of the posterior parameter distribution; SE = standard error; GRANT = approved grant sum; EXANTE = ex ante peer evaluation of a proposal; Bayesian reg. = Bayesian regression, ML SFA half = maximum likelihood half-normal SFA; Bayesian SFA half = Bayesian half-normal SFA; Bayesian SFA – Het($\sigma^2{}_e$) = Bayesian half-normal SFA – heterogeneity of the residual variance.

* $p < .05$.

\+ 0 is not in the 95% HPD credible interval (fixed effects).

5.3.2. CFACTOR

For the output variable CFACTOR, using the DIC, the assumption of discipline-specific production function models (M_2, M_3, M_5) could be clearly excluded. None of those models was better than the null model. Thus, different production functions or variance components of TIE for the 12 different disciplines were not assumed. However, there were discipline-specific error variance components (M_4). The disciplines differ in the variability of random shocks. In addition, with the exception

Table 5
Parameter estimation for selected models for number of publications (P).

		M_0		M_1		M_2		M_3	
		(Table 2, No = 1)				(Table 2, No = 2)		(Table 4, No = 13)	
		Bayes Pois reg.		ML SFA half		Bayesian SFA half		All covariates	
Variable	Par	$\bar{x}$	SD	Value	SE	$\bar{x}$	SD	$\bar{x}$	SD
Fixed effects									
Production function									
Intercept	β_0	−6.55*	1.07	−6.56	4.89	−5.44	3.57	−4.94	4.34
LOG(GRANT)	β_1	0.45*	0.03	0.32*	0.05	0.51+	0.06	0.58+	0.07
LOG(EXANTE)	β_2	0.87*	0.16	0.96*	0.42	0.81	0.53	0.68	0.64
Determinants of u (TIE)									
Grant sum	α_1							0.07	0.04
EXANTE	α_2							−0.00	0.04
Life Sciences & Med.	α_3							0.16+	0.04
Physical Sciences	α_4							−0.25+	0.05
Gender PI (female)	α_5							0.04	0.04
Age PI	α_6							0.01	0.03
Uni Vienna	α_7							−0.03	0.03
Project duration	α_8							−0.09+	0.03
Random effects									
u_i (TIE)	σ_u			0.00	5.04	1.60	0.05	1.49	0.08
e_{ji} (Error)	σ_e			0.66	0.01				
Pearson χ^2/df		5.22	0.07	–	–	1.24	0.06	1.28	0.06

Note: M = mean of the posterior parameter distribution; SD = standard deviation of the posterior parameter distribution; SE = standard error; GRANT = approved grant sum; EXANTE = ex ante peer evaluation of a proposal; Bayes Pois reg. = Bayesian Poisson regression; ML SFA half = maximum likelihood half-normal SFA with log(P); Bayesian SFA half = Bayesian half-normal SFA.

* $p < .05$.

+ 0 is not in the 95% HPD credible interval (fixed effects).

of the covariate 'University of Vienna – Yes or No,' no exogenous determinants of TIE appeared to result in a lower DIC than the DIC of the null model, although most models showed perfect convergence of the Markov chain Monte Carlo iteration process.

More detailed information on the chosen models is provided by the parameter estimations (Table 4); the table shows the different models in the columns and the parameter estimations for fixed and random effects in the rows (Greene, 2005). Across all models the main production factor was the approved grant sum ($\beta_1 = 0.10$); the EXANTE, in contrast, played no important role (not statistically significant; 0 is not in the highest posterior density credible interval; see Appendix B). When the same simple SFA model was estimated using the maximum likelihood (ML) procedure (M_1) and then estimated using a Bayesian approach (M_2), there were no noteworthy differences in the parameters. This meant that the results (mean values of the posteriori distribution) could in fact be interpreted like an ML parameter estimation. In model M_3, in addition to model M_2, heterogeneity of the random error components, σ^2_e, was allowed, i.e., σ^2_{ej}. = exp(θ_j), where θ_j is normally distributed across the 12 main disciplines with mean value α (=−4.57) and variance σ^2_θ (=0.03).

Of special interest was the last model (M_{13}), which performed relatively poorly in the model comparison but pointed to important exogenous determinants of TIE, the covariates gender of PI, age of PI, and project duration. Men, younger applicants and projects with above average project duration have lower TIE or higher technical efficiency (TE) than women, elder applicants, and projects with below average duration, respectively. For example, when a research project was headed by a woman, the TIE decreased by the factor h = exp(-0.17*1) = 0.84 as compared with a project headed by a man, where women are coded 1 and men are coded −1. With a man as PI, the TIE value of a project even increased by the factor h = exp(-0.17*-1) = 1.19. A project with, for example, a TIE value of −0.80 (basic distribution) and a woman as PI would have a TIE value of 0.84*-0.80 = −0.67 and in contrast with a man as PI a TIE value of 1.19*-0.80 = −0.95. Expressed in TE values (TE = exp(h*-u)), a project with a woman as PI would then have a TE value of 0.51, and a project with a man as PI a TE value of only 0.39.

No noteworthy effect on TIE was found, however, for the following covariates: discipline, size of university (University of Vienna or not), approved grant sum, and EXANTE.

5.3.3. Number of publications

In the model comparison (Table 3), as with CFACTOR, the assumption of discipline-specific production function models for number of publications (M_2, M_3, M_4, M_5) could be clearly excluded. Instead, some individual exogenous determinants of TIE were revealed to be important exogenous determinants of TIE with lower DIC than the null model.

The parameter estimations of selected models (Table 5) yielded more detailed information. Very unfavorable was a simple Bayesian Poisson regression with a Pearson χ^2/df ratio of 5.22, which differs significantly from 1.00 – that is, it indicates strong overdispersion. In contrast to the output variable CFACTOR, there were strong differences between the results of an ordinary SFA (M_1) and a Bayesian Poisson SFA half-normal (M_2): In the ordinary SFA, the variance of TIE ($\sigma^2{}_u$) was zero; in the Bayesian Poisson SFA it was 1.60. This difference could be explained by the positively instead of negatively skewed residuals, which was necessarily required in the case of an ordinary SFA.

Of particular interest was again the last model (M_3), which included all covariates for explaining TIE in a regression model. Important exogenous determinants of TIE were the following two: scientific discipline (Life Science and Medicine, Physical Sciences) and project duration.

6. Discussion

In connection with the bibliometric indicators that are commonly used for evaluation of research, there is criticism of size-independent indicators based on the mean number of citations (e.g., MNCS). The recommendation is to instead take input-output relations into account when examining research performance (Abramo & D'Angelo, 2016). From the perspective of research funding, the performance-based university research funding approach explicitly calls for similar questions about the effectivity and impact of funding and for efficiency analyses (Hicks, 2012), but without specifying how these efficiency analyses should be conducted.

The aim of this study was to present the statistical approach of stochastic frontier analysis for the analysis of input-output relations in research evaluation, taking the example of the FWF, Austria's leading funding organization for basic research. Up to now, mostly nonparametric DEA has been used for productivity and efficiency analysis in scientometrics. DEA does not as a rule consider the stochastic nature of bibliometric data. However, so far little use has been made of SFA in bibliometric studies (e.g., Bolli et al., 2016; Drivas, Balafoutis et al., 2014), which makes a detailed description of this approach necessary. This study connects up with a number of bibliometric studies that applied DEA (e.g., Abbott & Doucouliagos, 2003; Abramo et al., 2011).

This study focused on four questions on productivity and efficiency analysis: *(a) Can a production function, i.e., frontier of research performance, be identified for research projects? (b) How high is the TE of FWF-funded research projects? (c) What are possible exogenous determinants of TE? (d) How does the statistical procedure behave under different sampling conditions, such as sample size, scale level (continuous data, count data), and proportion of random noise?*

With regard to a frontier of scientific performance, we were in fact able to identify a production function for the two output variables CFACTOR (a latent dimension of the research output extracted from an analysis of research project reports) and number of publications. This reveals, despite random fluctuations, frontiers of production of research results that are especially dependent on financial investment (approved grant sum) but also on the ex-ante peer evaluation, at least regarding number of publications. Smaller discipline-specific differences in the SFA were found only for CFACTOR. The scientific disciplines differ in the size of the random shocks – that is, the random deviations from the production function. The lack of greater differences in the production function between the scientific disciplines can be explained by the fact that CFACTOR is a latent dimension that was estimated as part of a latent class analysis and is exempt from discipline-specific differences. The TE of FWF-funded research projects is high, in view of the CFACTOR with a median of 89%, which means that 50% of the projects achieve a TE value of over 89% of the maximum possible output. This output is specified by the production function for a given financial investment (approved grant sum) and a certain grade point average in the ex-ante peer evaluation. In a direct comparison, the median TE regarding the total number of publications is clearly lower at 0.27, which means that 50% of all funded research projects have a TE value above 0.27, and 5% of the research projects have a TE value above 0.80. This is also due to the skewed distribution, with few research projects having a very high number of publications.

With regard to CFACTOR, exogenous determinants of TIE or TE are gender of PI, age of PI, and project duration (hypotheses 1, 2, and 7). Female PIs, younger PIs, and projects with longer project durations have slightly higher TE values than male PIs, older PIs, and projects with shorter project durations. For number of publications, determinants are discipline (LCLUSTER) and also project duration (hypotheses 5 and 7): Physical Sciences have slightly higher TE values than other scientific disciplines, and projects with longer project durations have a slightly higher TE than projects with shorter project durations.

The simulation study showed that under most sampling conditions the Poisson SFA has only a small bias (<5%) for all estimated parameters, as does also the SFA for continuous outcomes. The negative binomial SFA, however, demands larger samples or a lower proportion of random shocks/random noise.

Finally, some limitations of this study should be mentioned:

- *Bibliometric data*: The data were available in anonymized form for each project, which limits bibliometric analysis (e.g., no access to single publications, no possibility of fractional instead of full counting of publications).
- *Sample*: The study examined a subsample of FWF-funded projects that published their findings mainly as journal articles (from research projects in Life Science and Medicine, Formal Sciences, Physical Sciences). This limits the generalizability of the results (for the FWF).
- *Production function*: The Cobb-Douglas production function with its multiplicative character (Limpert et al., 2001) proves to be especially appropriate for bibliometric efficiency analyses. In principle, also more complicated production functions are possible (e.g., translog), but this would entail stricter data requirements.

- *General restriction*: Production efficiency analysis in general has certain limitations, which are discussed by Abramo and D'Angelo (2014).

All in all, the Bayesian SFA modeling approach proves to be a very flexible instrument of data analysis: Different model assumptions (e.g., discipline specific production functions, distribution assumptions) can be tested, and the approach allows for random noise and can be used with raw bibliometric data. Beyond that, various model components can be combined (e.g., latent class analysis, regression analysis of TIE on a set of exogenous determinants, multilevel models). In addition, for each DMU (e.g., research project) it is possible to estimate TE as a parameter (with posterior distribution).

With regard to SFA for count data, it is recommendable to calculate a Poisson SFA first. If the coefficient Pearson χ^2/df does not deviate strongly from 1.00, a Poisson SFA can be estimated, and this largely independently of the particular sampling conditions. If the ratio Pearson χ^2/df deviates strongly from 1.0 (1.25 with moderated-sized samples, 1.05 with huge samples), a negative binomial SFA should be chosen, but this requires a larger sample and/or a higher γ coefficient.

Author contribution

Rüdiger Mutz: Conceived and designed the analysis, Collected the data, Contributed data or analysis tools, Preformed the analysis, Wrote the paper.

Lutz Bornmann: Conceived and designed the analysis, Collected the data, Wrote the paper.

Hans-Dieter Daniel: Conceived and designed the analysis, Contributed data or analysis tools, Wrote the paper.

Appendix A. and B Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.joi.2017.04.009.

References

Abbott, M., & Doucouliagos, C. (2003). The efficiency of Australian universities: A data envelopment analysis. *Economics of Education Review*, *22*, 89–97.

Abramo, G., & D'Angelo, C. A. (2014). How do you define and measure research productivity? *Scientometrics*, *101*, 1129–1144.

Abramo, G., & D'Angelo, C. A. (2016). A farewell to the MNCS and like size-independent indicators. *Journal of Informetrics*, *10*(2), 646–651.

Abramo, G., Cicero, T., & D'Angelo, C. A. (2011). A field-standardized application of DEA to national-scale research assessment of universities. *Journal of Informetrics*, *5*, 618–628.

Abramo, G., D'Angelo, C. A., & Murgia, C. A. (2016). The combined effects of age and seniority on research performance of full professors. *Science and Public Policy*, *43*(3), 301–319.

Archibald, R. B., & Feldman, D. H. (2008). Graduation rates and accountability: Regression versus production frontiers. *Research in Higher Education*, *49*, 80–100.

Athanassopoulos, A. D., & Shale, E. (1997). Assessing the comparative efficiency of higher education institutions in the UK by the means of Data Envelopment Analysis. *Education Economics*, *5*(2), 117–134.

Ballou, J., Mishkind, M., Mooney, G., & van Kammen, W. (2002). *National science foundation report on efficiency of grant size and duration*.. Retrieved from. https://www.nsf.gov/pubs/2004/nsf04205/mathematica_nsfrptfinal6.pdf

Bessent, A. M., Bessent, E. W., Charnes, A., Cooper, W. W., & Thorogood, N. C. (1983). Evaluation of educational-program proposals by means of DEA. *Educational Administration Quarterly*, *19*, 82–107.

Bolli, T., Olivares, M., Bonaccorsi, A., Daraio, C., Aracil, A. G., & Lepori, B. (2016). The differential effects of competitive funding on the production frontier and the efficiency of universities. *Economics of Education Review*, *52*, 91–104.

Bonaccorsi, A., & Daraio, C. (2004). Econometric approaches to the analysis of productivity of R&D systems. In H. F. Moed, W. Glänzel, & U. Schmoch (Eds.), *Handbook of quantitative science and technology research* (pp. 51–74). New York: Kluwer.

Bornmann, L., Stefaner, M., de Moya Anegón, F., & Mutz, R. (2016). Excellence networks in science: A Web-based application based on Bayesian multilevel logistic regression (BMLR) for the identification of institutions collaborating successfully. *Journal of Informetrics*, *10*(1), 312–327.

Charnes, A., Cooper, W. W., & Rhodes, E. (1978). Measuring the efficiency of decision making units. *European Journal of Operational Research*, *2*, 429–444.

Chen, C.-M., Delmas, M. A., & Lieberman, M. B. (2015). Production frontier methodologies and efficiency as a performance measure in strategic management research. *Strategic Management Journal*, *36*, 19–36.

Coelli, T. J. C., Rao, D. S. P., O'Donnell, C. J., & Battese, G. E. (2005). *An introduction to efficiency and productivity analysis* (2nd ed.). New York: Springer.

Daniel, H.-D. (1988). Methodische Probleme institutsvergleichender Analysen der Forschungsproduktivität [Methodical problems of institutional comparisons of research productivity]. In H.-D. Daniel, & R. Fisch (Eds.), *Evaluation von Forschung, Methoden – Ergebnisse – Stellungnahmen [Evaluation of research, methods – results – statements]* (pp. 215–241). Konstanz: Universitätsverlag.

Drivas, K., Balafoutis, A. T., & Rozakis, S. (2014). *Research funding and academic output: The case of Agricultural University of Athens (AUA working paper series No. 2014-5)*. Retrieved from. http://aoatools.aua.gr/RePEc/aua/wpaper/files/2014-5_Drivas.pdf

Drivas, K., Economidou, C., & Tsionas, E. G. (2014). *A Poisson Stochastic frontier model with finite mixture structure (MPRA paper No. 57485)*.. Retrieved from. http://mpra.ub.uni-muenchen.de/57485/

Dundar, H., & Lewis, D. R. (1998). Determinants of research productivity in higher education. *Research in Higher Education*, *39*(6), 607–631.

Ehlers, R. S. (2011). Comparison of Bayesian models for production efficiency. *Journal of Applied Statistics*, *38*(11), 2433–2443.

Fé, E., & Hofler, R. (2013). Count data stochastic frontier models, with an application to the patents–R&D relationship. *Journal of Productivity Analysis*, *39*(3), 271–284.

Faire, R., Grosskopf, S., & Margaritis, D. (2008). Efficiency and productivity: Malmquist and more. In H. O. Fried, C. A. K. Lovell, & C. H. Schmid (Eds.), *The measurement of productive efficiency and productive growth* (pp. 522–621). Oxford: Oxford University Press.

Falzagic, A. (2005). Measuring the intellectual capital of a university. In *Paper presented at the trends in the management of human resources in higher education* https://www.oecd.org/edu/imhe/35322785.pdf

Gingras, Y., Lariviere, V., Macaluso, B., & Robitaille, J. P. (2008). The effects of aging on researchers' publication and citation patterns. *Plos One*, *3*(12), e4048.

Glänzel, W. (2010). On reliability and robustness of scientometrics indicators based on stochastic models: An evidence-based opinion paper. *Journal of Informetrics*, *4*(3), 313–319.

Greene, W. H. (2005). Fixed and random effects in stochastic frontier models. *Journal of Productivity Analysis*, *23*(1), 7–32.

Greene, W. H. (2008). The econometric approach to efficiency analysis. In H. O. Fried, C. A. K. Lovell, & S. S. Schmidt (Eds.), *The measurement of productive efficiency and productivity growth* (pp. 92–250). Oxford: Oxford University Press.
Griffin, J. E., & Steel, M. F. J. (2007). Bayesian stochastic frontier analysis using WinBUGS. *Journal of Production Analysis*, *27*, 163–176.
Hicks, D. (2012). Performance-based university research funding systems. *Research Policy*, *41*, 251–261.
Hilbe, J. M. (2012). *Negative binomial regression*. Cambridge: Cambridge University Press.
Hilbe, J. M. (2014). *Modeling count data*. New York: Cambridge University Press.
Hoff, P. D. (2009). *A first course in Bayesian statistical methods*. New York: Springer.
Horta, H., & Lacy, T. A. (2011). How does size matter for science? Exploring the effects of research unit size on academics' performance productivity and information exchange behavior. *Science and Public Policy*, *38*(6), 449–460.
Hox, J. (2010). *Multilevel analysis – Techniques and applications*. New York: Routledge.
Johnes, J. (2004). Efficiency measurement. In G. Johnes, & J. Johnes (Eds.), *International handbook on the economics of education* (pp. 613–742). Cheltenham, UK: Edward Elgar Publications.
Johnes, J. (2006). Data envelopment analysis and its application to the measurement of efficiency in higher education. *Economics of Education Review*, *25*, 273–288.
Johnson, A. L., & Kuosmanen, T. (2012). One-stage and two-stage DEA estimation of the effects of contextual variables. *European Journal of Operational Research*, *220*, 559–570.
Kaplan, D. (2014). *Bayesian statistics for the social sciences*. London: The Guilford Press.
Kumbhakar, S. C., Wang, H.-J., & Horncastle, A. P. (2015). *A practioner's guide to stochastic frontier analysis using STATA*. New York: Cambridge University Press.
Lariviere, V., Ni, C., Gingras, Y., Cronin, B., & Sugimoto, C. R. (2013). Bibliometrics: Global gender disparities in science. *Nature*, *504*(7479), 211–2013.
Limpert, E., Stahel, W. A., & Abbt, M. (2001). Log-normal distributions across the sciences: Keys and clues. *BioScience*, *51*(5), 341–352.
Lindsey, J. K. (1995). *Modelling frequency and count data*. New York: Oxford University Press.
Liu, J. S., Lu, L. Y. Y., Lu, W.-M., & Lin, B. J. Y. (2013a). Data envelopment analysis 1978–2010: A citation-based literature survey. *Omega*, *41*, 3–15.
Liu, J. S., Lu, L. Y. Y., Lu, W.-M., & Lin, B. J. Y. (2013b). A survey of DEA applications. *Omega*, *41*, 893–902.
Lunn, D., Jackson, C., Best, N., Thomas, A., & Spiegelhalter, D. (2013). *The BUGS book – A practical introduction to Bayesian analysis*. New York: Chapman & Hall.
Mutz, R., Bornmann, L., & Daniel, H.-D. (2012). Does gender matter in grant peer review? An empirical investigation using the example of the Austrian Science Fund. *Zeitschrift für Psychologie*, *220*(2), 121–129.
Mutz, R., Bornmann, L., & Daniel, H.-D. (2013). Types of research output profiles: A multilevel latent class analysis of the Austrian Science Fund's final project report data. *Research Evaluation*, *22*(2), 118–133.
Mutz, R., Bornmann, L., & Daniel, H. (2016). Funding decision-making systems: An empirical comparison of continuous and dichotomous approaches based on psychometric theory. *Research Evaluation*, *25*(4), 416–426.
O'Hagan, A. (2008). The Bayesian approach to statistics. In T. Rudas (Ed.), *Handbook of probability: Theory and applications* (pp. 85–100). London: Sage.
O'Hara, R. B., & Kotze, D. J. (2010). Do not log-transform count data. *Methods in Ecology and Evolution*, *1*, 118–122.
Olivares, M., & Wetzel, H. (2014). Competing in the higher education market: Empirical evidence for economics of scale and scope in German higher education institutions. *CESifo Economic Studies*, *60*(4), 653–680.
Over, R. (1988). Does scholarly impact decline with age? *Scientometrics*, *13*(5–6), 215–223.
Parmeter, C. F. (2014). Efficiency analysis: A primer on recent advances. *Foundations and Trends in Econometrics*, *7*(3–4), 191–385.
Rabovsky, T. (2014a). Support for performance-based funding: The role of political ideology, performance, and dysfunctional information environments. *Public Administration Review*, *74*(6), 761–774.
Rabovsky, T. (2014b). Using data to manage for performance at public universities. *Public Administration Review*, *74*(2), 260–272.
Ramalho, E. A., Ramalho, J. J. S., & Henriques, P. D. (2010). Fractional regression models for second stage DEA efficiency analyses. *Journal of Production Analysis*, *34*, 239–255.
SAS Institute Inc. (2014). *SAS/STAT® 13.2 user's guide*. Cary, NC: SAS Institute Inc.
Simar, L., & Wilson, P. W. (2011). Two-stage DEA: Caveat emptor. *Journal of Production Analysis*, *36*, 205–218.
Sinuanystern, Z., Mehrez, A., & Barboy, A. (1994). Academic departments efficiency via DEA. *Computers & Operations Research*, *21*, 543–556.
Stringer, M. J., Sales-Pardo, M., & Amaral, L. A. N. (2010). Statistical validation of a global model for the distribution of the ultimate number of citations accrued by papers published in a scientific journal. *Journal of the American Society for Information Science and Technology*, *61*(7), 1377–1385.
Stroebe, W. (2010). The graying of academia. *American Psychologist*, *65*(7), 660–673.
Tsionas, E. G. (2002). Stochastic frontier model with random coefficients. *Journal of Applied Econometrics*, *17*, 127–147.
Waltman, L., van Eck, N. J., van Leeuwen, T. N., Visser, M. S., & van Raan, A. F. J. (2011). Towards a new crown indicator: An empirical analysis. *Scientometrics*, *87*(3), 467–481.
Wang, H.-J., & Schmidt, P. (2002). One-step and two-step estimation of the effects of exogenous variables on technical efficiency levels. *Journal of Productivity Analysis*, *18*, 129–144.
Wijk van, E., & Costas-Comesana, R. (2012). *Bibliometric Study of FWF Austrian Science Fund 2001-2010/2011*. Retrieved from. http://www.fwf.ac.at/fileadmin/files/Dokumente/Ueber_den_FWF/Publikationen/FWF-Selbstevaluation/Bibliometric-Study-FWF-2001-2010-Originalstudie.pdf
van den Broek, J., Koop, G., Osiewalski, J., & Steel, M. F. J. (1994). Stochastic frontier models. *Journal of Econometrics*, *61*, 273–303.