(How) Do research and administrative duties affect university professors' teaching?

Article

Accepted Version

It is advisable to refer to the publisher's version if you intend to cite from the work. See Guidance on citing.

To link to this article DOI: http://dx.doi.org/10.1080/00036846.2015.1037438

Publisher: Taylor & Francis

www.reading.ac.uk/centaur

CentAUR

(How) Do research and administrative duties affect university professors' teaching?

Aurora García-Gallego

LEE & Economics Department-Universitat Jaume I, Castellón, Spain

Nikolaos Georgantzís

School of Agriculture Policy and Development, University of Reading, UK

Joan Martín-Montaner

IEI & Economics Department-Universitat Jaume I, Castellón, Spain

Teodosio Pérez-Amaral*

Quantitative Economics Department-Universidad Complutense, Madrid, Spain

Abstract

We analyze the interaction between university professors' teaching quality and their research and administrative activities. Our sample is a high-quality individual panel data set from a medium size public Spanish university that allows us to avoid several types of biases frequently encountered in the literature. Although researchers teach roughly 20% more than non-researchers, their teaching quality is also 20% higher. Instructors with no research are 5 times more likely than the rest to be among the worst teachers. Over much of the relevant range, we find a nonlinear and positive relationship between research output and teaching quantity on teaching quality. Our conclusions may be useful for decision makers in universities and governments.

Keywords: teaching quality, research, administrative duties

* Corresponding author: Teodosio Pérez-Amaral, *Quantitative Economics Department-Universidad Complutense, Madrid, 28223, Spain,* teodosio@ccee.ucm.es

I. Introduction

Since the 12th century, universities have experienced significant changes regarding both their functions and the way they are run, as recently summarized in Martin (2012). As pointed out by Spiller and Zelner (1997) and Lindbeck and Snower (2003), for the last century and a half, there has been a widespread idea that universities encompass the presumably complementary functions of increasing current knowledge through research and spreading it to the new generations through teaching. However, this agreement on the functions of the university as an institution does not mean that there is only one way of running it. In fact, in some relevant cases, a strategy of teaching specialization has been adopted[1], while in other cases, research has been undertaken in research oriented institutes. An alternative approach to the provision of research and teaching would be a combination of institutional diversification together with individual specialization. In this scenario, although the university aims at providing both high-quality teaching and high-quality research, some faculty members specialize in one of them. In any case, institutional or individual specialization is rather exceptional and most universities world-wide seem to adopt the synergic approach, based on the complementarity between teaching and research, motivating their faculty members to perform a mix of the aforementioned tasks together with certain amount of administrative duties. Taking this observation as a positive indication on the perceived synergic relation between different types of academic duties, the normative question is whether the actual mix of the corresponding inputs and outputs is the optimal, or, alternatively, whether it could be improved by a redefinition of the obligations and objectives set by governments and university authorities.

Apart from the objectives and incentive strategies set at an institutional level, an individual instructor's skills in research and a researcher's teaching quality might also be determined by synergies arising from the joint dedication to both duties resulting in the provision of a higher quality output. From the interplay among intrinsic (individual, supply-driven) and extrinsic (institutional, demand-driven) motivators and the time constraint binding an academic's dedication to these activities, two opposing effects emerge regarding the complementarity or substitutability among tasks: (i) a substitution effect due to the fact that allocating more time to one activity leaves less time available for the other and (ii) a complementarity effect due to

---

[1] In various countries, we can find higher education institutions that have focused only on the teaching side, such as the *Grandes Écoles* in France, the liberal arts colleges in the United Kingdom or the *Fachhochschulen* in Germany.

possible synergies between activities. Choosing between diversification and specialization either at the institutional or the individual levels, requires confirming whether synergies arise at an individual level.

It is often assumed that research may have positive spillovers on teaching because it facilitates an up-to-date choice and deeper understanding of topics, and a more rigorous approach to the subjects taught. Furthermore, the time devoted by researchers to their students may have a higher quality than that devoted to them by non-researchers. To the best of our knowledge, research on the validity of these or similar statements is, so far, inconclusive. In fact, despite addressing similar research questions, a number of related studies differ in "the variables investigated, their measurement, as well as the investigated population" as pointed by Verburgh, Elen and Lindblom-Ylänne (2007), thus lacking a common framework. On the other hand, and despite this lack of consensus, most studies conclude that a very weak relation between research and teaching can be found, or even no relation at all, see Marsh and Hattie (2002).

In this paper, we turn to a particularly rich annual panel data set of 604 individual university professors, over the period 2002-2006. The panel was extracted from the staff files of the Universitat Jaume I in Castellón (Spain). The sample of professors covers a variety of different disciplines including humanities, social sciences, economics, management, natural sciences and engineering. Teaching quality evaluation is compulsory for all instructors. Thus, we minimize the self-selection problems, omitted variables biases and self-selection biases that have been common in some of the previous literature. We use several indicators of the quantity and quality of teaching, research output, and the amount of administrative duties performed.

Summarizing our results, we find that professors with a typical research output are somewhat better teachers than professors with less research. Moreover, non-researchers are five times more likely than researchers to be poor teachers. In general, the quality of university-level teaching is positively related with published research across most levels of research output. Our results may be useful for policy makers and decision makers in universities.

The remainder of the paper is organized as follows. Section II consists in a review of the previous literature and the current situation of the debate. Section III proposes a theoretical benchmark. Section IV presents the database. The econometric estimations and discussion of the main results are contained in section V. Section VI concludes.

II. Previous studies

The existence and the nature of a relation between research productivity and teaching performance of academics has been for years at the core of a strong debate concerning the design of universities all over the world. To what extent and how are these activities linked? Although the existing literature offers reasonable justifications for both a positive and a negative relation between them, to the best of our knowledge, there is little, if any, evidence on the links among all the professorial activities as required by the multidimensional description of the input and the output spaces.

On the one hand, common wisdom usually leads us to think that the best researchers should also excel in teaching, provided they are enlarging the bulk of knowledge on a given discipline that should be later taught at the classroom. Successful research also implies several individual characteristics such as higher interest in learning and more ability to motivate students. Besides, research requires being more organized, which in turns helps to be more effective to hand down knowledge. On the other hand, it could be argued that good research requires a high degree of specialization which is not compatible with the broad view that students expect to find in their instructors. In fact, Friedrich and Michalak Jr. (1983) conclude that the higher organizational capability of researchers does not compensate the fact that students perceive them as less knowledgeable than instructors that do not research. A more extensive discussion on this topic can be found in Friedrich and Michalak Jr. (1983) or in Marsh and Hattie (2002).

The answer to this dilemma, therefore, becomes an empirical issue. However, empirical research is not as extensive as one would have thought and has not provided yet conclusive results on this hot topic. In a classic meta-analysis, Feldman (1987) reviews 29 papers published between 1950 and 1984 to conclude that "on average, there is a very small positive association between the two variables". Hattie and Marsh (1996) also rely on a meta-analysis based on 58 contributions to obtain a more pessimistic conclusion, openly speaking of the relation between research and teaching as an "enduring myth".

More recent research does not establish anything close to a stylized fact, as it also provides mixed evidence. Thus, Noser, Manakyan and Tanner (1996) differentiate the effect of research depending on the level of the courses taught: a positive but small relation is found for undergraduate level, whereas mixed results appear at the graduate level. Shin (2011) finds different signs depending on the indicator used for research productivity. More conclusive, Marsh and Hattie (2002) find a close-to-null relation between the two activities. The papers above show that a crucial point in this empirical literature comes from the way research productivity and teaching performance are

measured. Regarding the case of teaching performance, student's evaluations of teaching (SET) are in most cases the only indicator available.

Under such circumstances, the usefulness of the indicator relies in its availability, despite several shortcomings that the empirical literature has highlighted through a large body of work. Perhaps the most obvious shortcoming is that it does not measure the improvement in the student's knowledge, but, in the best of the cases, the student's own perception of this improvement. This perception should obviously be affected by the teacher's performance, but there are also other issues which may highly correlate to evaluations. The two key elements which seem most likely to bias student's evaluations are the size of the class and the grade students expect to reach. With regard to the first issue, Bedard and Kuhn (2008) conclude that class size has a negative impact, although McPherson (2006) and McPherson, Jewell and Kim (2009) limit this impact to the principles level.

More relevant to assess the validity of SETs as an indicator of teaching quality is the impact that students' expected grades might have on their evaluation of teachers. Early research papers such as Heilman and Armentrout (1936) failed to observe any link between grades and evaluations. However, other authors have detected more recently some statistical evidence for this relationship. This is the case of Krautman and Sander (1999), Isely and Singh (2005), McPherson (2006), Langbein (2008), McPherson, Jewell and Kim (2009), Weinberg, Ashimoto and Fleischer (2009), or Ewing (2012). This outcome is explained in most cases through the increasing relevance of SETs to determine faculty's earnings and promotion, which would lead teachers to get high evaluations by achieving a reputation of leniency in grading. However, not all these papers take into account the endogeneity of expected grades. When they do so, the evidence is mixed. Seiver (1983) and Zangenehzadeh (1988) control for endogeneity and do not find evidence that teachers inflate grades to obtain high SET scores, whereas Krautman and Sander (1999) still do.

Despite being partially caused by the interest of faculty to obtain better SETs or not, grade inflation in higher education has been observed in the last decades. Krautman and Sander (1999) have pointed this out, whereas Babcock (2010) finds that higher nominal grades lead students to spend less time studying. If students are more sensitive to expected grades than to their own learning, active researchers (which are supposed to be stricter) could be punished in SETs despite having a positive effect in students learning. Babcock's (2010) results address to the impact that grade leniency has on student's effort and, therefore, their actual learning. Several papers in the last years have analyzed the relationship between SETs and the effective learning achieved by

students, with mixed results. Thus, Weinberg, Ashimoto and Fleischer (2009) find that SETs are positively related to current grades but unrelated to learning (measured by the grades students achieved in subsequent courses). Using a similar methodology, Carrell and West and (2010) and Braca, Paccagnella and Pellizzari (2011) find even a negative relationship. On the other hand, Beleche, Fairis and Marks (2012) use results on a common test taken by most students to measure their actual learning to find a positive link with SETs. After all, SETs still are and will most probably be widely employed to measure teaching quality. Empirical research, therefore, will focus on it, although all the mixed evidence quoted above suggests that we must be very cautious, since in many cases it is the only available but not necessarily a flawless indicator of teaching quality. However, when alternative indicators are available, as in Harrison, Douglas and Burdsal (2004), the degree of correlation among them seems to be very high.

With regard to research productivity, most of the published empirical research uses the number of publications as the main indicator. However a few papers opt for qualitative indicators such as citation counts (Stack, 2003, among the most recent) or account for the quality of the journals (Noser, Manakyan and Tanner, 1996). Not surprisingly, this is not an innocuous option. In Feldman (1987), papers using quantitative indicators show a positive but weak relation with SETs, whereas in most cases, those based on citation counts do not get any. Neither do subsequent papers such as Gomez-Mejia and Balkin (1992); however, these authors report a positive correlation between teaching evaluations and publications in top-tier journals, and no correlation at all in the case of books. The opposite outcome can be found in Shin (2011). This author performs a covariance analysis to find that teaching is positively related to books and domestic journal publication and negatively related to international journal publications.[2] Besides, this former association is found to be stronger in some academic disciplines.

The results above take to the forefront the heterogeneity existing in the research field. This heterogeneity has been analyzed in the literature at two levels: academic disciplines and individual researchers. First, two different papers focus on the existence of heterogeneity at academic discipline level, offering, once again, mixed evidence. Both Porter and Umbach (2001) and Marsh and Hattie (2002) rely on a multilevel model which allows to differentiate the relationship between teaching and research across departments. The results were hardly comparable, as the latter used SETs as the dependent variable, whereas the former tried to explain research

---

[2] This author interprets his result in terms of the narrowness associated to good researchers (who mostly publish in international journals). Therefore, according to him, this outcome is reflecting the role played by those abilities which characterize good and bad researchers/instructors.

productivity, being the teaching load one of the explanatory variables. The interesting point, however, is that academic discipline did matter for the former, whereas it did not for the latter.

A second aspect regarding heterogeneity in research productivity refers to the obvious fact that the number of publications is not evenly distributed across faculty. On the contrary, top researchers concentrate a high share of the total number of publications, which implies a skewed research productivity distribution. Stack (2003) reports a positive and significant link between research and teaching, by introducing nonlinearity in the estimation process against the linear relationship assumed by most empirical papers. This represents a promising line of research which will be followed in the present paper.

To summarize the literature reviewed in this section, we could conclude that, with very few exceptions, there is some evidence of a weak relation between teaching and research. Some recent papers have focused on asking why this weakness in the relationship. Answers have ranged from emphasizing methodological issues to pointing the relevance of context in explaining the variance in the relation. All these approaches raise interesting issues which are developed to different degrees in the analysis we present in the following sections.

III. Framework

A. *Professors as instructors, researchers and managers*

Most academic positions in universities require three types of duties, namely teaching courses, conducting research in their respective fields, and performing administrative duties. There is some flexibility as to how professors allocate their time among them. Let us summarize here our views on the interplay among the three types of duties that will help us define the theoretical framework for our analysis and, subsequently, interpret our results.

In the case of most Spanish universities, teaching load is uniformly distributed among faculty members with similar contracts, unless a reduction has been applied due to project management, top management positions, etc. Ultimately, the amount and quality of teaching is monitored by academic authorities like the Teaching Vice-rector, the Faculty Cloister, the Dean, Department Heads, etc.

In Spanish universities, teaching quality is unfortunately not a crucial issue for an academic's promotion. For example, student satisfaction surveys have a weight of approximately 5% in the

promotion from pre-tenure to tenure, and from tenure to full professor.[3] However, in the Universitat Jaume I, teachers have to achieve a minimum student satisfaction level in order to avoid triggering an uncomfortable process checking on their skills and teaching methodology.[4] Nevertheless, there seems to be no doubt that the extrinsic incentives for Spanish university instructors to deliver good lectures is much lower than the incentives to publish. This incentive structure is reproduced at more junior and non-tenure track contracts, with some extra requirements regarding a minimum teaching experience to access to assistant professorship. However, such requirements regard only quantity, not quality.

Like in most universities around the world, instructors in Spanish universities have strong incentives to perform and publish research because it has an important effect on their promotion. Unless they are already at the top of their academic careers, the salary is somewhat increased by their research output. Nevertheless, given that research quantity requirements cannot be as strictly specified as individual teaching loads, many full time university teachers do little or no research.[5] Although some internal rules contribute to rewarding good research with reductions in teaching load even very high quality research does not guarantee having to teach significantly less.

The administrative load of academics in the Spanish university system is variable and negotiable. It usually requires acquiescence in the part of the instructor. However, top administration duties usually carry a salary increase, a statutory teaching load reduction and, sometimes, an additional guaranteed sabbatical year and a higher status within the university community.

B. A model

Consider the following straightforward extension of Becker's (1975) professorial decision making model, adding to the original research-teaching setting a third type of output, university administration.

---

[3] The Spanish National Agency for the Evaluation of Quality and Accreditation (ANECA) includes a guide for applicants that ask for promotion (http://www.aneca.es/Programas/ACADEMIA).

[4] At the Universitat Jaume I, teachers are invited to submit written reflections to the vice-rector of teaching if they have worryingly low scores in the student satisfaction survey.

[5] There are many reasons for not doing research, like lack of training, lack of infrastructure, lack of a research culture and group atmosphere, lack of interest, personal inability or lack of incentives (e.g. a high opportunity cost of time).

The model assumes that a professor i maximizes the utility function $U_i(Q_{T_i}, Q_{R_i}, Q_{A_i}, Q_{C_i})$, whose arguments correspond to the four outputs Q_{ji}, (where $j = Teaching,\ Research,\ Administration,\ Consumption$) yielded by the professor's actions. The professor decides on the time spent on each one of the four activities.[6] The four outputs monotonically increase in the time spent on each one of them, although potentially synergic activities like teaching and research may mutually benefit each other and the time spent on one of them may raise the output of the other. In fact, due to this interaction between potentially synergic activities, net complementarity or substitutability may be observed among them, despite the clear cut net substitution effect among the times dedicated to the generation of the four outputs as dictated by the time constraint: $T_{R_i} + T_{T_i} + T_{A_i} + T_{C_i} \leq T_i$, in which T_i corresponds to the total amount of active time available to i. Note that, originally, the model's comparative statics aimed at yielding testable hypotheses for empirical and normative analysis of different rewards to teaching and research. Nevertheless, the model's implications for technological shocks affecting the relative shadow prices of the outputs to an individual professor are immediately applicable to the heterogeneity usually observed among individuals or groups of professors, provided that their idiosyncratic and institutional differences are interpreted as between-professor changes in the parameters α_{jji}, α_{jki} of the expression $Q_{ji} = \alpha_{jji} \cdot T_{ji} + \alpha_{jki} \cdot T_{ki}$, specifying professor-specific direct and cross-activity impacts of the time dedicated, respectively, to activities j and k on the output of j. As noted by Becker (1975), the indeterminate effect of one activity on the other allows for multiple predictions concerning the substitutability or complementarity among a professor's outputs, which here include administrative tasks. Thus, it is left open for empirical research to determine the signs of such cross-activity effects.

IV. The data

The dataset is a yearly panel for 604 individual instructors of the Universitat Jaume I of Castellón (Spain). The panel data contains information on 69 academic variables for five years, 2002-2006.[7] During this period, the university has systematically compiled individual information on a large number of variables concerning each faculty member's teaching quality and quantity, research performance and administrative duties. Since all the instructors of the university are required to

---

[6] Consumption represents all other activities in life that a professor needs to combine with his main professional duties. Therefore, consumption represents resources dedicated to leisure, housekeeping, purchasing, etc.

[7] The sample of instructors in our study is relatively stable during this period. Mobility in Spanish universities is low. Therefore, while some instructors have joined and others have left the university during our reference 5-year period, the average time of a subject in our sample is 4.19 years.

undergo teaching evaluations, we avoid a possible self-selection bias. The first 23 variables correspond to indices of teaching. The next 14 are indices related to administrative duties, and the last 32 indices of research. The 604 professors are from 25 academic departments covering a variety of different disciplines including humanities, social sciences, economics, management, biomedical, natural sciences and engineering.

As we are interested in relating the quality of teaching with administrative duties, research, and other variables, we use the following synthetic indices to approximate the concepts of interest:[8]

Teaching quality: Denoted by *Teachqual,* teaching quality is the outcome of the students' opinion university survey for each instructor in a given academic year. It is obtained from students' responses to an overall satisfaction survey question using a 0 to 9 Likert scale. In a specific module, an instructor's score is the average of the responses received. Given that each teacher usually teaches more than one module, the scores used here are each teacher's average evaluation from different modules. It is included in our dataset as a real number between 0 and 9, with two decimals.

Research outcome: Denoted by *Research1,* the research attainment of each academic in the university is available through an online application managed by each department. An independent panel assesses all outputs and applies a similar set of rules for all departments as to what counts as research productivity. The contributions are calculated using a quality adjusted index that weights more heavily research published in well ranked international peer-reviewed journals than other types of publications. A moving sum of the previous four years is calculated for each individual so that, for each relevant year of the panel, information accounts only for the most recent research. Several other variations of the research outcome measures were also considered.

Quantity of teaching: Denoted by *Teachcours,* the teaching of each academic is the weighted sum of the number of undergraduate and postgraduate courses taught. Other measures that include, for instance, the number of courses taught abroad were also considered.

Administrative duties: The quantity of heavy administration duties, denoted by *Admin1* is measured by the sum of the number of years in top management positions held by each academic within the university. These positions, such as chancellor, vice-chancellor, dean, vice-dean, or

---

[8] The abbreviations of the variables reported here correspond to a larger set of variables tested in alternative model specifications. The notation and numbering is aimed at distinguishing them from other variables (not reported here) measuring similar magnitudes.

head of department, entail mandatory reductions in the quantity of teaching. We also consider other less demanding tasks denoted by *Adminlight* measured as the sum of the number of years served in less demanding administrative positions held by each academic within the university. These positions, involving membership to university or college boards, tutorship of students abroad or similar tasks, do not have a mandatory reduction in the teaching load.

With respect to other variables, we compute the *Participation in Academic Committees* as the number of academic committees to which each instructor belonged in a specific year. These are internal quality evaluation committees. They usually involve small *ad hoc* teams of faculty and administrative staff, preferably volunteers, reviewing or evaluating specific processes in accordance to a broader national or international set of rules regarding academic or administrative innovation.

The *Publication of teaching materials* is the number of books and multimedia teaching materials with ISBN published by each instructor during the previous four years of each year in the sample. For each of the 25 departments of the university, the *Department* is coded using a binary variable, 1 if the instructor belongs to a given department and 0 otherwise. And finally, the *Gender* is coded using a binary variable (Female) taking the value 1 for females and 0 for males.

In Figure 1 we show the histograms of the indices mentioned.

Figure 1. Histograms of education indexes
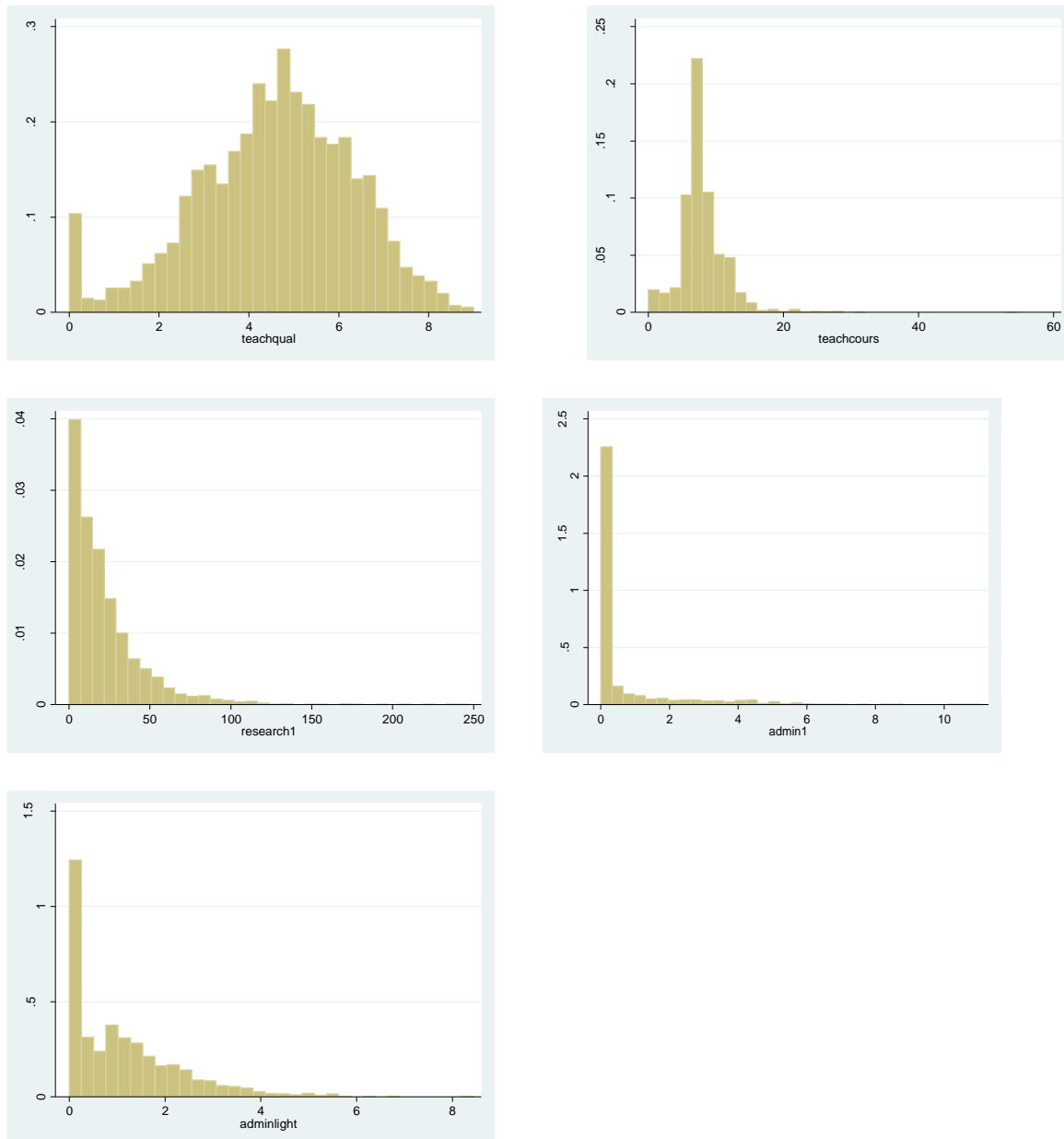
The figures suggest that there is sufficient variation in the indices to make valid inferences. Several of them show accumulations at zero, reflecting that many professors do not do any research or administration. They also show that a substantial number of professors obtain a very low teaching evaluation.

Figure 2 shows the histograms of teaching quality, divided into two categories. The left panel shows the scores for those professors who can obtain some research output (*Research1*>0). The right hand side panel contains the scores of those professors with no research output. Notice that we find a higher accumulation of individuals with very low scores in the group of non-researchers.

Figure 2. Teaching quality of researchers and non-researchers
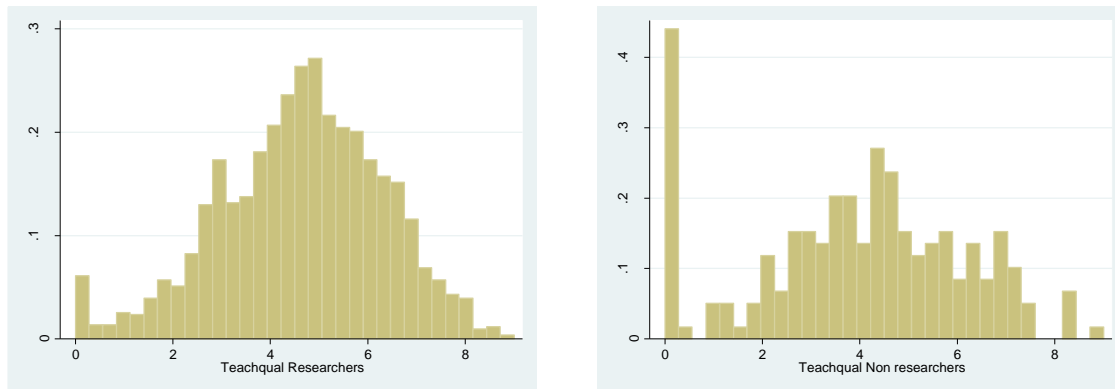
Researchers have, on average, higher teaching loads than non-researchers. The average teaching load for researchers with more than 5 teaching quality points is 8.35, while the average for non-researchers is 6.87. Therefore, we find that, on average, researchers teach 21.5% more than non-researchers. This is possibly due to the existence of part time lecturers who, in general, perform little if any research. The inclusion of this group of teachers in the analysis allows to avoid sample selection biases and provides and interesting comparison with the rest of the professors in the sample.

Figure 3: Teaching quality vs quality-adjusted research. Non parametric fit
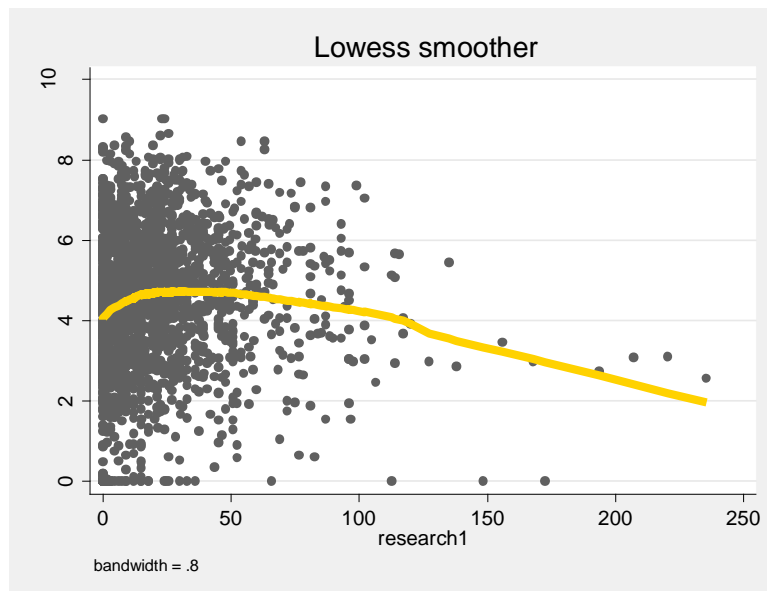
In Figure 3 we observe that when we pool all observations and perform a non-parametric adjustment using the Lowess smoother, we observe a nonlinear relationship between teaching quality and published research, with a maximum of 4.91 (over 9) *Teachqual* points on the

students' satisfaction scale at, approximately, *Research1* = 23 (quality-adjusted publication record), using a window of 5 observations. This suggests that the quality of teaching increases with the amount of published research up to a maximum level beyond which teaching quality decreases slightly up to 50 quality-adjusted publication points, remaining in all cases above the teaching quality of low research performers. Only for the upper 10% of our sample's research performance the relation between teaching quality and research becomes negative. Although these results do not imply causality in either direction, they reveal the existence of an empirical relationship that admits different interpretations. Thus, this exploratory analysis needs to be completed by considering other relevant variables in a multivariate model presented in the following section.

V. Econometric analysis and empirical results

In this section, we present an econometric model estimated from the yearly panel dataset using the indices shown in the previous section.

Some key questions that we would like to answer in this paper are: (1) How does research relate to teaching quality? (2) How do administration duties relate to teaching quality? (3) How does the teaching load relate to teaching quality? (4) How is the quality of teaching correlated with complementary pedagogic activities, such as participating in committees, publishing teaching materials or taking teaching improvement courses? (5) Are there gender or department-specific differences?

For answering these questions we will use the following equation that relates the quality of teaching with its determinants:

$$Teachqual_{it} = \alpha_i + \beta_1\, Adminlight_{it} + \beta_2\, Research1_{it} + \beta_3\, Research1sq_{it} + \beta_4\, Teachcours_{it}$$

$$+\beta_5\, Teachcourssq_{it} + \beta_6\, AcadCommissions_{it} + \beta_7\, Books_{it} + \beta_8\, EducationDept_{it} + \beta_9\, Female_i + u_{it}$$

$i = 1, ..., 604,\ individuals.$
$t = 2002, ..., 2006,\ years.$

This equation represents the perceived quality of teaching measured by the SET for a given teacher *i* in year *t*. We assume that the instructor maximizes the perceived quality of his teaching, *Teachqual*, in year *t* conditional on his/her circumstances as measured by the independent variables in year *t*.

$α_i$: the different constants represent the individual specific effects. They capture individual characteristics that do not change along time as well as the effects of variables that vary across individuals but do not change, or evolve slowly along time, such as the seniority of the teacher, the class size, the ability of the instructor and others. We expect different constants for different individuals in our sample.

$β_j$: for $j = 1,…, 9$, are the coefficients of each explanatory variable, which measure the partial effect of each variable on the dependent variable. They are unknown and need to be estimated from our data.

Adminlight: the amount of light adminstrative tasks performed by the teacher. This can be considered exogenous, since it is determined by previous performance as administrator, seniority and other considerations, but it is not affected by the current quality of teaching. We expect either no effect, or a negative sign of the estimated coefficient.

Research1: the amount of accumulated high quality research which is a result of ability, dedication, effort and success in the previous years. It is exogenous relative to the results of maximizing the current years´ teaching quality. We expect a positive sign of its estimated coefficient.

Research1sq: is the square of *Research1*. This term allows for more flexibility in the response of the quality of teaching to *Reseach1*, that is, a non-linearity, which is consistent with common sense and the previous exploratory analysis. Since *Research1* is exogenous, this square term is also exogenous. We expect a negative sign, compatible with the existence of a maximum.

Teachcours: the quantity of teaching in a given year is also exogenous. In this university it is determined by the type of contract and dedication, as well as the teaching reductions obtained in this year. It is not contemporaneously determined by the quality of teaching. We expect a positive sign, compatible with a maximum.

Teachcourssq: is the square of *Teachcours*. This term allows for more flexibility in the response of the quality of teaching, in this case a non-linearity, which is consistent with common sense. Since *Teachcours* is exogenous, this square term is also exogenous. We expect a negative sign.

AcadCommissions: in the university under analysis, the participation in academic committees is normally voluntary, and it is decided prior to observing the teaching quality in a specific year. Participating in academic committees may be a signal of a genuine interest in the quality of the

teaching. There may be a common factor that drives both teaching quality and the participation in academic committees, such as more interest in teaching. However, delivering higher teaching quality in a year does not necessarily relate with past and present belonging or having belonged to such a committee. Therefore, in our model, the participation to these committees is predetermined with respect to teaching quality. We expect a positive sign.

Books: these are teaching related books written by instructors. The decision of writing a manual is taken previously to the determination of teaching quality in a specific year. This variable is a signal of interest in teaching. Writing teaching related books is a predetermined variable, and a potential predictor for the quality of teaching. We expect a positive sign.

EducationDept: this variable takes the value of 1 when the instructor belongs to the Department of Education and 0 otherwise. This variable permits to measure the difference in the quality of teaching attributable just to the fact that the instructor belongs to that particular department. The decision to belong to a given department is previous to determining the quality of teaching, so we can also consider this variable as predetermined with respect to teaching quality. We expect a positive sign.

Female: this is a dummy variable that takes the value of 1 for females and 0 otherwise. It is an exogenous variable, since the quality of teaching at year t does not influence the gender of a person, which was determined biologically before his/her birth. However, the gender of the instructor may affect contemporaneously the perception of teaching quality by students. We have no particular expectation regarding the sign of its coefficient.

u_{it}: is a common random error affecting all the observations. This error may be due to several circumstances, as well as specification errors.

From these considerations we conclude that the above equation of *Teachqual* does not exhibit simultaneity and can be specified as a standard single equation panel data model, linear in the parameters, which can be estimated consistently and efficiently using the standard estimators for these models. In our equation causality runs in one direction only, from each of the explanatory variables to teaching quality but not vice-versa. We test this hypothesis empirically at the end of this section.

Nevertheless, one might argue that there is joint production of research, teaching and administrative service, and thus it may be that the causality between teaching and research (for example) works the other way around.

However, this is not the case here, since the quality of teaching, which is typically observed with a lag, cannot possibly influence current research output, which is the result of past decisions to do research, submit originals for evaluation and eventually publish, which is typically a lengthy process. Current quality of teaching cannot possibly affect current research output. It could only affect future research output, and with a significant lag of two to four years. This possibility may deserve further attention, and would need to be explored with a longer data set.

However, it is unlikely that a significant effect of teaching quality on research output can be found. Doing research requires specific skills over and above those required for being a good teacher. They include being at or near the frontier of knowledge at an international level, identifying relevant research topics, having the up-to-date technical knowledge required, possessing the needed writing skills, having access to funding (and other research resources) and having adequate co-authors. It is unlikely that being only a good teacher causes higher research output.

In any case, in this paper we focus on the current effect on teaching quality of variables that are measured contemporaneously and that can be among its plausible determinants. We estimate the model shown in Table 4, reporting our favorite equation for teaching quality. The random effects panel data model estimated by generalized least squares allows for nonlinearity in Core published research, as suggested by Figure 3 and also nonlinearity in the teaching load.[9]

Table 4. Teaching quality vs. its determinants. Random effects panel data model estimated by Generalized Least Squares

Dependent variable: Teaching quality

Explanatory variables	Estimate (*z*-statistic)
Light administration duties	-0.054 (-2.09)
Core published research	0.0085 (2.54)

---

[9] The equation can be thought of as a multivariate generalization of Figure 3. The Hausman test (Hausman, 1978) for estimating the equation with panel data by covariance or using the generalized least squares estimates, takes a value of 11.93 with a *p*-value of 0.15. Therefore, we do not reject the null of no correlation between individual effects and regressors, so we can use the Generalized Least Squares estimates in Table 4.

Core published research square	-0.000069 (-2.54)
Teaching load	0.469 (9.95)
Teaching load square	-0.031 (-10.07)
Participation in committees	0.121 (2.89)
Books, multimedia teaching materials with ISBN	0.029 (3.62)
Education department	0.668 (2.23)
Female	0.277 (2.24)
Constant	2.679 (13.31)
Number of observations	2008
Estimation technique	Generalized Least Squares
Joint significance	$\chi^2(9)=177.25$ Prob> $\chi^2=0.0000$

NOTE: *z*-statistics are distributed as approximately normal $N(0,1)$ under the null. All estimates are significant at the 1% except for light administration duties and females, which are significant at the 1.4 and 1.7%, respectively. STATA 12 is used for all calculations.

The nonlinear relationships are confirmed by the positive coefficient estimates of the linear terms in both equations, and the negative coefficient estimated of the square terms, all of them being statistically significant. As a whole, these estimates suggest that, first, for some (low) values of teaching load, more teaching may be positively associated with teaching quality, while for higher levels, too much teaching may be related to lower quality. More specifically, the quantity of teaching has a relates positively with teaching quality for relatively low teaching loads, which is compatible with the significantly lower teaching evaluations obtained by part-time teachers. The quality peaks around 7.62 courses. In summary, moderate amounts of teaching are positively associated with the teaching quality, while higher teaching duties apparently represent an excessive load which is negatively related with teaching quality. Since the average teaching load is 8h per week in the period under study, the observed arrangement in this university does not seem to be far from the optimum.

In fact, it may be unconceivable to think that research or teaching load will always be positively related to teaching quality for all possible values of research and teaching load. Searching for a linear or monotonic relationship, like has been done in the previous literature like in Marsh and Hattie (2002) may be incorrect and lead to flawed results.

Second, research output is positively associated with teaching quality for small amounts of research, below the median performance in our sample, until reaching a maximum at 57.43. Core research contributes positively to teaching quality but this effect decays slowly for high values of core research. That is, low to medium amounts of research are positively related to teaching, while high core research output may be associated with lower teaching quality. In both cases, our results allow for an intuitively appealing link between research and teaching quality. Thus, the "positive segment" for low to medium research activity could indicate that researchers are able to adapt their research achievements to teaching purposes. With regard to high research output, our indicator can reflect two -not necessarily alternative- different situations: researchers that publish a huge amount of articles per year, or researchers that publish mostly in top journals. Publishing a lot could imply less time (or interest) in teaching, whereas publishing in top journals sometimes means top research is too specific to be easily translated into better teaching, despite time availability and interest from the researcher's side. In this latter case, it would be very interesting to discriminate among quality of teaching by level of courses, since top research is likely to influence teaching quality at the graduate level, but less so at the undergraduate level. However, our data set includes only students' satisfaction on aggregate.

A third set of interesting results are obtained with respect to the effect of administrative tasks. In this study, we have distinguished between heavy and light administration duties. Our results confirm that the quality of teaching is not related to the latter. This may be due to the fact that such admin appointments entail mandatory reductions in the teaching load. However, the quality of teaching is negatively related to light administrative duties (an effect of -0.05), which do not entail a compensating reduction in teaching loads. Therefore, evidence seems to point towards administrative tasks distracting faculty from their teaching duties, unless they are released from part of their teaching load.

The positive link observed for the participation in academic committees deserves a specific comment. The estimate (0.121, with a significant z-statistic of 2.89) seems to contradict the conclusions above. However, we should keep in mind that membership to these committees and commissions is not compulsory and, although it represents certain amount of administrative

workload, it is often strongly related to teaching innovation. In other words, people participate in these commissions because of a previous interest in the improvement of teaching quality.

Some other indicators referred to interest in improving teaching quality are a mixed bag. On the one hand, the quality of teaching is positively associated to the elaboration of books or multimedia teaching materials with ISBN (the point estimate is 0.29 with a significant z-statistic of 3.62), whereas courses of pedagogic enhancement of the Teaching Support Unit have no statistically significant effect when entered in the equation of Table 4.

Finally, we analyze the existence of some specific effect by field of specialization and by gender. Faculty members at the Education Department are the only ones who display a specific effect, obtaining better teaching results than people with equal characteristics in other departments (they obtain, on average, a score that is 0.668 points higher). As far as gender effects are concerned, we find that female professors obtain better teaching results than their male counterparts (they obtain, on average, a score that is 0.277 higher than that of males).

C. Advantages of panel data

Authors like Mundlak (1987), Hsiao (2003) and Wooldridge (2010) provide evidence that the use of individual panel data typically allows a more sophisticated analysis than cross section data.[10] First, with panel data, the number of observations available is usually higher than typical cross sections, allowing for more degrees of freedom and a more precise estimation of the parameters of interest. Second, and most important, they allow us to control for individual heterogeneity, such as individual ability, by using individual specific dummies. The use of individual specific dummies also allows us to reduce the bias due to omitted variables. For instance, omitted variables that are constant or evolve slowly over time -such as the seniority, number of students per group, etc.-, can be captured by the different individual constants.

D. Omitted variables tests

We have performed several tests to check for the possible omission of relevant terms such as a quadratic one for *Adminlight,* which turns out to be insignificant with a z value of 0.26, and both heavy and medium administrative duties with z values of 0.16, which are also insignificant. We have also tested for the omission of other candidate explanatory variables that measure activities related to the quality of teaching or directly designed to boost teaching quality. All of them are

---

[10] This type of data has been used in most of the previous literature.

insignificant. The variables include courses of pedagogic enhancement, publishing teaching notes with ISBN, teaching articles published in journals or in the web, and the participation as instructor in experiences of European harmonization and advising.

F. Exogeneity

The exogeneity of the explanatory variables of the equation is necessary for the consistence of the generalized least squares estimates. The exogeneity is tested using Hausman tests that compare the chosen specification estimated by generalized least squares, that assumes exogeneity, with instrumental variables estimates that are consistent in case of endogeneity (although inefficient). The instruments used are one year lags of possibly endogenous variables: core published research, teaching load, participation in commissions and books, multimedia teaching materials with ISBN.

Two instrumental variables estimates are used. The first is applied to the last three years of the sample and the second uses the last two years. In both cases, the null hypothesis of exogeneity of the regressors is not rejected by the Hausman test which behaves under the null hypothesis as chi-square with 8 degrees of freedom, with computed values of 9.68 and 5.65, and exhibit p-values of 0.28 and 0.68, respectively. Since we find no evidence against exogeneity, the preceding estimates can be used for inference.

G. *Robustness*

The equation has been tested for robustness to outliers by restricting the values of *Research1* to be, alternatively, below 120 and 100, and it was found that the curvature with respect to *Research1* (as well as the rest of the coefficients) remains the same after this perturbation.

The equation has also been tested for robustness to the choice of sample by not considering the instructors with no research contribution, since they may behave differently, especially if it is decided that they may not be relevant for the comparison. It was found that the estimates of the alternative model remain very similar to those of the original equation after this perturbation.

As suggested by one the referees, robustness to the definition of the variable *Adminlight* was tested by incorporating a variety of administrative duties, difficult to identify, which were into a new variable *Adminlight2*. A point estimate of -0.006, was obtained, which is insignificant at the 50% instead of -0.05, which was barely significant at the 4%, which we obtained previously. This suggests that the effects of these variables are small or insignificant in our sample.

VI. Conclusions

Academic jobs require simultaneous performance of a variety of duties, including teaching, research and administration captured in a large number of alternative performance measures available. Assuming that teaching is a university's major contribution to the society, we have addressed the question whether teaching quality is related with other academic duties. Our results show that the answer to this question is not as straightforward as assumed in previous studies restricting attention on linear or monotonic responses. We find non-linear effects of the major variables associated to teaching quality, among which are research performance and the professor's teaching load. These effects and the role of a number of other variables are identified due to our particularly rich dataset that allows for a more sophisticated statistical treatment of the relevant information. Our dataset consists of the whole population of instructors at a medium-sized public Spanish university, which belong to a wide variety of fields of knowledge, and their evaluation is compulsory. This minimizes the self-selection problems, omitted variables biases and self-selection biases that have been common in some of the previous literature.

We estimate a panel dataset where the dependent variable is the quality of teaching (as measured by student's evaluations of teachers). The use of panel data allows us to control for different kinds of heterogeneity, such as the individual characteristics of the professors, their natural abilities, their department, and also to avoid the biases due to omitted variables such as the availability of teaching assistants, the size of the class and the seniority of the professor, among others.

The explanatory variables include an index of published research and its square, and several indicators for administrative tasks. On the one hand, the nonlinear estimate suggests that teachers who also carry out some levels of research achieve better evaluations by their students; however, a high degree of specialization in research apparently worsens the quality of teaching. On the other hand, students' evaluations of teachers who engage into administrative duties are not affected if those duties imply some reduction in the teaching load. Those evaluations though, are clearly worse for those teachers who do not see their teaching load reduced, although the administrative duties are considerable lighter than the latter ones.

Taken together, these results point towards some benefits if some type of light specialization is adopted. The most successful researchers and those more able for administrative tasks should reduce their teaching and concentrate in those activities. However, this should not imply increasing the teaching load for non-researchers and/or non-managers. Our estimations also detect an optimum teaching load, beyond which teaching quality declines.

For those indicators referred to faculty's efforts to improve the quality of teaching, we get mixed results. Thus, participation in committees and commissions related to teaching and publication of teaching materials are linked to better teaching, whereas courses of pedagogic enhancement of the Teaching Support Unit have no relation at all. Since the courses are contemporaneous to the student's evaluations, it is not unlikely that the complete effects could appear in future evaluations, but it is a bit surprising that no positive effects are observed at the same time they are being implemented.

In the first lines of this paper, we brought up recent debates regarding the way universities should organize themselves in order to carry out their two core functions: teaching and research, plus their own internal management. Although it seems clear that some degree of specialization should be adopted, it is not clear whether the institutional specialization is preferable to the individual one. In this sense, our results stand for the latter option, as it seems clear that some research activity is associated with higher quality of teaching, except in the case of top researchers. In any case, they make clear that the relation between these activities is not as straightforward, monotonically positive or negative, as presented in most of the relevant literature. Besides, there are no obvious reasons why these results would vary substantially in other universities or countries. However, it would be interesting to look for more universities which systematically collect similar data on their professors' teaching and research performance to make cross country comparisons. There is an urgent need for more studies of this kind and comparison across different universities and national systems, in order to replace anecdotal evidence and unfounded beliefs with systematic and rigorously analyzed information.

Acknowledgements

VII. References

Bedard, K., & Kuhn, P.J. (2008). Where class size really matters: Class size and student ratings of instructor effectiveness. Economics of Education Review, 27 (3), 253-265.

Becker, W.E. Jr. (1975). The university professor as a utility maximizer and producer of learning, research, and income. The Journal of Human Resources, 10(1), 107-115.

Beleche, T., Fairris, D., & Marks, M. (2012). Do course evaluations truly reflect student learning? Evidence from an objectively graded post-test. Economics of Education Review, 31 (5), 709-719.

Carrell, S. E., & West, J.E. (2010). Does professor quality matter? Evidence from random assignment of students to professors. Journal of Political Economy, 118 (3), 409-432.

Ewing, A.M. (2012). Estimating the impact of relative expected grade on student evaluations of teachers. Economics of Education Review, 31 (1), 141-154.

Feldman, K.A. (1987). Research productivity and scholarly accomplishment of college teachers as related to their instructional effectiveness: A review and exploration. Research in Higher Education, 26 (3), 227-298.

Friedrich, R.J., & Michalak, S.J. Jr. (1983). Why doesn't research improve teaching?: Some answers from a small liberal arts college. The Journal of Higher Education, 54 (2), 145-163.

Gomez-Mejia, L.R., & Balkin, D.B. (1992). Determinants of faculty pay: An agency theory perspective. The Academy of Management Journal, 35 (5), 921-955.

Harrison, P.D., D.K. Douglas & C.A. Burdsal (2004). The relative merits of different types of overall evaluations of teaching effectiveness. Research in Higher Education, 45 (3), 311-323.

Hattie, J., & Marsh, H.W. (1996). The relationship between research and teaching: A meta-analysis. Review of Educational Research, 66 (4), 507-542.

Hausman, J. A. (1978). Specification tests in econometrics, Econometrica, 46 (6), 1251-1271.

Hsiao, C. (2003). Analysis of Panel Data, 2nd Edition, Econometric Society Monographs, Cambridge University Press.

Isely, P., & Singh, H. (2005). Do higher grades lead to favorable student evaluations? Journal of Economic Education, 36 (1), 29-42.

Krautman, A.C., & Sander, W. (1999). Grades and student evaluation of teachers. Economics of Education Review, 18 (1), 59-63.

Lindbeck, A., & Snower, D.J. (2003). The firm as a pool of factor complementarities. IZA Discussion Paper 882.

Marsh, H.W. (1987). Students' evaluation of university teaching: Research findings, methodological issues, and directions for future research. International Journal of Educational Research, 11 (3), 263-253.

Marsh, H.W., & 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.

Martin, B.R. (2012). Are universities and university research under threat? Towards an evolutionary model of university specialization. Cambridge Journal of Economics, 36 (3), 543-565.

McPherson, M.A. (2006). Determinants of how students evaluate teachers. Journal of Economic Education, 37 (1), 3-20.

Mundlak, Y. (1978). On the pooling of time series and cross section data. Econometrica, 46 (1), 69-85.

Noser, T.C., Manakyan, H., & Tanner, J.R. (1996). Research productivity and perceived teaching effectiveness: A survey of economics faculty. Research in Higher Education, 37 (3), 299-321.

Porter, S.R., & Umbach, P.D. (2001). Analyzing faculty workload data using multilevel modeling. Research in Higher Education, 42 (2), 171-196.

Shin, J.C. (2011). Teaching and research nexuses across faculty career stage, ability and affiliated discipline in a South Korean research university. Studies in Higher Education, 36 (4), 485-503.

Spiller, P.T., & Zelner, B.A. (1997). Product complementarities, capabilities and governance: A dynamic transaction cost perspective. Industrial and Corporate Change, 6 (3), 561-594.

Stack, Steven. (2003). Research productivity and student evaluation of teaching in social science classes: A research note. Research in Higher Education, 44 (5), 539-556.

Verburgh, J.E., & Lindblom-Ylänne, S. (2007). Investigating the myth of the relationship between teaching and research in higher education: A review of empirical research. Studies in the Philosophy of Education, 26 (5), 449-465.

Vidal, J., & Quintanilla, M.A. (2000). The teaching and research relationship within an institutional evaluation. Higher Education, 40 (2), 217-229.

Walstad, W., & Allgood, S. (2005). Views of teaching and research in economics and other disciplines. American Economic Review, 95 (2), 177-183.

Wooldridge, J. M. (2010). Econometric analysis of cross section and panel data. 2nd edition, The MIT Press Cambridge, Massachussetts, London, England.

Appendix

For completeness, an alternative estimation technique, ordered probit, was also conducted, with the following results.

Ordered probit	regression				Number of obs	=	2008
					LR chi2(9)	=	129.16
					Prob > chi2	=	0.0000
Log likelihood	= -12451.205				Pseudo R2	=	0.0052

teachqual	Coef.	Std. Err.	z	P>z	[95% Conf.	Interval]
adminlight	.0268294	.020396	1.32	0.188	-.0131459	.0668047
research1	.0108844	.0019207	5.67	0.000	.0071199	.0146489
research1sq	-.0000865	.000015	-5.77	0.000	-.0001158	-.0000571
teachcours	.1581845	.0375258	4.22	0.000	.0846352	.2317337
teachcourssq	-.0116868	.0022763	-5.13	0.000	-.0161482	-.0072253
d14	.0912555	.0360453	2.53	0.011	.020608	.1619029
d06	.0290495	.0081167	3.58	0.000	.0131411	.0449579
DEP18	.4491291	.1005483	4.47	0.000	.2520579	.6462002
female	.1945427	.0468012	4.16	0.000	.1028139	.2862714

Although the results are not directly comparable, the estimated coefficients are of the same sign and similar in magnitude as in the panel equation estimated by GLS shown in the paper in section V. We cannot present tables of misclassification due to the large number of categories involved in the definition of the dependent variable, which takes values between 0 and 9 using two decimals. This estimation constitutes another test of robustness of the equation. In this case it is relative to an alternative estimation technique that considers the dependent variable as an ordinal variable rather than a cardinal variable.