

Economies of scope in research and teaching: a non-parametric investigation

Kristof De Witte
Nicky Rogge
Laurens Cherchye
Tom Van Puyenbroeck

HUB RESEARCH PAPERS 2012/14
ECONOMICS & MANAGEMENT
FEBRUARI 2012



Economies of scope in research and teaching: a non-parametric investigation

Kristof De Witte^{‡δ1}, Nicky Rogge^{‡*}, Laurens Cherchye^{‡†} and Tom Van Puyenbroeck^{‡*}

([‡]): Katholieke Universiteit Leuven (KULeuven)
Faculty of Business and Economics
Naamsestraat 69, 3000 Leuven (Belgium);

(^δ): Maastricht University
Top Institute for Evidence Based Education Research
Kapoenstraat 2, MD 6200 Maastricht (the Netherlands)
Kristof.dewitte@econ.kuleuven.be

([†]): Tilburg University
Department of Econometrics and OR
P.O. Box 90153, 5000 LE Tilburg (The Netherlands);

(^{*}): Hogeschool-Universiteit Brussel (HUBrussel)
Centrum voor Duurzaam Ondernemen (CEDON)
Center for Educational Research & Development (ER&D)
Stormstraat 2, 1000 Brussels (Belgium)

Abstract

We propose a non-parametric methodology to study the presence of economies of scope between teaching and research (i.e., the teaching-research nexus). In particular, the paper advocated a conditional version of the ‘benefit-of-the-doubt’ approach to estimate the relationship between the professors’ overall academic output, measured by a composite measure of multi-dimensional and importance-adjusted scores of teaching effectiveness and research productivity, and the time devoted to teaching and to research. The methodology is illustrated with a dataset of professors working at a Business & Administration department of a university college where the time allocation of teaching and research was assigned exogenously. The outcome of the analysis indicates the presence of limited scope economies for professors with an extensive research time.

Keywords: Teaching-research nexus, Data envelopment analysis, Conditional efficiency, Economies of scope, Higher education.

JEL-classification: C14, C25, I21

¹ *Corresponding author.* Tel.: 003216326566

E-mail address: Kristof.Dewitte@econ.kuleuven.be

1. Introduction

The debate on the precise relationship between teaching and research - in particular whether teaching tasks hinder good research or whether research benefits the teaching effectiveness - is a long-standing and highly controversial one. A key reason for this controversy seems to be the involvement of multiple university stakeholders (such as students, professors themselves, university management, and policy makers) with sometimes diverging interests. Broadly speaking, there are two main streams in the debate: the believers in a positive teaching-research nexus and the sceptics. The believers argue that teaching and research go together in an essential and symbiotic way (Neumann, 1992; Kremer, 1990; and Westergard, 1991). Therefore, they believe that it is unwise to separate teaching and research (i.e., assigning activities among professors and/or institutions such that there are only full-time teachers/teaching institutions and full-time researchers/research institutions). The sceptics strongly contest this alleged positive teaching-research nexus (e.g., Hattie and Marsh, 1996; Feldman, 1987, Marsh and Hattie, 2002). In their opinion, this positive nexus is just a notion invoked by certain stakeholders who benefit from the presence of this perceived link. Most of the skeptics even question whether there exists any relationship at all between teaching and research. Hattie and Marsh, for example, posited that *“the widely held belief that teaching and research are inextricably entwined is an enduring myth. At best, research and teaching are very loosely coupled”* (Hattie and Marsh, 1996 p. 529). Other non-believers (e.g., Barnett, 1992; Massy and Zemsky, 1994; etc.) see the combination of teaching and research as a potential source of conflict with professors being forced to focus on one activity while partially neglecting the other.

The previous studies on the teaching-research nexus can be largely classified into two groups according to whether they used qualitative or quantitative approaches to examine the relationship between both core activities of professors. Qualitative studies frequently use semi-structured, in-depth interviews to collect information about the perceptions, attitudes, or opinions of the different university stakeholders on the teaching-research nexus. The quantitative studies (Fox, 1992; Friedrich and Michalak, 1983; etc.) typically compute the linear correlation coefficient between the outputs of the teaching and research activities, respectively research productivity and teaching effectiveness. In general, both types of studies produced different results. Qualitative studies usually report a strong belief among university stakeholders that teaching and research are positively related. Specifically, most respondents indicate that this positive relationship predominantly works in one way, with the impact of research on teaching being far more important than the other way around. Conversely, most quantitative studies found that teaching and research (i.e., the output of both activities) are at maximum marginally correlated. So, findings from quantitative and qualitative studies seem contradictory. This should perhaps not be a surprise considering (1) the intrinsically complex nature of both the teaching and research activity as well as of the nexus between both activities and (2) the large differences between the qualitative and quantitative approaches used to capture this complexity.

Some recent studies, however, pointed to a number of important limitations of previous quantitative studies. A first important shortcoming is related to the use of simple correlation analysis in the majority of these studies. Particularly, whereas the lack of a priori precise knowledge on the true relationship between teaching and research calls for a methodology that is sufficiently flexible and does not make too strong assumptions, the correlation coefficient analysis implicitly assumes a linear relationship between both activities. Among others, Hattie and Marsh (1996), Marsh and Hattie (2002), Stack (2003), and Locke (2004), argued that the attempt to establish linear relationships between the outputs of both teaching and research activities may be flawed.

A second limitation relates to the proxies used to capture research productivity and teaching effectiveness. Typically, quantitative studies gauge the output of professors in teaching and research by either using unidimensional measures (e.g., students ratings on one global questionnaire item proxy the professor's teaching effectiveness, and publication counts proxy the professor's research output) or either simple summations of the professors' performances on underlying research and teaching criteria (so multidimensional outcome measures without any correction for differences in the importance of the underlying teaching and research criteria). Both an extensive academic literature (e.g., Feldman, 1996, 1997; Massy and Wilger, 1995; De Witte and Rogge, 2010, 2011; and Marsh, 2007) and practical evidence suggest that measures of teaching effectiveness and research productivity of professors should be multidimensional (comprise the multiple aspects of respectively the teaching and research activity) and should be adjusted for value/importance differences of the underlying performance criteria.

A third important limitation of most quantitative studies is that they do not correct the estimations of the teaching-research association for the impact of factors that are often not (or only limitedly) controllable by the professors but may nevertheless influence their opportunities to teach effectively and do good research.

This paper contributes to the literature by examining scope economies using a quantitative approach that accommodates for the shortcomings just indicated. With respect to the first limitation, in contrast to most previous quantitative studies that employed parametric approaches such as computing simple linear correlation coefficients to examine the teaching-research relationship, this paper advocates using a non-parametric analysis approach. Particularly, we propose the 'benefit-of-the-doubt' approach (BoD, after Melyn and Moesen, 1991) to examine whether teaching, research or their combination foster academic performance. The BoD-methodology is related to the popular Data Envelopment Analysis-methodology (DEA, hereafter), a non-parametric technique originally developed by Farrell (1957) and put into practice by Charnes *et al.* (1978), to evaluate the efficiency performance of observations (e.g., companies, organizations, individuals, etc.) in complex settings. The belief is that the BoD-approach, thanks to its non-parametric nature, is well-suited to capture the complexity in the

teaching-research nexus. The version of the BoD-model that is used in this paper is based on new insights of Cazals *et al.* (2002), Daraio and Simar (2005, 2007a, 2007b), and De Witte and Kortelainen (2009). To our best knowledge, this is the first paper to explore the relationship between teaching and research in a non-parametric fashion.

Regarding the second limitation, contrary to most previous quantitative studies, the output variables used in this paper to measure teaching effectiveness and research productivity consist of composite performance measures that comprise the multiple underlying teaching and research output criteria and account for differences in the importance of these underlying criteria. In particular, we construct importance-adjusted composite performance measures of teaching effectiveness and research productivity by using stakeholder opinions the selection and the weighting of the appropriate criteria of teaching and research.

The usefulness of the BoD-methodology to examine the teaching-research nexus is illustrated on a sample of academics working at a Business & Administration faculty of a large university college. This application is attractive for two reasons. First, the sample consists of a homogenous group (only Business and Administration faculty) such that biases due to differences across departments can be avoided. Second, the university college can be compared to the ‘new’ polytechnic universities in the UK and the colleges in the US where a teaching university is only recently research oriented. In this transformation, the research and teaching time has been allocated exogenously and did not depend on the professor’s actual and past performance. This makes an attractive setting for studying the causal relationship between teaching and research.

The remainder of this paper unfolds as follows. The next section provides a brief literature review of previous quantitative studies. In a third section, we discuss at length the aforementioned limitations of most previous quantitative studies. Section 4 discusses the data for the professors working at the faculty of Business and Administration of the university college under study. Section 5 presents the methodology to estimate the existence of scope economies. Section 6 presents the findings for our data set. In a final section, we make some concluding remarks and provide some directions for further research.

2. Teaching-research nexus: A review of the literature

The relationship between teaching and research attracted significant attention during the 1970s and the 1980s. Recently, stimulated by the increased interest of both practitioners and policy makers, interest in the subject has been renewed. In this section, we present the most important findings of quantitative studies.

Five literature reviews gather the quantitative studies, up to the mid 90s (i.e., Faia, 1976; Feldman, 1987; Hattie and Marsh, 1996; Allen, 1996; and Braxton, 1996), and three of these are

meta-analyses. A first review of the empirical literature on the teaching–research nexus was performed by Faia in 1976. This review covered 11 studies that appeared between 1952 and 1975 and found indications of two possible relationships between teaching and research: a zero relationship (4 studies) and a positive, yet very weak association (7 studies).

Feldman (1987) included 29 studies in his meta-analysis (14 other studies were excluded for various reasons, such as lack of information to retrieve the exact direction of the reported association and no sufficient data to adequately compute the relationship between teaching and research). Feldman reported correlations between teaching effectiveness and research productivity varying from a low of -0.31 to a high of + 0.39. More in particular, he found a majority of the studies (i.e., 18 of the 29 studies) reporting average correlations that are not significantly different from zero, ten studies with average correlations that are significantly positive, and only one study reporting a significantly negative average correlation. After aggregating the correlations of the 29 studies (thereby combining the significance tests of the several studies into an overall pooled test statistic), Feldman obtained an average correlation of + 0.12, hence, a positive, yet weak teaching-research nexus.²

Both Hattie and Marsh (1996) and Allen (1996) replicated Feldman's (1987) meta-analysis for a larger sample of studies. More precisely, Hattie and Marsh (1996) increased the number of studies from 29 to 58. Allen (1996) repeated the analysis for a sample of 46 studies.³ For Hattie and Marsh (1996), this resulted in a sample of 498 teaching-research correlations (collected over the 58 studies), with an overall correlation of +0.06 and a 95% confidence interval between +0.061 and +0.066. After correcting for overemphasis on a few studies with a large number of computed correlations, average correlations rose to around +0.11. Allen (1996) reported a similar result with an average correlation of +0.107.⁴

Braxton (1996) reviewed 30 studies, i.e. the 29 studies also considered in the meta-analysis of Feldman (1987) and the study of Voeks (1962), using a vote-counting methodology.⁵ Braxton categorized the studies in three different groups: students reporting average correlations

² Feldman (1987, p. 240) also found that average correlations were quite consistent for the various types of indicators used to measure research productivity.

³ Both Hattie and Marsh (1996), and Allen (1996) included studies in their meta-analyses that were previously excluded by Feldman (1987) due to data limitations or inaccuracies.

⁴ Note that the average correlations of Hattie and Marsh (1996), and Allen (1996) are very close to the one reported by Feldman (1987). This should not be a surprise taking into account that there is considerable overlap among the studies reviewed in the three meta-analyses (with both Hattie and Marsh (1996) and Allen (1996) covering the 29 studies reviewed by Feldman (1987)).

⁵ Braxton (1996) defended the use of the vote-counting methodology over the meta-analytic approaches. In his opinion, the method of averaging the correlations found in the several studies might mask important variation in the findings (thereby preventing a simultaneous appraisal of the different perspectives on the teaching-research association). However, others, like Feldman (1987) for instance, argued that such a 'vote counting' approach is not advisable due to its very low analytic power.

respectively higher than +0.10 (11 studies), between +0.09 and -0.09 (19 studies), and -0.10 (1 study). A simple count of the studies in the different groups favours the hypothesis of teaching and research not being intertwined.

3. Issues in quantitative estimations

A first shortcoming of most previous quantitative studies was discussed by, among others, Marsh and Hattie (2002), Locke (2004), and Stack (2003), and is related to the use of a simple correlation analysis in the majority of the quantitative studies. The correlation coefficient implicitly assumes linearity in the relation between the variables of interest (i.e., teaching effectiveness and research productivity). As, e.g., Zaman (2004) has noted, there are no objective reasons to assume that such a linear relationship actually holds between the teaching and research measures. The link between measures of research productivity and teaching effectiveness may be non-linear. For instance, there is the empirical finding that the distribution of research outputs is typically heavily and negatively skewed among professors, with a minority of the professors producing the majority of a department's research output.⁶ Both Stack (2003, p.540) and Hattie and Marsh (1996, p. 532) indicated the skewed distribution of research output as one of the reasons why linear correlations between teaching and research outputs are small or insignificant and perhaps artificially suppressed. Stack (2003), e.g., did find a significant positive correlation between research productivity and teaching effectiveness, but only after correcting the distribution of research productivity for skewness. In other words, even when linear correlations between teaching and research appear to be very low or even zero, this need not imply that there is no relationship whatsoever between both activities. Clearly, this demonstrates the value of methodologies which allow for flexibility in the exploration of teaching-research associations. However, to our knowledge, up to now, only a few studies used a flexible methodology to explore the teaching-research nexus (notable exceptions are, for instance, Stack, 2003; and Marsh and Hattie, 2002).

A second important issue is related to the applied measures of research productivity and teaching effectiveness. Previous literature mainly used aggregated measures that are a summation (or an arithmetic average) of the professors' performances on respectively the underlying teaching and research criteria. More precisely, studies measured teaching effectiveness as a simple arithmetic mean of student ratings on multiple questionnaire items (measuring several aspects of teaching) and research productivity as a simple count of the number of realized publications.⁷ In essence, this implies that in the computation of both composite measures all underlying criteria are

⁶ In the literature, one also refers to this finding as Lotka's Law (after Lotka, 1926)

⁷ In their meta-analysis, Hattie and Marsh (1996) found that approximately 80% of the studies they surveyed looked at research productivity by simply counting the number of realized publications.

assumed to be of equal importance. Whether such equal weights are appropriate is questionable. Indeed, there are quite some indications suggesting that measures of research productivity and teaching effectiveness should not only be multi-dimensional but also importance-adjusted. For instance, Feldman (1996, 1997) and Marsh (2007) provide empirical findings of differences in importance of underlying criteria for measures of teaching effectiveness, while Massy and Wilger (1995) report a similar phenomenon for measures of research productivity. Practical experience from professors themselves additionally indicates that some underlying performance criteria are relatively more important in the evaluation of a professors' teaching effectiveness or research productivity. For instance, most professors will agree that publications in international peer-reviewed journals should weigh more heavily compared to articles published in national journals in the composition of a measure of research productivity. A similar remark applies to the evaluation of teaching effectiveness. Educational experts (e.g., Pritchard et al., 1998) and teachers typically believe that some teaching aspects are more crucial to the development of excellent teaching practice than others. Consequently, several authors (Feldman, 1996, 1997; Marsh and Hattie, 2002; and Stack, 2003) have been critical to measures of research productivity and teaching effectiveness that do not correct for differences in importance of the underlying teaching and research criteria.

A third limitation of most of the previous quantitative studies is that they do not account for the impact of exogenous factors in the estimation of the relationship between teaching and research. This is somewhat surprisingly given the substantial evidence and practical experience of professors suggesting that several factors though being non-controllable to the professors can significantly influence research productivity (e.g., Sax et al., 2002; Hattie and Marsh, 1996; Chen et al., 2006; etc.) and teaching effectiveness (see, among others, Cashin, 1995; Centra and Gaubatz, 2000; Feldman, 1996; Marsh, 2007) and, thereby, the nexus between both key roles of professors. Examples of such background factors include time spent on teaching and time devoted to research, at least to the extent that it is the university management (e.g., faculty board, research council) that determines for each professor the teaching and research duties. Therefore, as the literature indicates that both teaching time and research time may influence the professors' opportunities to teach effectively or to be productive in research (Fox, 1992; Fox and Milbourne, 1999; Maske et al., 2003; Chen et al., 2006; etc.), it is important to account for the factor time in the evaluations of the professor's output in both activities and, thus, in the estimations of the teaching-research nexus.

To account for the limitations of most previous quantitative studies, Section 5 proposes a nonparametric and non-linear methodology which allows (1) examining the teaching-research nexus without requiring strong assumptions on how the outcomes on both activities interrelate, and (2) correcting for differences among professors in background factors (such as the teaching duties and the available time to do research). First, the next section describes the case study used in this paper to illustrate the methodology.

4. Data

The analysis is applied at a large university college in Brussels. This university college resembles in many ways to the ‘new’ (polytechnic) universities in the UK and the colleges in the US. In particular, it used to be an educational institution with a predominant focus on teaching, whereas recently, following the Bologna reforms, the university is increasingly research oriented.

The sample of observations consists of 155 professors working at the department of Business & Administration of this university college. In this sample, three types of professors can be distinguished. There is a group of professors who only focus on teaching (‘full-time teachers’), there is also a group of professors who exclusively focus on research (‘full-time researchers’), and there is a group of professors who combine both activities (‘the researcher-teachers’). Further, the homogeneity of the set of observations in our analysis (i.e., only professors from the same department) guarantees that results will be less biased (for instance, by departmental characteristics, factors related to the educational program, etc.). The data were collected for the period 2006-2008 from three different sources: departmental research evaluations, questionnaire data administered by students to evaluate teaching effectiveness (i.e., SETs), and administrative records.

The official research evaluations performed at the university college provide a measure of research productivity for each professor in the sample. Contrary to the typical, single-criterion measures of research performance often used in the academic literature and practice (such as publication counts, citation counts, etc.), the university college under study preferred to use a multi-criteria Research Evaluation Score (RES-score) that accounts for importance differences over research output criteria. In particular, the department’s Research Council, in dialogue with the researchers, decided on the selection of research output criteria and their associated importance so as to reflect the research policy priorities of the faculty. This resulted in a composite RES-score comprising professor performances in 9 research output criteria, each with a corresponding importance weight (see Appendix 1). For example, a publication in an international peer-reviewed journal was valued more than a publication in a national peer-reviewed journal (15 points vs. 7 points). For a more detailed description of the composite, importance-adjusted RES-score, we refer to De Witte and Rogge (2010).

We measure teaching effectiveness by computing composite SET-scores for lectured courses using evaluation questionnaires filled in (anonymously) by students attending the courses. The questionnaire contained 16 statements (all measured on a five-point Likert scale that corresponds to a coding rule ranging from 1 (I completely disagree) to 5 (I completely agree)) to evaluate the multiple aspects of the professor’s teaching performance (the 16 questionnaire items are listed in Appendix 2). Contrary to traditional practice, SET-scores were not calculated as arithmetic averages of the class average student item responses (which would imply equal importance of all underlying teaching criteria). Instead, importance weights were assigned to the underlying

teaching criteria according to their importance to effective teaching (after consulting the two parties most involved in the teaching process: the students and the teachers). In total 530 of such SET-scores were calculated for college courses taught at the department of Business & Administration. Professors who taught several courses will have several teaching effectiveness scores (SET-scores), i.e. one for each evaluated course. To obtain a global measure of teaching effectiveness per professor, the SET-scores of professors were further weighted by the contact hours.⁸ We refer to Rogge (2011) and De Witte and Rogge (2011), for a more detailed description of the questionnaire data and the differential weights used to construct the SET-scores.

The official employee records contain information about each professor's teaching load and available time for doing research. In the literature, both variables have been shown to be potentially associated with teaching and research performance (e.g., Sax et al., 2002; Marsh and Hattie, 2002; etc.).⁹ The teaching load is proxied by the official percentage of time that a professor should invest in teaching. The available time for doing research is proxied by the official percentage of time that the professor is ought to spend at doing research.

It is important to point out that research time and teaching time allocated to an individual professor do not depend on the professor's actual research and teaching performance, i.e., the time allocation can be considered as exogenous. To see this, we note that the period we consider here corresponds with the transition period in which the university college transformed from an almost exclusively teaching-oriented institution into an institution with both a teaching and research orientation. During this transition, both the board and the research council of the department decided on the teaching and research time for the professors without considering the professors' past teaching and research performances in detail. Professors who were already active in doing research in economics or management-related research areas received automatically 50% research time. Junior professors received a 60%-40% division of research-teaching percentages. Professors who were already active in doing research but in other than economics-related research areas, were granted 20% research time.¹⁰ Next to the professors with both research and teaching interests, there are also professors that are full-time teacher (i.e., no research time) or

⁸ The rationale for using contact hours is that, in general, courses with a higher contact hour ask more preparation, organization, etc. Therefore, they should weigh more heavily in an academic's teaching performance evaluation.

⁹ This set of background variables is not exhaustive. The literature indicated also other background variables (such as research collaboration (e.g., Maske et al., 2003; and Ramos et al., 2007), family situation (e.g., Bellas & Toutkoushian, 1999; Hunter & Leahey, 2010), etc.) that have the potential to influence the teaching and research performance of academics. However, due to data limitations we were unable to account for these variables in our analysis.

¹⁰ Note that the percentages can be adjusted according to the total contract time. As an example, for a professor who is only working half-time at the department (and with a new interest in research), instead of having research and teaching percentages of 20% and 80%, percentages will be set equal to 10% and 40%.

full-time researchers (i.e., no teaching duties). Here, full-time teachers are typically professors with a long track record at the institution and no interest in doing research, while full-time researchers are young academics with an exclusive focus on doing research (i.e., they are not responsible for providing lectures).¹¹

In sum, for each professor in the sample, the data sources provide a measure of his/her output in both key activities, i.e., a measure of teaching effectiveness and a measure of the research productivity, as well as of the available time for both activities. Summary statistics are presented in Table 1. Because we have full-time teaching and full-time research professors, the minimum- and maximum values for the teaching and research time percentages equal 0 and 1. Correspondingly, the SET-scores (respectively RES-scores) equal 0 for the full-time researchers (respectively full-time teachers). Further, we find that the minimum-, first quartile-, and median value of the RES-scores equal 0. This reflects the negatively skewed distribution of the research output among the professors (Lotka's Law, see above). One also observes that 45% of the professors in the sample set are female and that the median age of the professors is 43 years.

< Table 1 about here >

< Figure 1 about here >

5. Are there any scope economies between teaching and research?

5.1. A 'benefit-of-the-doubt' (BoD)-model

Before we can estimate whether or not economies of scope exist between teaching and research, we need to aggregate the outputs realized by each professor in both activities into an overall performance score. To do so, we rely on the nonparametric 'benefit-of-the-doubt'-model (BoD, after Melyn and Moesen, 1991) that is rooted in the Data Envelopment Analysis (DEA; Charnes et al., 1978) literature. Essentially, the BoD-model differs from the standard DEA-model as it exclusively focuses on aggregating outputs y ($y \in \mathfrak{R}^q$, where q is the number of outputs).¹² The application at hand consists of two outputs (hence, $q = 2$), i.e., the multidimensional and importance-adjusted SET-scores ($i = 1$) and RES-scores ($i = 2$) of the professors (with SET- and RES-scores discussed above).

¹¹ Although all professors are included in the analysis below, robustness test with subgroups (e.g., excluding full time teachers and researchers) point to similar results.

¹² We refer to Melyn and Moesen (1991) and Cherchye et al. (2007) for a detailed discussion of this point.

To aggregate the different outputs, the BoD-model determines the weights endogenously. Particularly, the basic idea of the BoD-model is to put, for each output i ($i=1,\dots,q$), the performance of a professor k ($y_{k,i}$) in a relative perspective to the performances of the other professors j ($y_{j,i}$) $j=1,\dots,k,\dots,n$ in the set of observations Υ , and look for the outputs of relative strength and of relative weakness. A good relative performance on an activity of the evaluated professor (i.e., a relatively high SET-score and/or RES-score) then indicates that this professor is either relatively good in this activity (a relative strength) or assigns more time to it. Given this, the output realized in that activity should weigh more heavily in the professor's overall performance evaluation which means that a high endogenous weight is assigned to it. The opposite reasoning holds for the activities on which the professor realizes an output that is relatively low compared to the colleagues (i.e., a relatively low SET-score or RES-score). The BoD-model assigns a low endogenous weight to these output criteria.

The notion of 'benefit-of-the-doubt' thus enters into the specification of the output weights. The BoD-model looks for each evaluated professor k for the set of weights $w_{k,i}$ that maximize the impact of outputs of relative strength and minimize the influence of outputs of relative weakness. If the evaluated professor is performing better in teaching than in research compared to the colleagues, the BoD-model assigns a high relative weight to the output realized in the teaching activity (SET-score) and a low relative weight to the research output (RES-score). On the other hand, if the evaluated professor realizes the highest relative output in the research activity, the BoD-model assigns a high relative weight to the RES-score and a low relative weight to the SET-score. In other words, for each professor the BoD-model looks for the highest overall performance score possible by adapting the most favourable weights. If this overall performance score $E_k y$ falls below the maximum of 1, this 'benefit-of-the-doubt' model assures that no other weights for the two outputs result in a higher performance measure. The ensuing model can be formally translated in linear programming terms, as follows:¹³

$$\begin{aligned}
 E_k y &= \max_{w_{k,i}} \sum_{i=1}^q w_{k,i} y_{k,i} && 1 \\
 \text{s.t.} &&& \\
 \sum_{i=1}^q w_{k,i} y_{j,i} &\leq 1 && j = 1, \dots, k, \dots, n && 1a \\
 w_{k,i} &\geq 0 && i = 1, \dots, q && 1b
 \end{aligned}$$

¹³ See Cherchye et al. (2007) for details. All program codes (written in R) are available from the authors upon request. The codes can be easily customized to the specificities of other evaluation settings.

From restriction 1a, we can deduce that, for all evaluated professors, the overall performance score $E_k y$ will lie between 0 and 1 with higher values indicating a better relative overall performance. If $E_k y < 1$, this indicates that the evaluated professor k could perform better. Indeed, there are other professors in the sample Υ who realize higher overall performance scores even when applying the evaluated professor's most favourable weights $w_{k,i}$ (i.e., weights which are probably less favourable than their own optimal weights). Only if $E_k y = 1$, the evaluated professor obtains the highest overall performance score relative to the other professors. In this case, he/she is not outperformed by other observations $j \quad j = 1, \dots, c, \dots, n$ when applying his/her best possible weights $w_{k,i}$.

5.2. Measuring scope economies by a BoD-model

The model above allows for estimating the overall performance of professors (based on the outputs realized in both the teaching and research activities), but not the influence of scope economies. Using insights of Daraio and Simar (2007a,b) and De Witte and Kortelainen (2009), we tailor the BoD-model such that it allows us to estimate the presence of scope economies. We proceed in two steps.

5.2.1. A robust BoD-model

We have presented the BoD-model as a deterministic model in our above discussion (see 1 – 1b). This has some drawbacks. First, the model does not allow us to define statistical properties of the overall performance measures $E_k y$. For example, we cannot compute confidence intervals for these measures. Second, professors with outlying performance data or performance data affected by measurement error (if present in the sample set Υ) can heavily influence the results of all observations in the sample set. Outlying observations can arise from atypically high output attainments (for instance, professors attaining very high SET-scores and/or realizing a very high research output). While some may argue that outlying observations are highly interesting, and require additional research and inspection, we choose to mitigate their impact in our application; this also falls in line with usual practice in parametric applications.

To account for these issues, we extend the above BoD-model by using the insights of Cazals et al. (2002) for obtaining robust, so-called 'order- m ' overall performance measures.¹⁴ In what follows,

¹⁴ Cazals et al. (2002) initially proposed the insights behind the robustification of the performance scores in the setting of the DEA-models. However, given the formal similarity between the BoD-model of Cherchye

we restrict ourselves to briefly sketching the basic idea of the robust measurement procedure. We refer to Cazals et al. (2002) for a more detailed explanation of the method; this also includes a discussion of attractive statistical properties (which carry over to our setting).

The robust measurement procedure does not consider the full set of observations Υ for defining the overall performance score E_k . Instead, it repeatedly considers subsamples Υ_k^m of only m observations for defining robust performance scores E_k^m . In each iteration, the method proceeds in two steps. In a first step, m observations are drawn (with replacement) from the set Υ (denote the corresponding reference set by Υ_k^m). In a second step, the method computes E_k^m by means of the linear programming problem in model 1 – 1b after replacing Υ by Υ_k^m .

Robust measures are obtained by repeating these two steps B times (with B a large number, *in casu* 500). For each observation k , the robust order- m performance measure is then computed as the average value of E_k^m defined over the B iterations. As outlying observations do not form part of the set Υ_k^m in every draw, the impact of outlying observations on these order- m performance values is effectively mitigated. Moreover, because of the re-sampling procedure, we can construct confidence intervals and standard deviations for the performance measures.

A final point of attention concerns the choice of the parameter m . As pointed out by Cazals et al. (2002) and Daraio and Simar (2006), too high and too low values of m should be avoided. In our application, we use $m = 50$. Sensitivity analysis points out that the results are robust with respect to alternative choices of value of m (i.e. we also considered $m = 20, 30, \dots, 120$). (For a more comprehensive discussion of the role of the parameter m , see both aforementioned studies.)

5.2.2. A conditional robust BoD-model

In our following empirical analysis, we will use one further extension of the robust BoD-model. Cazals et al. (2002) discussed how the robust order- m model could be modified to account for the operational environment (i.e., to account for (observable) heterogeneity). These ideas were further operationalized by Daraio and Simar (2005, 2007a, 2007b) for continuous variables and by De Witte and Kortelainen (2009) for both continuous and discrete environmental variables.

Specifically, these authors propose a methodology that obtains so-called conditional performance measures, which condition the performance measurement on exogenous factors, which we capture by the vector z . The computation of these conditional measures involves a slight modification of the robust order- m procedure outlined above. Whereas the original procedure

et al. (2007) and existing DEA-models, these insights are readily adapted to the current context (i.e., the BoD-model).

draws observations from Υ with uniform probability, conditional measures are obtained by drawing observations with a probability defined on the basis of a kernel density function evaluated at the location of the exogenous factors for the observation k . The idea is that observations get a greater probability of being drawn if their exogenous factors (i.e., the teaching and research environment) are more similar to the ones of the evaluated observation k . We denote the estimates of the conditional BoD-model by $E_k y|z$.

In summary, conditional performance measurement accounts for the differences in exogenous factors among professors by comparing likes with likes.¹⁵ In the application at hand, this means that in the estimations of the overall performance scores for professors, the conditional BoD-model will predominantly compare professors with colleagues that are allocated similar teaching time z_1 and research time z_2 . As such, this conditional measurement procedure acknowledges that the available time to do research and teaching may influence the professors' research productivity and teaching effectiveness (and, hence, the nexus between the two core activities of professors).

5.2.3. The non-parametric regression framework

The conditional robust BoD-framework allows for an interpretation of the impact of the exogenous factors by non-parametrically regressing the ratio $E_k y|z / E_k y$ on the exogenous factors z . By doing so, we can learn (1) whether z is on average statistically significantly related to the overall performance scores $E_k y$, and (2) whether this relationship is favourable or unfavourable. Daraio and Simar (2005, 2007a,b) and Badin et al. (2010) also showed how the conditional robust BoD-framework enables one to visualize the estimated relationships between the exogenous factors z and the ratio $E_k y|z / E_k y$. When z is univariate, the visualization is clear-cut: a scatter plot with on the horizontal axis the exogenous factor z and on the vertical axis the ratio $E_k y|z / E_k y$. However, when z is multivariate, the visualization is somewhat more demanding, however, partial regression plots where only one exogenous factor is allowed to vary while all other exogenous factors are kept at a fixed value (*ceteris paribus*) provide an appealing solution.

In the interpretation of outcomes of the non-parametric regression analysis, if z is significantly and positively related to the overall performance scores of professors, the regression coefficient will show a negative sign in the non-parametric regression analysis or a decreasing regression line

¹⁵ Note that, in many ways, the conditional BoD-model corresponds to the traditional idea of 'matching' in which a control group and experimental group are constructed.

in the visualization of this estimated relationship. In that case, we can argue that there is a favourable impact of the exogenous factors z that fosters research and teaching performance (and, thus, the overall performance score $E_k y$) such that the ratio $E_k y|z / E_k y$ is decreasing for larger values of z . Conversely, an unfavourable influence of z between teaching and research would act as a limitation for output such that the ratio between $E_k y|z / E_k y$ is increasing for larger values of z .

The next section illustrates how the conditional, robust BoD-methodology in combination with the non-parametric regression analysis and visualization allows one to assess the influence of the exogenous factors ‘time to teaching’ (z_1) and ‘time to research’ (z_2). Specifically, we investigate the presence of scope economies by both a visualisation of the effects that we measure and by assessing the statistical significance of these effects.

5.3. The results: presence of scope economies?

To evaluate the presence of scope economies in the case study, the conditional robust order- m BoD-model is estimated using the data on the 155 professors in our sample set. The output variables consist of the multidimensional and importance-adjusted teaching and research performance scores (i.e., the composite SET- and RES-scores) as described in Section 4. As discussed before, the background (exogenous) factors on which we condition the performance estimations are percentage of teaching time (z_1) and percentage of research time (z_2). We will discuss each of these factors in turn.

We first consider the influence of the exogenous factor teaching time z_1 on the overall performance of professors. The results are presented in Figure 2a and the first column of Table 2. In Table 2, the coefficient and the p -value indicate that the overall impact of teaching time on academic performance is statistically significant and unfavourable (coefficient of 0.4628 with p -value of 0.0360). Figure 2a yields more detailed insights. When teaching time is either low or zero, we observe in ‘Zone 1’ of Figure 2a a favourable effect of attaching more importance to teaching; here, an increase in teaching time improves overall academic performance. Once 20% of the official amount of working time is allocated to teaching (‘zone 2’), doing more teaching is unfavourable for overall academic performance. This remains the case as long as no more than 78% of time is devoted to teaching, after which a further increase in teaching duties is favourable again (‘zone 3’). In sum, Figure 2a reveals the existence of scope economies between teaching and research for professors who did not teach before. It seems that either a little bit of teaching (up to 20% of time) or a strong specialization in teaching (at least 78% of total time) favours overall academic performance (as measured by $E_k y$). Next, for teaching time between 20%

and 78%, we find a similar effect as, among others, Feldman (1987) and Fox and Milbourne (1999): in such a situation, additional teaching time seems to imply an important constraint on improving academic performance (in particular, research productivity). A quick look at our raw data confirms this finding. In general, professors with research percentages equal to or below 20% perform rather poorly in terms of research productivity. In fact, most of these professors did not succeed in producing any research output whatsoever (i.e. their RES-score equals 0).

Let us then investigate the relationship between research time (z_2) and overall academic performance. Both ‘zone 4’ in Figure 2b and the coefficient and p -value in the second column of Table 2 suggest an overall unfavourable influence of doing more research in case a considerable proportion of time is invested in teaching (i.e., coefficient of 0.1028 with p -value of 0.000). Only if at least 58% of time is allocated to research, a further increase of research time leads to improved academic performance (see ‘zone 5’).¹⁶

Note that Figures 2a and 2b provide complementary information. Specifically, both figures reveal that an increase in research time harms overall academic performance if teaching time is substantial (compare zone 3 in Figure 2a with zone 4 in Figure 2b). Next, the two figures lead us to conclude that, if the allocated research time is considerable, more teaching duties harms academic performance (compare zone 2 in Figure 2a with zone 5 in Figure 2b). In sum, our results suggest the existence of scope economies (at least, to a limited extent) between teaching and research for full-time researchers. In these cases, overall academic performance benefits from a little bit of teaching.

< Figure 2 about here >

< Table 2 about here >

6. Summary and conclusions for policy and further research

We have examined the presence of scope economies in teaching and research by using a non-parametric methodology. We have argued that this Data Envelopment Analysis (DEA)-based methodology accommodates for the following shortcomings of previous quantitative studies: (1) the assumption of linearity in the teaching-research relationship, (2) the use of unidimensional or simplistic measures of the professor’s teaching effectiveness and research productivity, and (3) the lack of correction for the impact of mediating factors that are often not (or only limitedly) controllable to the professors. In particular, the paper advocated a conditional version of the ‘benefit-of-the-doubt’ approach to estimate the relationship between the professors’ overall

¹⁶ Note again that due to unobserved heterogeneity (e.g., ability) one cannot interpret this result as causal.

academic output, measured by a composite measure of multi-dimensional and importance-adjusted scores of teaching effectiveness and research productivity, and the time devoted to teaching and to research (i.e., the teaching and research percentages). We used a tailored BoD-model that builds on recent insights of Cazals *et al.* (2002), Daraio and Simar (2005, 2007a, 2007b) and De Witte and Kortelainen (2009).

Our results indicate that professors who are full-time researchers (and, hence, currently have no teaching duties) could improve their overall academic performance by investing a small proportion of their time in teaching. However, this favourable effect of increasing teaching time is fairly modest (and only holds if teaching time is less than 20% of total time). In fact, once teaching time exceeds 20%, further increases in teaching duties seem to harm the overall academic performance. On the other hand, we observe that specialization in teaching and research correlates with better academic performance.

We believe these findings provide useful input to the university management. For instance, it could be interesting to give permission to full-time researchers to teach a course subject that is related to the own research interests or, somewhat similarly, to allow certain members of the teaching staff to set aside a number of hours for “guest lectures” by full-time researchers (e.g., to present their research topic). The general finding that specialization in teaching and research correlates with better academic performance seems to suggest that it is in the interest of academic performance to avoid professors with mixed activities. However, we believe that this finding should be interpreted with caution. A detailed study of our data shows mixed outcomes, with some professors succeeding in balancing a considerable research task with moderate teaching loads (realizing good SET-scores and a high research output) and others being less successful in combining both activities (realizing only good SET-scores, or a high research output, or neither). Next, our results clearly suggest that granting only marginal research time to professors is inefficient in terms of improving overall academic performance. In fact, a large majority of the professors with low research percentages did not succeed in realizing any research output. Apparently, a minimal amount of research time is needed for a professor to be able to produce research outputs. This finding seems intuitively plausible.

The results reported in this paper should, however, be interpreted with some caution. First of all, one should not generalize the results found here to other evaluation settings at other faculties or universities without further study. For instance, when the university faculty board and research council explicitly consider the professor’s opinion and/or past teaching effectiveness and research productivity in the decision on the research and teaching percentages, both teaching and research time variables are no longer exogenous. In such settings, the outcomes of the conditional BoD-model do not imply any causal interpretation. The conditional BoD-model then allows us to detect ‘only’ correlations between performance and scope. More generally, Hattie and Marsh (2002, p. 606) warned that departmental ethos and other characteristics of the department may

affect research productivity, teaching effectiveness, and, hence, the exact economies of scope between both core activities. Second, and somewhat related to this first remark, because of limited data availability, we were not able to explicitly account for a number of exogenous factors which have been found to significantly impact on teaching effectiveness and/or research productivity (and, thus, the teaching-research nexus). Here, we think of the family situation of the professor (e.g., children versus no children; see, for instance, Bellas & Toutkoushian, 1999, and Hunter & Leahey, 2010), the (teaching and research) ability of the professor (e.g., Marsh and Hattie, 2002; and Stack, 2003), whether or not the professor has research collaborations with other academics (e.g., Maske *et al.*, 2003; and Ramos *et al.*, 2007), etc. As such, the results might be (partially) driven by unobserved heterogeneity (i.e., omitted variable bias).

We see multiple avenues for follow-up research. First, as noted above, the exact nexus between teaching and research (i.e., both its existence and strength) likely varies with the particular (exogenous) circumstances and conditions. Therefore, it seems interesting to apply the proposed methodology in other teaching and research environments, to check for recurring patterns in the results. In the same vein, it would be useful to apply the non-parametric robust and conditional BoD-model that we presented here to data that have been investigated in previous studies (which typically used parametric approaches, such as correlation analysis, regression analysis, etc.). Specifically, we can then verify whether the BoD-model, which imposes no a priori assumptions on the teaching-research relationship, generates very different results than the more traditional approaches used in existing studies. A final possible direction for future research consists of expanding our study with other background (exogenous) variables that have been found to influence the link between teaching and research (family situation, ability, etc.; see above).

Literature

- Allen, M. (1996). Research productivity and positive teaching evaluations: Examining the relationship using meta-Analysis. *Journal of the Association for Communication Administration* 2: 77-96.
- Badin, L., Daraio, C., and Simar, L. (2010). Optimal bandwidth selection for conditional efficiency measures: A data-driven approach. *European Journal of Operational Research* 201(2): 633-640.
- Barnett, R. (1992). Linking teaching and research: A critical inquiry. *The Journal of Higher Education* 63(3): 619-636.
- Bellas, M.L., and Toutkoushian, R.K. (1999). Faculty time allocations and research productivity: Gender, race and family effects. *The Review of Higher Education* 22: 367-390.
- Braxton, J.M. (1996). Contrasting perspectives on the relationship between teaching and research. *New Directions for Institutional research* (Vol. 2). San Francisco: Jossey-Bass: 5-14.
- Brown, R.B., and McCartney, S. (1998). The link between research and teaching: Its purpose and implications. *Innovations in Education and Teaching International* 35(2): 117-129.
- Cashin, W. E. (1995). Student ratings of teaching: the research revisited. *IDEA Paper* Nr.32.
- Cazals, C., Florens, J.P., and Simar, L. (2002). Nonparametric Frontier Estimation: A Robust Approach. *Journal of Econometrics* 106 (1): 1-25.
- Centra, J.A., and Gaubatz, N.B. (2000). Is there gender bias in student evaluations of teaching. *Journal of Higher Education* 71(1): 17-33.
- Charnes, A. Cooper, W.W., and Rhodes, E. (1978). Measuring the efficiency of decision making units. *European Journal of Operational Research* 2: 429-444.
- Chen, Y., Gupta, A., and Hoshower, L. (2006). Factors that motivate business faculty to conduct research: An expectancy theory analysis. *Journal of Education for Business* (March/April): 179-189.
- Cherchye, L., De Rock, B., and Vermeulen, F. (2008). Analyzing cost-efficient production behavior under economics of scope: A nonparametric methodology. *Operations Research* 56(1): 204-221.
- Cherchye, L., Moesen, W., Rogge, N., and Van Puyenbroeck, T. (2007). An introduction to 'benefit of the doubt' composite indicators. *Social Indicators Research* 82: 111-145.

- Clark, S. (1986). The academic profession and career: Perspectives and problems. *Teaching Sociology* 14: 24-34.
- Daraio, C., and Simar, L. (2005). Introducing Environmental Variables in Nonparametric Frontier Models: A Probabilistic Approach. *Journal of Productivity Analysis* 24 (1): 93-121.
- Daraio, C., and Simar, L. (2007a), Advanced robust and nonparametric methods in efficiency analysis: Methodology and applications. *Series: Studies in Productivity and Efficiency*. Springer.
- Daraio, C., and Simar, L. (2007b), “Conditional Nonparametric Frontier Models for Convex and Nonconvex Technologies: A Unifying Approach”, *Journal of Productivity Analysis*, Vol. 28, pp. 13-32.
- De Witte, K., and Kortelainen, M. (2009). Blaming the exogenous environment? Conditional efficiency estimation with continuous and discrete environmental variables. *CES Discussion Paper Series DPS 08.33, MPRA Paper 14034*.
- De Witte, K., and Rogge, N. (2010). To publish or not to publish? On the aggregation and drivers of research performance. *Scientometrics* 85(3): 657-680.
- De Witte, K., and Rogge, N. (2011). Accounting for exogenous influences in performance evaluations of teachers. *Economics of Education Review*. In Press.
- Faia, M.A. (1976). Teaching and research: Rapport or mesalliance. *Research in Higher Education* 4: 235-246.
- Farrell, M.J. (1957). The measurement of productivity efficiency. *Journal of the Royal Statistical Society, Series A CXX Part 3*: 253-290.
- Feldman, K. (1987). Research productivity and scholarly accomplishment of college teachers as related to their instructional effectiveness: A review and exploration. *Research in Higher Education* 26: 227-298.
- Feldman, K.A. (1996). Identifying exemplary teaching: Using data from course and teacher evaluations. In M.D., Svinicki and R.J., Menges (Eds). *Honoring Exemplary Teaching: New Directions for Teaching and Learning* No. 65. San Francisco: Jossey-Bass.
- Feldman, K.A. (1997). Identifying exemplary teaching: Evidence from student ratings. In Perry, R.P. and Smart, J.C. (Eds.). *Effective Teaching in Higher Education: Research and Practice*. Agathon Press. New York: 93-143.
- Fox, M.F. (1992). Research, teaching, and publication productivity: Mutuality versus competition in academia. *Sociology of Education* 65(4): 293-305.

- Fox, M.F., and Milbourne, R. (1999). What determines research output of academic economists? *Economic Record* 75: 256-267.
- Friedrich, R., and Michalak, S. (1983). Why doesn't research improve teaching? Some answers from a small liberal arts college. *Journal of Higher Education* 54: 145-163.
- Hattie, J., and Marsh, H.W. (1996). The relationship between research and teaching: A meta-analysis. *Review of Educational Research* 66(4): 507-542.
- Hunter, L.A., and Leahey, E. (2010). Parenting and research productivity: New evidence and methods. *Social Studies of Science* 40(3): 433-451.
- Kremer, J. (1990). Construct validity of multiple measures in teaching, research, and service. *Research in Higher Education* 32: 351-361.
- Lee, S., and Bozeman, B. (2005). The impact of research collaboration on scientific productivity. *Social Studies of Science* 35(5): 673-702.
- Locke, W. (2004). Integrating research and teaching strategies: Implications for institutional management and leadership in the United Kingdom. *Higher Education Management and Policy* 16(1): 101-120.
- Lotka, A.J. (1926). The frequency distribution of scientific productivity. *Journal of the Washington Academy of Sciences* 16(12): 317-324.
- Marsh, H.W., and Hattie, J. (2002). The relation between research productivity and teaching effectiveness, complementary, antagonistic, or independent constructs. *The Journal of Higher Education* 73(5): 603-641.
- Marsh, H.W., and Roche, L. (2000). Effects of Grading Leniency and Low Workload on Students' Evaluations of Teaching, Popular Myth, bias, validity, or innocent bystanders? *Journal of Educational Psychology* 92(1): 202-228.
- Marsh, H.W. (2007). Students' evaluations of university teaching: dimensionality, reliability, validity, potential biases and usefulness. in R.P. Perry and J.C. Smart (Eds.). *The Scholarship of Teaching and Learning in Higher Education: An Evidence-Based Perspective* (pp. 319-383), Springer.
- Maske, K.L., Durden, G.C., and Gaynor, P. (2003). Determinants of scholarly productivity among male and female economists. *Economic Inquiry* 41: 555-564.
- Massy, W. F., and Wilger, A.K. (1995). Improving Productivity. *Change*: 10-20.

- Massy, W. F., and Zemsky, R. (1994). Faculty discretionary time. *Journal of Higher Education* 65(1): 1-22.
- Melyn, W., and Moesen, W. (1991). Towards a synthetic indicator of macroeconomic performance: Unequal weighting when limited information is available. Public Economics Research Paper 17. CES KU Leuven, Press.
- Neumann, R. (1992). Perception of the teaching-research nexus: A framework for analysis. *Higher Education* 23(2): 159-171.
- Pritchard, R.D., Watson, M.D., Kelly, K., and Paquin, A.R. (1998). *Helping Teachers Teach Well: A New System for Measuring and Improving Teaching Effectiveness in Higher Education*. The New Lexington Press, San Francisco.
- Ramos, R., Royuela, V., and Suriñach, J. (2007). An analysis of the determinants in Economics and Business publications by Spanish universities between 1994 and 2004. *Scientometrics* 71(1): 117-144.
- Rogge, N. (2011). Granting teachers the ‘benefit-of-the-doubt’ in performance evaluations. *International Journal of Educational Management* forthcoming.
- Sax, L.J., Hagedorn, L.S., Arredondo, M., and Dicrisi, F.A. (2002). Faculty Research Productivity: Exploring the role of gender and family-related factors. *Research in Higher Education* 43(4): 423-446.
- Stack, S. (2003). Research productivity and student evaluation of teaching in social science classes: A research note. *Research in Higher Education* 44(5): 539-556.
- Voeks, V.W. (1962). Publications and teaching effectiveness. *Journal of Higher Education* 33: 212-218.
- Westergard, J. (1991). Scholarship, research and teaching: A view from the social sciences. *Studies in Higher Education* 16: 23-28.
- Zaman, M. Q. (2004). Review of academic evidence on the relationship between teaching and research in higher education. Research Report No. 506. Department of Education and Skills. London.

Table 1: Summary statistics of variables

	Time investment (Exogenous input to the professor)		Output	
	Teaching time	Research time	Teaching score	Research score
Minimum	0.00 %	0.00 %	0.00	0.00
First quartile	64.48 %	0.00 %	2.99	0.00
Median	100.00 %	0.00 %	3.51	0.00
Third quartile	100.00 %	25.00 %	3.91	6.00
Maximum	100.00 %	100.00 %	4.76	418.00
St. deviation	30.36 %	28.68 %	1.13	61.65
Number of obs.	155	155	155	155

Table 2: Non-parametric bootstrap to test the presence of scope economies

Z	Teaching	Research
Coefficient	0.4628	0.1028
St. error	0.0305	0.0085
<i>p</i> -value	0.0360	2.22 E-16
R-squared	0.3175	0.2269

Figure 1: Teaching and research percentage

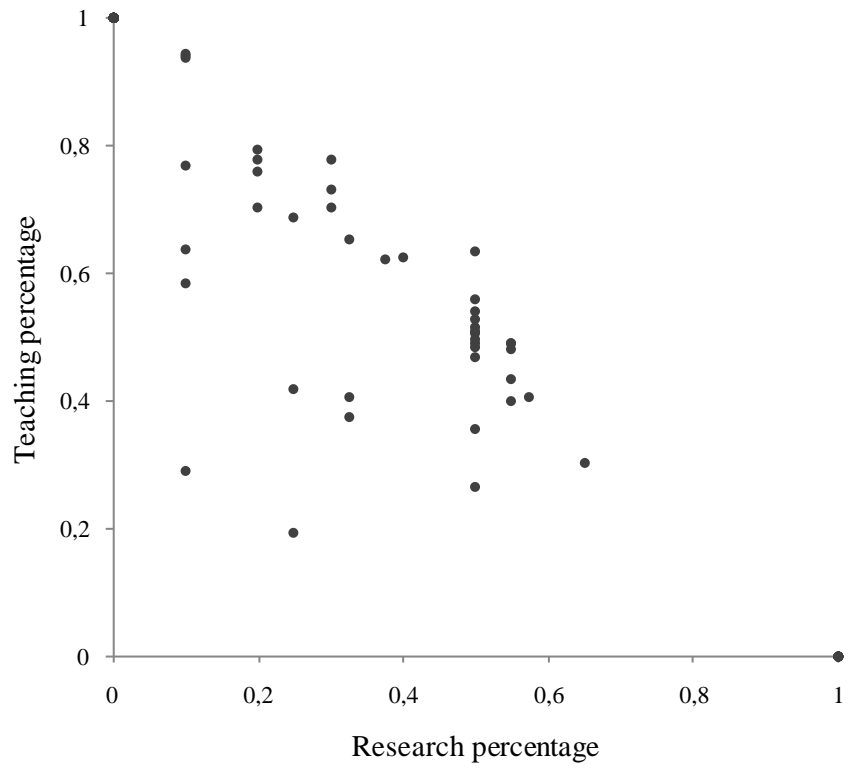


Figure 2: Graphical test for the presence of scope economies

Fig. 2a: $z_1 =$ Teaching time

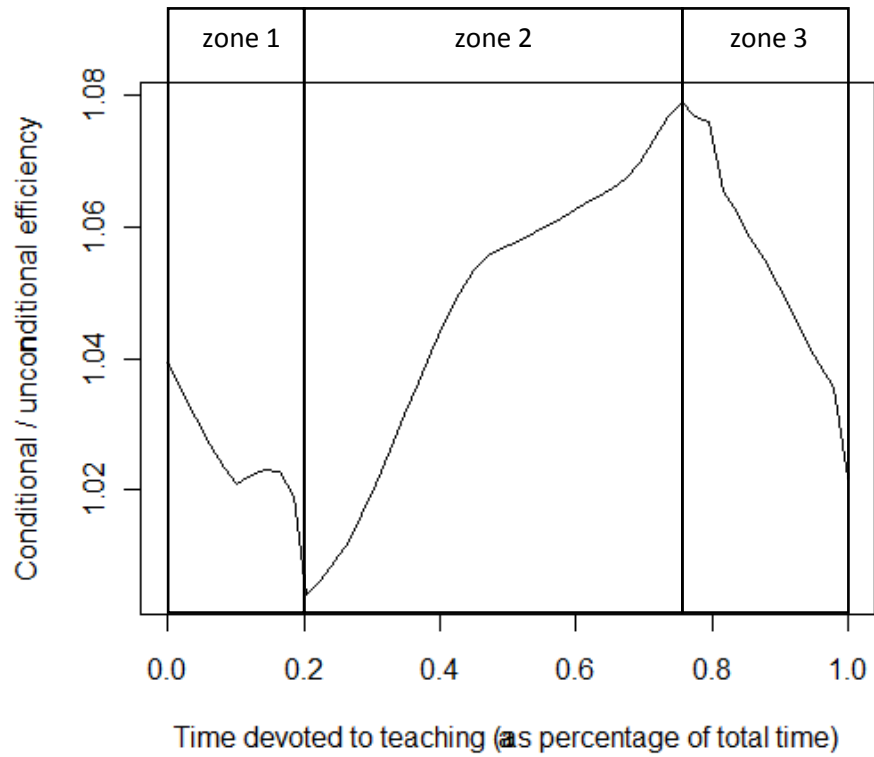
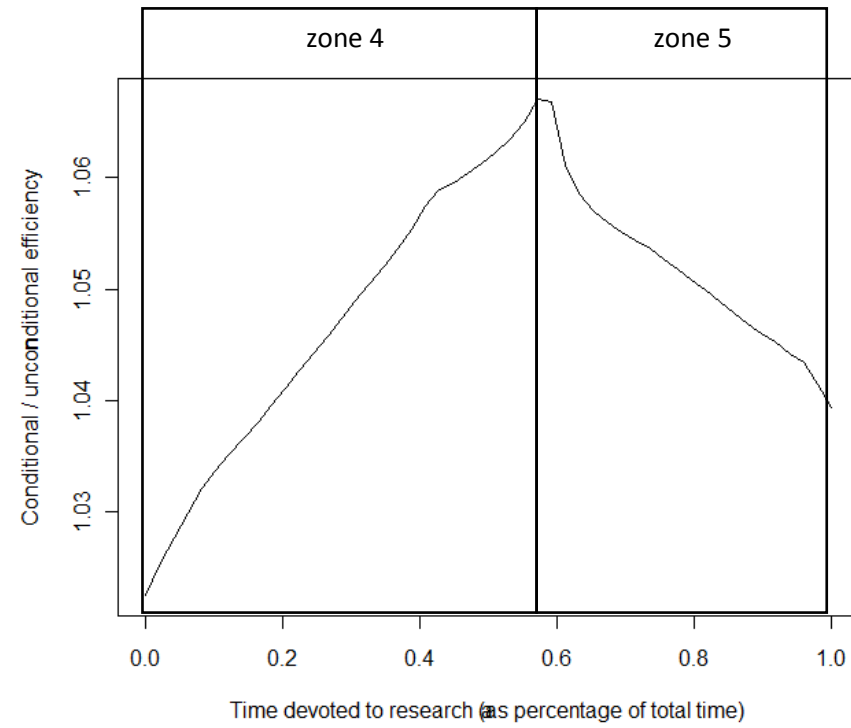


Fig. 2b: $z_2 =$ Research time



Appendix 1: Nine research output criteria for the RES-score

Research output criterion	Importance weight
1. International publication (Thompson Master List)	15
2. International book (based on own scientific work) as an author	15
3. National book (based on own scientific work) as an author	10
4. Finished research report (externally funded by a commissioner)	10
5. National scientific journals, chapter in international scientific book, Complete article in international proceedings (in all cases peer-reviewed)	7
6. Promoter of an externally funded project	5
7. Complete article in national proceedings (peer-reviewed)	7
8. Chapter in a scientific national book	7
9. Research/discussion paper	5

Appendix 2: The student questionnaire: 4 key dimensions of teaching with 16 questionnaire items

Key dimensions & Questionnaire items
Learning & Value
<ul style="list-style-type: none">- The lecturer justifies this part of the schooling in function of our formation.- I have learnt a lot from this course.- In general, the classes are good.
Examinations & Assignments
<ul style="list-style-type: none">- Exam requirements and evaluation criteria are clear.
Lecture Organization
<ul style="list-style-type: none">- The lecturer takes my knowledge and skills into consideration.- Learning materials (syllabuses, textbooks, texts, electronic resources) are well-organized and properly written.- The lecturer makes use of didactic aids (black board, transparencies, video, computer, language lab, skills lab, CDs, CD-ROMs,...) in a functional way.- The classes encourage students to reflect on the subject matter and assimilate it actively.- Lectures are well-structured.- The teaching tempo is good.
Individual Lecturer Characteristics
<ul style="list-style-type: none">- The lecturer deals with questions, suggestions and critical comments in a serene and constructive way.- The lecturer has a good contact with the students.- The lecturer speaks sufficiently loud and clear.- The lecturer treats each student respectfully.- The lecturer offers useful examples, applications, and exercises in class.- The lecturer explains the subject matter well.