ISSN 2042-2695

CENTRE *for* ECONOMIC PERFORMANCE

CEP Discussion Paper No 1386

November 2015

Multitask Agents and Incentives: The Case of Teaching and Research for University Professors

Marta De Philippis

THE LONDON SCHOOL OF ECONOMICS AND POLITICAL SCIENCE

E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL

Abstract
This paper evaluates the behavioural responses of multitask agents to the provision of incentives skewed towards one task only. In particular it studies the case of strong research incentives for university professors and it analyzes their effects on the way university faculty members allocate effort between teaching and quantity and quality of research and on the way they select into different types of universities. I first obtain different individual level measures of teaching and research performance. Then, I estimate a difference in difference model, exploiting a natural experiment that took place at Bocconi University, which heavily strengthened incentives towards research in 2005. I find evidence that teaching and research efforts are substitutable in the professors' cost function: the impact of research incentives is positive on research activity and negative on teaching performance. The effects are driven by career concerns rather than by the monetary incentives and are stronger for low ability researchers. Moreover, under the new incentive regime lower ability researchers tend to leave the university. Since I estimate that teaching and research ability are positively correlated, this implies that also bad teachers tend to leave the university. These results are consistent with a model of incentives where agents allocate effort between two substitute tasks and ability is multidimensional.

Keywords: Multitasking, incentives, teaching
JEL codes: I2; J41; M5

This paper was produced as part of the Centre's Education Programme. The Centre for Economic Performance is financed by the Economic and Social Research Council.

I thank Steve Pischke for very precious guidance, supervision and encouragement. I thank Marco Agliati, Esteban Aucejo, Oriana Bandiera, Marco Bertoni, Tito Boeri, Alessandra Casarico, Stephan Maurer, Michele Pellizzari, Giovanni Pica, Alfonso Rosolia, Stefano Verzillo and Giulia Zane and participants to the LSE work in progress seminar and labour workshop, the second fRDB workshop and the XIII Brucchi Luchino workshop for providing me with very useful comments and information. Finally I am indebted to Mariele Chiruli, Enrica Greggio, Erika Palazzo, Cherubino Profeta and Gianluca Tarasconi for precious help and information on the data.
Marta De Philippis, London School of Economics and Centre for Economic Performance and Bank of Italy.

Published by
Centre for Economic Performance
London School of Economics and Political Science
Houghton Street
London WC2A 2AE

1 Introduction

The study of principal-agents relationships and the design of the optimal incentive provision systems have a long tradition in economics. A particularly complex and very common situation arises when agents have to allocate their time and effort among different tasks. In this case, the provision of incentives on one task only may distort multitask agents' behavior: individuals may respond by increasing effort in the activities subject to incentives, crowding out time and energy from other uses. This is especially the case if performance in other tasks is not easy to measure, and if there are no other reasons, such as social pressure or intrinsic motivation, to perform them in any case [Holmstrom and Milgrom, 1991, 1994, Brüggen and Moers, 2007, Fehr and Fischbacher, 2002, Bandiera et al., 2010, 2007, Benabou and Tirole, 2003, Prendergast, 2008].

While the theory related to multitask agents is very well-developed, starting from the seminal work by Holmstrom and Milgrom [1991], empirical tests to the size and the sign of the behavioural responses predicted by this type of models are difficult to implement because of very heavy data requirements, first of all the need of an individual measure of performance for each task, that is often not easily observable. The empirical literature is therefore very scarce and the actual economic cost of standard incentives for multitasks agents is still largely unknown. In practice, it depends on how the different tasks interact in the agent's production and cost functions and it is therefore specific to the actual tasks taken into consideration.

This paper analyzes one of the leading examples of multitask agents: the case of university professors. Faculty members allocate time among many activities, mostly teaching and research. Incentives in most countries are however strongly skewed towards research: the 'Publish or Perish' paradigm is the most popular criterion for faculty hiring and promotion decisions in universities.This paper analyzes the overall consequences of strong research incentives on teaching and research outcomes. It evaluates, first, the direct impact of research incentives on research performance itself and, second, it studies the indirect effect of research incentives on teaching quality. Moreover, in order to understand the overall impact on teaching and research performance, it analyzes how the composition of professors changes under an incentive scheme strongly skewed towards research. Finally, by analyzing the correlation between teaching and research, it evaluates what may be the costs and benefits of separating teaching and research careers for university professors.

Using a standard model of incentives where agents allocate effort between two different tasks and ability is multidimentional, I show that the effect of stronger research incentives on teaching and sorting of professors depends on two main parameters: on whether teaching and research are substitute or complement in the professors' cost function and on whether teaching and research ability are correlated (i.e. whether good researchers are also good teachers). I then estimates the sign of these parameters. I overcome many of the standard identification issues by studying the case of Bocconi University, an Italian private institution of tertiary education based in Milan. Its institutional setting provides a unique opportunity to test the effect of research-oriented incentives on teachers' allocation of effort between multiple activities and the overall effect on the university's teaching and research outcomes.

Three features of Bocconi's institutional setting are crucial for my analysis. First, I can construct a measure of teaching performance using a value added approach that is the standard one used to evaluate teachers in primary or secondary schools [Rothstein, 2010, Rockoff, 2004, Aaronson et al., 2007, Rivkin et al., 2005]. It is usually impossible to apply this method to universities because students self-select into courses, exams and teachers therefore, while the usual assumption that, conditional on previous test scores, allocation is random is credible in primary and secondary schools, it does not hold in the university context. Bocconi students instead are randomly assigned to teachers in each academic year: within a degree program, if the number of enrolled students requires it, students are randomly split in different classes, each of which taught by a different lecturer, while the exam, the syllabus and the type of classrooms are identical for all students. Therefore, I can compare the average class grade of students taught by different professors teaching the same course.[1] Second, Bocconi sharply changed its faculty's incentive regime in 2005, shifting the focus explicitly towards research, by strengthening research requirements for promotion decisions and introducing monetary incentives based on quality and quantity of publications. Third, the large heterogeneity of Bocconi teaching faculty's contracts provides a natural control group: many teachers are not fully hired by Bocconi and act only as external teaching faculty. They have the same teaching responsibilities but are not subject to Bocconi changes in promotion strategy and incentives.

This paper therefore estimates a difference in difference equation, evaluating teachers' performance before and after 2005 and using external teachers as control group. Finally, I check the robustness of my results by using two alternative control groups: (i) faculty members who became tenured just before 2005, and are therefore not exposed anymore to career concerns and (ii) faculty members of another Italian university (Bologna) very similar to Bocconi in terms of quality and quantity of research.[2]

My main results are as follows. First, I find that the new incentive regime improved both the quality and the quantity of published papers. After the change in the incentive scheme Bocconi faculty members started to publish, on average, 25% more papers than before. Moreover, the effect is mostly driven by young faculty members, whose career concerns are stronger since they are not tenured yet. Both the magnitude and the sign of this result are perfectly in line with the literature on piece rate incentives [Lazear, 2000]. Second, the introduction of incentives towards research had a negative impact on teaching performance, as measured by time-varying teacher fixed effect. In particular teaching quality decreased by 7% of a standard deviation under the new incentive regime. The effect is, again, mostly driven by young faculty members and more negative for students at the bottom of the ability distribution. Combining the two estimates on teaching

[1]In particular I estimate time varying teachers fixed effects, controlling for yearly shocks at the couse level-such as shocks to the exam papers or to the syllabus. In principle I do not need to control for students' characteristics such as previous test scores, because of randomization of students across classes within the same course: if the number of enrolled students requires it, students are randomly split in different classes, each of which taught by a different lecturer, while the exam, the syllabus and the type of rooms are identical across classes.

[2]This second strategy can only be applied to Research outcomes, because I do not have information on teaching performance for the university of Bologna. I chose Bologna, because in terms on quality of research as evaluated by the Italian Institute of University Research Evaluation (ANVUR) it is the most similar to Bocconi University, in terms of dimension of the department and quality of the research outcome between 2004-2010. www.anvur.org/rapporto/files/Area13/VQR2004-2010_Area13_Tabelle.pdf

and research I find that, overall, one extra publication reduces teaching quality by one third of a standard deviation. This suggests that, at least for the type of courses I am considering, teaching and research are substitutes, not complement in the teachers' cost functions. Third, I find evidence of some positive sorting effects: the new incentive scheme induced low ability researcher to leave. Forth, I document that teaching and research abilities are positively correlated, this implies that if a university manages to attract good researchers, it will also attract good teachers. The overall effect on teaching quality is therefore ambiguous: on one side, since teaching and research effort are substitutes, teaching quality of incumbents decreases, on the other side the policy pushed away the worst researchers and, since research and teaching ability are positively correlated, also the worst teachers.

This paper fits into the literature that investigates behavioral responses to incentives, in particular in the context of multitask agents. As mentioned before, there is little empirical evidence of the actual cost of not optimally designed incentive schemes for multitask agents, mostly because of data limitations and because performance in many tasks is difficult to measure, for instance because it refers to components that are not observable or because it is difficult to disentangle the individual contribution to the final outcome. Few exceptions, that usually analyze the quantity-quality trade off for the same activity, are Dumont et al. [2008], Feng Lu [2012], Hong et al. [2013], Johnson and Reiley [forthcoming]. In the education literature Jacob [2005], Fryer and Holden [2013] analyze the impact of accountability policies on test-specific skills and students' effort in high-stake versus low-stake exams.

My paper contributes to the incentive literature, first, by providing a well-identified estimate of how multiple tasks interact in the agents' cost function. While most of the existing papers look at the quality-quantity trade off of performing the same activity, I analyze the effect on the performance in two different activities, when it is not clear a priori whether the tasks are substitute or complement in the agents' cost function. Second, to my knowledge this is the first paper that combines estimates of the effort substitution effect with an analysis of how multitask agents sort in different types of firms, depending on the incentive schemes. This is key in order to evaluate the overall effect for the principal on each task. Sorting effects may be very relevant and may countervail the direct effort substitution effect so to revert the sign of the overall impact of changes in the incentive scheme. Third, I am able to disentangle the pure effect of monetary incentives from the effect generated also by career concerns: this is extremely useful in order to understand the main drivers behind different responses and to be able to efficiently reproduce the effects to other settings.

My paper is also related to the rather thin education literature on teachers' contracts and incentives. Some papers evaluate the effect of teaching contracts on teaching performances [Figlio et al., 2013, Bettinger and Terry, 2010] and find that students learn more from non-tenure line professors. Since non-tenure line faculty is less focused on research, this may suggest that their results are driven by differences in teachers' incentive schemes. Still, it is impossible from these analyses to disentangle whether the effect they find is instead driven by selection into non-tenure line jobs. For what concerns incentives, two papers look at the trade-off between teaching and research, by analyzing the effect of increased teaching incentives on research and teaching outcomes. Brickley and Zimmerman [2001] use a single difference strategy to study the consequences of the introduction of teaching performance incentives at the University of Rochester Business School. The authors find a substantial and almost immediate jump in teaching ratings, measured by students' evaluations, and a corresponding decline in research output. Payne and Roberts [2010] analyze this same issue but using between, not within, university variation. They exploit US state variation in the adoption of teaching performance measures and find that research activity decreased in quantity but improved in quality in non-flagship universities.

This paper contributes to the education literature in two ways. First, it is the only one, to my knowledge, to test the other side of the relationship between teaching and research: the effect of strong research incentives. This type of analysis is crucial given the extremely wide adoption of research incentives in universities. Moreover, it is likely that the extent of effort reallocation generated by research incentives is larger than for teaching incentives because teaching effort is more difficult to measure and monitor and peer pressure on excellence in teaching is much weaker than in research. Second, this paper provides the first piece of evidence on the sign of the correlation between teaching and research ability. The positive correlation between teaching and research has important implications for the design of professors' incentives and hiring schemes. For example, policies aimed at increasing teachers' specialization that propose to dedicate part of the faculty exclusively to teaching and part of it exclusively to research, should take into consideration that good researchers are also good teachers, on average.

The structure of the paper is as follows. Section 2 provides a simple conceptual framework that rationalize expected results; Section 3 outlines the identification strategy; Section 4 describes the data and the institutional setting; Section 5 presents my empirical results and Section 6 shows how my results are robust to alternative control groups. Finally, Section 7 briefly characterizes the policy implications of my results and concludes.

2 Conceptual Framework

This section presents a very simple framework with the aim of organizing and rationalizing expected findings. The working of the model in the spirit of Holmstrom and Milgrom [1991] and it is similar to the model presented by Fryer and Holden [2013].[3]

An agent, upon accepting the contract, takes two non-verifiable actions e_r and e_t, which I call research and teaching effort respectively. Each action takes values in $\mathbb{R}_+$ and generates a performance measure $m_i = \alpha_i e_i$ where $i = r, t$ and α_i is unknown to the principal. I refer at α_i as the type of the agent on task i (her ability level).

I assume that the principal offers a linear incentive scheme of the form $x = s + b_r m_r + b_t m_t$. If the agent accepts, she makes her effort choices, the performance measure is realized and the principal pays the agent accordingly.

[3] I will not model why the university decided to increase research incentives, i.e. I do not make assumptions on the university objective function, I only analyze what are the agents' responses to an increase in research incentive, in the spirit of Lazear [2000].

I also assume that the agent's preferences can be represented by the following CARA utility function:

$$u(x,e) = -exp[-\eta(x - \frac{1}{2}(e_r^2 + e_t^2) - \delta e_r e_t)] \tag{1}$$

where x is the monetary payment and δ is the degree of substitutability between the tasks r and t in the cost function ($0 < \delta < 1$). Let $\underline{U}$ be the agent's outside option if he does not work. Moreover, I assume that there is a minimum teaching performance $\underline{m}_t$ and research performance $\underline{m}_r$ required by the university.

The agent therefore maximizes utility with respect to e_r and e_t, subject to the participation constraints ($u(x,e) > \underline{U}$ and $m_r > \underline{m}_r$ and $m_t > \underline{m}_t$). Note that when $m_r^* < \underline{m}_r$ or $m_t^* < \underline{m}_t$ each individual will choose whether to stay and exert effort level $\underline{e}$ or to leave, depending on whether $U(x_{\underline{m}}, e_{\underline{m}})$ is larger or smaller than $\underline{U}$.[4] If it is smaller, she will decide to leave (or be fired). Otherwise, she will be induced to exert more effort, even if very costly, in order to stay in the university.

The new incentive scheme, that took place at Bocconi in 2005 as I will describe with more details in Section 3, implied an increase in b_r, the monetary return to research activity, and in $\underline{m}_r$, the minimum research performance required, but only for professors not tenured yet. Changes in b_r act mostly on the intensive margin (how much research effort to exert), changes in $\underline{m}_r$ instead mostly affect decisions also on the extensive margin (whether to stay in university or not).

2.1 Effects on teaching and research performances

This section shows what happens to m_r^* and m_t^* (and therefore e_r^* and e_t^*) if the university increases b_r and $\underline{m}_r$ and professors stay in the university.

In appendix A I solve the model (for internal solutions) and I show that the equilibrium effort level is:

$$e_r^* = \frac{b_r\alpha_r - \delta b_t\alpha_t}{1-\delta^2}; \quad e_t^* = \frac{b_t\alpha_t - \delta b_r\alpha_r}{1-\delta^2} \tag{2}$$

It is clear that e_r^* increases if b_r increases, while the sign of the derivative of e_t^* with respect to b_r depends on the sign of δ.

Proposition 1 *An increase of b_r, the marginal return on research performance, leads to an increase in e_r.*
The response of e_t depends on the value of δ: $\begin{cases} \frac{\partial e_t}{\partial b_r} < 0 & if \delta > 0 \text{ } (e_r \text{ and } e_t \text{ substitute}) \\ \frac{\partial e_t}{\partial b_r} > 0 & if \delta > 0 \text{ } (e_r \text{ and } e_t \text{ complement}) \end{cases}$

The policy, moreover, increased $\underline{m}_r$.

Proposition 2 *When $m_r^* > \underline{m'}_r$: an increase in $\underline{m}_r$ does not have any effect.*
When $m_r^* < \underline{m'}_r$ and $U(x_{\underline{m'}_r}, e_{\underline{m'}_r}) > \underline{U}$, $\underline{m}_r$ and/or $\underline{m}_t$ are binding and professors exert $e_{\underline{m}_r,r}$ or $e_{\underline{m}_t,t}$ even if above their optimal level.

[4] $e_{\underline{m}}$ and $U(x_{\underline{m}}, e_{\underline{m}})$ are respectively the effort need to exert in order to obtain $\underline{m}$ and the utility level when $\underline{m}_r$ and or $\underline{m}_t$ are binding.

2.2 Sorting effects

Whether agents will decide to continue working under the new regime or tp leave, depends on $\underline{U}$, the utility provided by leisure.

Increases in b_r, do not have any effect on the decision to continue working because, at most, the agents will not change their behaviour. Increases in $\underline{m}_r$, instead, may have effects on the decision to stop working.

Proposition 3 *If $m_r^* < \underline{m'}_r$ and $U(x_{\underline{m'}_r}, e_{\underline{m'}_r}) < \underline{U}$, professors will leave the university and enjoy utility $\underline{U}$*

Therefore, overall, for individuals whose $m_r^* > \underline{m}'_r$, the effect of the policy comes entirely from variations in b_r and therefore from evaluating the sign of the derivatives of e_r^* and e_t^* with respect to b_r.

For individuals whose $m_r^* < \underline{m}'_r$, the effect depends on whether $U(x_{\underline{m}}, e_{\underline{m}})$ under the new $\underline{m'_r}$ is larger or smaller than $\underline{U}$. If it is smaller, again, they will decide to leave and exert no effort. Otherwise, they will be induced to exert more research effort, even if very costly, and stay in the university.

I now evaluate how this effect varies by agent's ability. It is important to keep in mind that $\frac{\partial e_r}{\partial \alpha_r}|_{m_r=\overline{m}} < 0$: research effort is more costly for low α_r individuals. An increase in research incentives, therefore will be much more beneficial for high ability researchers. Instead, those more likely to leave because of an increase in $\underline{m}_r$ are low ability researchers.

Proposition 4 *When $m_r^* > \underline{m}_r$: an increase in b_r, leads to a larger increase in e_r for individuals with high α_r and to a larger response of e_t for individuals with low α_r.*
For low α_r agents, it is more likely that $\underline{m}'_r$ is binding, and are therefore induced to leave.

The predicted response of stronger b_r along the distribution of α_r is therefore that: (i) for teaching, the effort substitution effect is stronger for low ability researchers (as long as $\delta > 0$); (ii) for research the effect is instead U shaped. Very low ability researchers will leave the university; of those staying, the lowest ability ones (those whose $m_r^* < \underline{m}'_r$) will increase effort on research in order to reach $\underline{m}'_r$; the others (those whose $m_r^* > \underline{m}'_r$), will increase e_r proportionally with their ability α_r.

3 Empirical strategy

This section develops my empirical strategy, aimed at estimating the causal effect of increasing incentives towards publishing on teaching and research performance.

I use administrative data from Bocconi university archives to estimate two Difference-in-Difference models, one for teaching and one for research, exploiting the sharp change in Bocconi research incentives and using external faculty as control group.

I begin this section by describing in more details the reform in Bocconi's incentives regime announced in July 2005 (Section 3.1). Sections 3.2 and 3.3 present my empirical model for the evaluation of the effect on teaching performance and on research activity respectively. Finally, section 3.4 describes how I estimate sorting effects.

3.1 The new incentive policy

In 2005, Bocconi University unexpectedly announced the adoption of a new policy of hirings and promotions. The Board of Directors called for the Rector to make Bocconi University one of the top five universities in Europe. As a consequence, the old hiring and promotion strategies, mainly based on national competitions and seniority, were replaced with new practices based on international standards. Since then[5], an independent committee, composed of faculty members from all disciplines, has been in charge of recruiting and promotions. Decisions have been centralized at the university level, making exceptions impossible. Moreover, the importance of research outcomes in promotion decisions was clearly stated in all internal faculty contracts.

The goals of the New Strategic Plan, as announced in July 2005, were the following: (i) recruiting at least 50% of new faculty on the international job market; (ii) improving the systems to evaluate research produced by each professor (through the creation of an independent evaluation committee and the internationalization of evaluation criteria); (iii) adopting clear incentives on research (both monetary[6] and career-based); (iv) creating mechanisms to "attract and keep the best researchers worldwide".

The focus switched explicitly towards research, tenure decisions started to be based almost entirely on scientific productivity and the requirements on quantity and quality of research started to be much tighter.

3.2 Research Performance

I first evaluate whether incentives on publishing have an impact on research quality or quantity.

I use three different measures of research performance: (i) the number of publications; (ii) a proxy of the index actually used by Bocconi to evaluate teachers (which is computed as the sum of the number of articles published by each teacher, weighted by the quality of the journals as classified by Bocconi[7], divided by the number of coauthors) and (iii) the number of working papers and published papers (from Google Scholar).

I collect publication data from the Web of Science website. In particular, I count professors' yearly publications in the fields categorized by Web of Science as 'business', 'maths' and 'economics'. Unfortunately, for less recent years, the Web of Science database only reports the author's first name initial and not the full name. As such, I run a search only using the authors' first name initial, together with their surname.

[5]The actual implementation of the policy was in 2007, but throughout the analysis I will consider the year of the announcement, 2005, as the treatment year. Be aware that the full effect will be in place starting from 2007.

[6]Even if previously anticipated, Bocconi started to actually provide monetary incentives to its internal faculty in the academic year 2008. In particular there are three types of incentives: (i) the possibility of getting "research profile", with less teaching duties; (ii) research premia that depend on the number and the quality of publications; (iii) research funds, given to everybody who has reached a minimum level of research productivity in the previous two years. Publications were weighted depending on the quality of the journal)

[7]Bocconi divides journals into 3 categories: A+ journals (i.e. Econometrica), to which it assigns a weight of 15; A journals (i.e. Economic Journal), weighted 7; B journals (i.e. Economic Letters), weighted 3. I classified journals using the list valid for the year 2007, available upon request.

I use Google Scholar as a source for the number of working papers. In particular, I use a web scraping program which makes automatic searches (one for each year/professor combination) from the Google Scholar website. I restrict my research on the Google Scholar website to the following fields: 'social sciences, arts, and humanities' and 'business, administration, finance, and economics'. In this case, data on full names are available for all years. I thus look for full names.[8]

I then implement a Difference-in-Difference model by estimating the following equation for the years between 2001 and 2010:

$$pub_{pt} = \theta_t + \theta_p + \gamma_{res}(internal_p * post2006_t) + \gamma_4 X_{pt} + \eta_{pt} \tag{3}$$

where pub_{pt} are publications of professor p in year t; $internal_p$ is the internal status (in 2005); θ_t are time fixed effects; θ_p are teacher fixed effects; X_{pt} are teacher characteristics (age, age squared) and η_{pt} is the error term. I cluster standard errors by professor.

For sake of consistency, I include only teachers who were teaching classes I can use to estimate the teaching equation (see below equation 5).[9] Moreover, in order to exclude endogenous status switches from internal to external or viceversa after the introduction of the policy, I classify teachers as internal if they were internal in 2005. In my robustness checks (Section 5.3) I check my results are not driven by this choice, by running the same analysis using contemporaneous status instead of status before 2006 as treatment, therefore including endogenous 'switches' in the effect. Moreover, I drop internal lecturers. Lecturers are internal professors (fully hired by Bocconi) but with only teaching duties.[10] On one side, monetary research incentives are not provided to lecturers but, on the other side, the way lectureship decisions are taken has probably changed after 2006. They therefore do not represent a good control group. In a robustness check (Section 5.3), I include lecturers and interact them with the treatment. Fibally, I drop law professors and law courses: law's exams are usually oral exams so the set of questions is not the same for all students. It is therefore difficult to use average grade as a measure of teaching quality.

3.3 Teaching Performance

Second, I estimate my empirical model for the effect on teaching in two steps.

The first step uses microdata from the student academic curriculum database and it is aimed at computing the average grade at the class level, conditional on students' high

[8] This procedure does not eliminate the possibility that the same working paper is counted more than once, if published in two different versions. However, this is still a measure of the effort one puts in that specific research. Moreover, this measure also contains the published version of the working papers. Accessed in Dec 2011.

[9] The difference in the number of observations is given by those teachers who were teaching more than one class per year or by the fact that some teachers do not teach compulsory undergraduate courses all years, but I still include those year observations in my analysis, for consistency over time.

[10] The difference between the position of lecturers and assistant/full professors is clear from how their contracts. The contract for assistant professors states "responsibilities include teaching and, most importantly, productivity in research". The contract for lecturers, instead, states that only teaching duties are expected from lecturers. Research activity is not even mentioned.

school final score and demographics.[11] Students taking the same course are all taught the same syllabus and are all examined on the same questions, independently of the class to which they are (randomly) assigned. Some variations in the material and in the exam across degree programs are allowed (this is why I correct for the full interaction of courses, degree programs and years). Usually a senior member of the faculty acts as the course coordinator: he establishes the material to teach, manages possible complications and prepares the exam paper. Grading is instead generally delegated to the individual teachers, who typically are supported in the marking by teaching assistants.

I estimate the following equation:

$$grade_{ipct} = \beta_0 + \beta_1 HSgrade_i + \beta_2 X_i + \alpha_{ptc} + u_{ipct} \tag{4}$$

where $grade_{ipct}$ is the grade obtained by student i, with teacher p[12], in year t, in course c (standardized at the course-year level to have mean 0 and standard deviation 1); $HSgrade_i$ is student i high school final grade; X_i are the students' individual characteristics (gender, age, whether Italian, whether from Milan, type of high school attended). u_{ipct} is the error term. α_{pct}, the year specific teacher fixed effect, is my parameter of interest.

The second step evaluates how the teacher fixed effects α_{ptc} evolve over time, in response to the change in incentive regime. I implement the same Difference-in-Difference estimation as in Section 3.2, changing the dependent variable. In particular, I estimate the following equation:

$$\widehat{\alpha_{ptc}} = \delta_p + \delta_{tc} + \gamma_{teach}(internal_p * post2006_t) + \gamma_2 X_{pt} + \epsilon_{ptc} \tag{5}$$

where $internal_p$ is a dummy equal to one if the professor was internal before the change in incentives; $post2006_t$ is a post reform dummy; δ_p are teacher fixed effects[13]; δ_{tc} are fixed effects for the full interaction between academic years, courses and degree programs[14]; X_{pt} are time-varying professor characteristics (age, age squared, experience in teaching undergraduate courses in Bocconi) and ϵ_{ptc} is the error term. I cluster standard errors by professor.

γ_{teach} quantifies the change in teaching performance of incumbent professors under the new incentive scheme more focussed towards research.

The economics literature usually measures teacher quality by estimating a teacher fixed effect in equation 4. Here, differently from most of the previous analyses, I allow teacher effects to vary over time and I analyze how they change in response to the positive shock in research activity.

[11] To reduce computational burden, I exploit randomization of students to teachers and I do not include students fixed effects.

[12] Since in around 40% of the cases more than 1 professor teaches the same class the actual meaning of p in this first case is the "professor mix" of the class.

[13] Notice that in this case p represents a single teacher. Therefore if a class was taught by multiple teachers I impute the (unique) class fixed effect to both teachers.

[14] Courses may have the same code but programs and exams may be different for different degree programs. Interacting also with degree programs allows me to exploit variation across teachers' performance when syllabus and exam papers are exactly the same (and over which the randomization of students to teachers takes place).

I overcome many of the standard identification problems because: (i) I eliminate concerns related to time constant factors by including teachers' fixed effects in my regressions. I only analyze how teaching performance evolves over time; (ii) the presence of a unique final exam and the randomization of students to teachers eliminates concerns on time varying endogenous matching. There has been a debate (Rothstein [2010, 2009], Ishii and Rivkin [2009], Kane and Staiger [2008], Chetty et al. [2013]) about whether value-added models perform weakly in the absence of randomization. Teachers fixed effects may also identify endogenous matching between teachers and students. Results are mixed. Most recently Kane and Staiger [2008], Chetty et al. [2013] use primary school data to show that this problem can be eliminated controlling for previous year test score. However, the problem of endogenous matching is likely to be much worse in the university context, where students self-select into courses and therefore teachers.

Finally, I estimate the same effect running the analysis directly at the student level. I therefore estimate the following equation:

$$grade_{iptc} = \zeta_p + \zeta_{tc} + \zeta_{teach}(internal_p * post2006_t) + \zeta_2 X_{ipt} + v_{iptc} \tag{6}$$

where all the variables are defined as before and v_{iptc} is the error term.

While my preferred specification is the estimation of equation 5, because it is more easily interpretable as changes in teaching quality, this last specification will allow me to evaluate how the main effect is heterogeneous with respect to students' characteristics, in particular with respect to students' ability, measured by their final high school grade.

3.4 Sorting Patterns

To have a complete picture of the overall effect of the change in incentives on research and teaching quality, I analyze how the composition of workers changed after the new regime was introduced. As shown in Section 2, the change in minimum research requirements should push low ability researchers away. Whether this translates into maintaining also better teachers, it will depend on how teaching and research ability are correlated.

I analyze selection effects in two ways: first, I compare estimates with and without professors' fixed effects; second, I obtain direct estimates of the underlying teaching and research abilities and I analyze how the ability composition of teachers varies over time, looking both at teachers sorting in and sorting out.

In order to analyze sorting patterns, I need estimates of teaching and research ability. I obtain these estimates estimating professors' fixed effects from the following equations. For teaching:

$$\widehat{\alpha_{ptc}} = \theta^t_p + \delta_{tc} + \gamma_2 Q_{pt} + \epsilon_{ptc} \tag{7}$$

where α_{ptc} is the conditional average grade of professor p, teaching course c in year t; δ_{tc} are fixed effects for every course-year; Q_{pt} are professor characteristics (age, age squared, years of experience at Bocconi); ϵ_{ptc} is an error term. Finally, θ^t_p are professor fixed effects, my estimate of underlying teaching ability.

Analogously, for research:

$$pub_{pt} = \theta^r_p + \zeta_t + \zeta_2 Q_{pt} + \eta_{pt} \tag{8}$$

where pub_{pt} is the number of papers published by professor p in year t; ζ_t are year fixed effects, that absorb any possible time trend in how difficult it is to publish papers over time; Q_{pt} are professor characteristics (age, age squared); η_{pt} is an error term. Again, θ_p^r are professor fixed effects, my estimate of underlying research ability.

One first concern may be that, since incentives are muted under the new scheme, it is not clear whether fixed effects based on teaching or research productivity after 2006 are a good proxy for ability. This would imply one should only use fixed effects evaluated before 2006. However, it would be impossible to test whether the new policy managed to attract high ability professors, since I would not be able to estimate a teacher fixed effect for faculty members who entered under the new incentive regime. In Figure 4, I follow Lazear [2000] and I show that, for professors who were teaching also before 2006, there is a strong positive correlation between fixed effects evaluated in the period before 2006 and those estimated for the period after 2006. Whenever it is possible (for sorting out effects), I will run my regressions also using fixed effects estimated on the pre-2006 period only.

Before showing the specifications, one caveat need to be borne in mind. First, for this analysis I do not run a proper diff-in-diff strategy because I do not know the entire employment history of external teachers. I only observe whether they were teaching undergraduate compulsory courses in each year between 2001 and 2011, but I do not observe their exact year of entry/exit. For internal professors, instead, I know exactly their year of entry, every change in their contracts and their year of exit, including the reason for leaving. Moreover, it is very unlikely that external teachers represent a good control group for the analysis on sorting: the way they are selected is very different from the selection process of internal faculty and it varies substantially depending on specific departments and academic years.

I evaluate how average teaching and research ability change, depending on the year teachers enterd/exited Bocconi.

For sorting out, I estimate the following equation:

$$\widehat{\theta_p^j} = \alpha_1 exitpost2006_p + \alpha_2 exitpre2006_p + \alpha_3 X_p + \delta_e + u_p \tag{9}$$

where: $j = r, t$; δ_e are year of entry fixed effects; $exitpost2006_p$ is a dummy equal to one if teacher p left Bocconi after 2005; $exitpre2006_p$ is a dummy equal one if professor p left Bocconi before 2005[15]; X_p are time-invariant professors' characteristics (age of entry, gender) and u_p is an error term. I only include teachers leaving Bocconi for reasons different from retirement.

Symmetrically, I obtain the effects on sorting in of teachers, by estimating the following equation:

$$\widehat{\theta_p^j} = \psi_1 entrypost2006_p + \psi_3 X_p + \psi_4 f(e) + \omega_p \tag{10}$$

where: $j = r, t$; $f(e)$ is a linear and squared trend for year of entry; $entrypost2006_p$ is a dummy equal to one if teacher p entered after 2005; X_p are time-invariant professors' characteristics (age of entry, gender) and ω_p is an error term. To make the two groups

[15]the omitted category are those staying

of teachers more comparable, I estimate equation 10 only for teachers who entered after 2000.

4 Data and descriptive statistics

4.1 Students

This paper uses the administrative records of individual students and teachers from Bocconi University, an Italian private institution of tertiary education based in Milan. Bocconi offers degree programs in Economics, Management and Law. I only consider compulsory undergraduate courses between 2001 and 2011. My sample includes around 700 teachers and 30,000 students, who take on average 20 compulsory exams over the 3 years of study.

My data cover in detail the entire academic history of students, including their basic demographics (gender, place of residence and place of birth), high school leaving grades as well as high school type (whether focusing on humanities, on sciences or technical/vocational subjects). Information is also provided on the grades in each single exam together with the date when the exams were sat. Moreover, I have access to the random class identifiers of students, which allows me to determine in which class each student attended each course.[16]

Table 1 reports descriptive statistics for students. Most of the students are Italian, one fourth is from Milan. They are positively selected among the population of high school graduates: the average high school final grade is very high (0.9 out of a maximum of 1[17]). On average there are 5 classes per course, of about 110 students each, and 20 compulsory undergraduate courses per year. Each student sits on average 7 exams per year. The degree program in Management is the one with the highest number of classes (7 on average).

4.2 Teachers

Together with student data, I have access to administrative data on Bocconi faculty. In particular, I have information on teachers' demographics (date of birth, gender, full name), type of contract, department of affiliation and number of teaching hours in each course and class. I am therefore able to match students with teachers.

I classify each teacher as internal or external. Table 2 lists all different teaching contracts available at Bocconi over the years I consider and the way I group them into five categories: assistant professors-junior researchers, associate professors, full professors, non academics and professors from other universities. I define teachers in the first three

[16] Students students who did not sit the exam in the academic year they were supposed to, are randomly allocated to a new class and the records on the initial class allocation are overwritten in the administrative database. I therefore include them in the new class, including a dummy equal to one if the student took the exam in a different year from what expected. However, this is a very small group (about 3% of students).

[17] Given that I know the maximum final high school grade each foreign student can take, I standardize high school final grades of foreign students to be between 0.6 and 1, so that they are comparable with grades of Italian students.

categories as internal, treated by incentives, and teachers in the last two categories as external, my control group.

Table 3 reports descriptive statistics of teachers. Column (1) reports descriptives for internal teachers, column (2) for external teachers and columns (3) reports the difference of the two groups. In total, in my sample, I observe 681 teachers for 5 years on average. Internal teachers tend to be slightly older and to teach more hours at Bocconi. Most teachers are hired by the Management or Economics department. Finally, assistant professors and external faculty members represent 30% of the sample each, while associate and full professors represent around 15% each.

4.3 Students Teachers Randomization

The randomization of students to teachers is performed every year via a simple random algorithm that assigns a class identifier to each student, within each degree program[18]. Table 4 provides evidence that teachers were actually randomized to students. Following Braga et al. [2011] I show results of a regression of class (student) average characteristics on teacher characteristics and dummies for the full interaction of courses, academic years and degree programs.[19] The null hypothesis under consideration is the joint significance of the coefficients on teacher characteristics. The F statistics are always very low, suggesting there is no significant correlation between students' and teachers' characteristics.

5 Empirical results

5.1 Results for Research

The sign of the effect on research is expected, from Section 2, to be positive and stronger for young professors not tenured yet, since they are affected both by the monetary incentives and to progress in their careers.

Table 5 shows some descriptive statistics for the number and the quality of publications and working papers for internal and external teachers before and after 2006.

The first panel analyzes the total number of publications (books or journal articles) of professor p in year t, as collected from the Web of Science database. The second panel looks at the number of publications, weighted by the importance they have in terms of Bocconi's new incentive regime. This allows me to evaluate quality as well as quantity of research production.[20] Finally, the third panel evaluates the effect on the number of working papers from Google Scholar.[21] The first column reports the mean and the standard deviation of publications for internal and external teachers. The second and the third columns break down the number of publications for the period before and after 2006. Finally, the number in the bottom-right corner represents the simple

[18]The university administration adopted the policy of repeating the randomization for each course with the explicit purpose of encouraging wide interactions among the students.

[19]This is the level at which randomization takes place

[20]This variables moreover eliminates much more the problem of homonymity because all journals where Bocconi faculty publishes should be inserted in the list. Therefore only homonymous people in exactly the same sub-filed may be considered.

[21]Accessed in july 2011.

difference in difference, without any control. Standard errors, clustered at the teacher level, are reported in parenthesis. The Table shows that the number and the quality of publications increased after 2006 and they increased much more for internal professors than for external professors.

Table 6 shows results from equation 3, using the same three dependent variables.

Columns (1) and (2) report estimates without teacher fixed effects. The effect is positive and significant in all three panels. Once I include teacher fixed effects (columns (3) and (4)) the effect is still positive and significant. After the introduction of research incentives, the number of publications increased by 0.14 (36% over the mean) for internal faculty and the index used by Bocconi to evaluate teachers increased by 0.13 (16% over the mean). The number of working papers of internal professors is also 0.15 (10% over the mean) higher than it would have been otherwise. Moreover, while columns (1) and (3) look at the aggregate effect, columns (2) and (4) separately evaluate the effect for for assistant professors and associate professors (which I call junior faculty) and full professors. The aggregate effect is mostly driven by junior faculty, as their career concerns are stronger. The magnitude of the effects I estimate is in line with what is found and predicted by the general incentive theory [Prendergast, 1999, Lazear, 2000].

Finally columns (5) and (6) report results from estimating equation 3, using as dependent variable the square root of the number of publication. This is to try tackle simultaneously the presence of possible outliers and of a lot of zeros.[22]

Figure 1 displays the evolution of the difference in average number of publications between internal and external faculty.[23] The dotted lines refer to the 10% confidence interval boundaries. While the difference is rather stable before 2005, it gets larger after the introduction of research incentives. Moreover, given the long time needed to publish papers in most disciplines, after 2006 there is a clear change in trends but there is not a sharp jump.

5.2 Results for Teaching

As shown in Section 2, the sign of the effect of stronger research incentives on teaching quality depends on whether teaching and research efforts are complements or substitutes in the professors' cost function (δ smaller or larger than 0 respectively). The effect moreover is expected to be stronger for junior professors, exposed both to the change in monetary incentives and to the change in the minimum number of publications required.

Table 7 presents the results obtained from estimating equation 4. Exam grades are standardized to have mean 0 and standard deviation 1 within the same course-year.[24]

Results show that being male, with a higher final high school grade, Italian and from Milan is associated with higher university exam grades[25].

[22]Moreover I dropped the 5/1000 highest values for each dependent variable. It is very likely that most outliers are generated by homonymity.

[23]This graph plots the coefficient γ_t of the following equation:

$$pub_{pt} = \theta_t + \theta_p + \gamma_t(internal_p * \theta_t) + \gamma_4 Q_{pt} + \eta_{pt}$$

.

[24]This is in order to make the estimated α_{pct} comparable because not dependent on the difficulty of a particular exams.

[25]Grades in Italy go from 18 (pass) to 31 (excellence).

Table 8 reports some summary statistics of the estimated α_{pct} for internal and external teachers, before and after 2006. While before 2006, the teaching performance of the two groups was very similar, after 2006 it improved much more for external teachers than for internal teachers. Again, the bottom-right corner reports the diff in diff, without any control.[26]

Table 9 displays results from estimating equation 5. The first two columns show results without teacher fixed effects. Column (3) and (4) add teachers fixed effects. Teaching quality of internal teachers is 0.04 (around 7% of a standard deviation) lower after the change in incentives than it would have been otherwise. This suggests that teaching and research are substitutes in the professors' cost function. Again, the effect is stronger for young faculty members, more exposed to the policy.

Panel a and b of Figure 2 show the evolution of the different performance of external and internal teachers (panel a) and external and assistant professors (panel b) over time[27]. The difference is rather stable before the academic year 2005/2006 (named 2006 in the graph). Right after the adoption of the new incentive regime there is a drop in the quality of teaching for internal professors. In the following years, the performance is still slightly worse than before the reform, but better than in 2006. This may be because internal professors understood the consequences of their effort reallocation and partially readjusted their behaviour. Alternatively, they just started being more generous with their grading standards.

Table 10 reports results from the student level regression (equation 6).[28] As expected, results are very similar. What differentiates columns (1) and (2) of Table 10 from columns (3) and (4) of Table 9 is the way observations are weighted and coefficients should be interpreted. Table 10 implicitly weights observations by the number of students in each class: the coefficients should be interpreted as effects on average students' performance. Table 9 weights observations by teachers and the coefficients should be interpreted as effects of average teachers' performance.

Columns (3) and (4) of Table 10 explore whether the main results of Table 9 mask some important heterogeneity at the student level. I estimate equation 10, interacting the main effect with a proxy for students' ability. In particular I use high school final grade as proxy.[29] My omitted category are high ability students. Results show that the negative effect is mostly borne by low ability students.

This result suggests that there is room for policies aimed at matching professors to students in order to reduce the overall negative effect of stronger research incentives on teaching. This would mean in this case to match young researchers, more affected by

[26]notice that, because of some sampling error generated by the fact that $\hat{\alpha_{pct}}$ are estimated, the reported standard deviation may be larger than the standard deviation of the true α_{pct}.

[27]This is obtained by plotting the coefficients γ_t obtained from the following equation:

$$\alpha_{ptc} = \delta_p + \delta_{tc} + \gamma_t(internal_p * \delta_t) + \gamma_2 Q_{pt} + \gamma_3 Z_{pct} + \epsilon_{ptc}$$

Year 2001 (and the interaction between 2001 and internal) is omitted. The dotted lines refer to the 10% confidence interval bands.

[28]In this case, whenever a class was taught by more than one teacher, the observations for each student were doubled, such that each student was imputed to every teacher he was assigned to.

[29]I divide it into 3 categories: (i) high ability (omitted)= those students whose final high school grade was between 1 and 0.9; middle ability = between 0.8 and 0.9 and low ability: below 0.8.

the change in incentives, to higher ability students, who are less damaged by their lower teaching quality.

5.3 Robustness Checks for the effect on teaching

Table 11 presents a first set robustness checks for the estimation of the teaching equation. First, I estimate equation 5 excluding the academic years 2008/2009, 2009/2010 and 2010/2011. Starting from 2008/2009, internal faculty was exposed not only to research incentives, but also to teaching performance monetary awards. In particular, Bocconi University created a commission in charge of awarding a premium of 20,000 euros for the best 20 teachers who voluntarily apply. Decisions are based on students' evaluations. This new policy may attenuate the effect of research incentives (Holmstrom and Milgrom [1991, 1994]). Column (1) of Table 11 shows that results are almost unchanged. Second, in Column (2) I also include lecturers in my sample and I estimate a different treatment effect for lecturers. The effect on internal professors is similar. The effect on lecturers, even if not significant because of the small number of observations, is negative. Column (3) includes endogenous switches from internal to external status after the policy: it uses the contemporaneous status, not the status before 2006 as in Table 9, to define internal status. The control group includes in this case also, for instance, professors who switched from internal to external as a consequence of the policy. The coefficient is still negative and significant, but the magnitude is smaller. This means that Bocconi promotions from external to internal and viceversa where positively correlated with teaching quality. In column (4) I weight my regression by the number of hours taught by each professor in each class. Results are very similar.

I now discuss three possible confounding factors, that may undermine my identification strategy. The first is that students might not comply with the random class assignment and they might endogenously decide to attend classes with different lecturers. For example, they may match to the best professors, or attend classes with their closest friends. Unfortunately, I do not have any direct information on these unofficial switches of classes.[30] Braga et al. [2011] analyze whether the direction of class switches at Bocconi University is correlated with professors' ability. They use data on students' answer to an item in the student evaluation forms asking them about the level of congestion in their classroom. They estimate the degree of class switches as the difference in congestion level between the most congested and the least congested classes for each course. They find that, overall, course switching is not related to teacher effectiveness in any direction. Therefore, if the process of class switching is unrelated to teachers or students quality, then it will just affect the precision of my estimated class effects. Moreover, if the process is constant over time, the effect will go away with professors' fixed effects. Finally, even if course switching does affect my results, it would probably bias them against finding a negative effect on teaching performance. It is likely that students, if anything, will react by attending classes with the best teachers, who after the change of incentives will more likely be external faculty members. This would reduce the negative effect of the incentive policy on teaching.

[30] Bocconi decided to hand in evaluation forms to a subsample of professors only exactly in the year 2005 and 2006, making it impossible to look at students' evaluations for the period I am interested in.

Another concern is that teachers may change the way they grade students' exams as an effect of observing worse performances of their students. There is not a common rule on how exams are graded in Bocconi: in some cases exam papers are randomly given to class teachers to be graded, in some other cases each professor is in charge of grading his own group. I do not have information on how exam papers are actually graded in each course. In columns (1) and (2) of Table 12 I look at the effect on teaching quality for exams that are more objectively-graded, such as math, statistics or quantitative finance. Results show that, even if the effect is slightly smaller and less precise for this types of courses, it remains negative.[31] Moreover, again, if anything, I expect internal teachers to start being more lenient towards their students, therefore I expect this type of bias to go against finding a negative effect on teaching performance.

In columns (3) and (4) of Table 12, I check whether the new incentive regime induced internal teachers to change their teaching load and duties. I estimate equation 5 using as dependent variables a dummy equal to one if professor p was the course coordinator in year t and the number of teaching hours taught by professor p in year t, respectively. Results show that there is no significant change in the type of teaching loads and duties before and after the change in the incentive regime. This suggests that the change in teaching quality was not driven by other, simultaneously related, changes in how teaching was organized and distributed.

Finally, in Table 13, I check whether my results may be driven by a recommendation letter sent by Bocconi University in 2006 to the entire teaching faculty, asking for higher homogeneity of grades across classes. This may affect my analysis, if internal and external teachers responded to this request differently. Table 13 displays the standard deviation of average class grades across classes belonging to the same course and degree program, by academic year and by whether the teachers were internal or external. The variability of grades between classes did not decrease right after 2006, as a consequence of such recommendation, either for internal and for external teachers.

5.4 Teaching and Research ability

Understanding the sign of the correlation between α_r and α_t, as defined in Section 2, is crucial both to have a full picture of potential sorting effects and to understand the plausible cost of separating careers of teachers and researcher in university.

Figure 5 and Table 15 correlate the two sets of fixed effects as estimated from equation 7 and 8 and show that teaching and research ability are strongly positively correlated: good researchers are also good teachers. This is an important result that has not been estimated before. Columns (1) and (3) include all teachers in my sample. Columns (2) and (4) try to address the fact that teacher fixed effects represent noisy measures of the true teaching and research abilities and sampling error may bias the coefficients of columns (1) and (3). I exploit the fact that sampling error decreases substantially if the analysis is performed on a subsample of teachers with a large number of observations. I therefore estimate the correlation, including only teachers for which I can estimate the fixed effects with more that 5 observations.[32] Results are very similar but, as expected, after the

[31] Notice that it may be that what generates these results is just the fact that teaching are research efforts are more complement for math subjects than for other subjects.

[32] Notice that I always estimate the research fixed effect with 10 (yearly) observations. For the teaching

correction the coefficients are larger, because not affected anymore by the attenuation bias.

This result has crucial policy implications. First, comparing the standard deviation of the fixed effects plotted in Figure 5, which quantify the effectiveness of the time-invariant part of teaching quality, with the coefficients obtained in Tables 6 and 9, it is clear that sorting effects may potentially have much larger and substantial consequences on the overall productivity than substitution effects. Keeping the composition of teachers constant, the reform of the incentive structure improved research productivity by around 20% (in line with most of the literature on piece-rate incentives) and decreased teaching quality by 7% of a standard deviation. Instead, when we allow the composition of teachers to change and we incorporate the fact that universities will, as a consequence, attract (push away) the best (worst) researchers, the average productivity may potentially increase by much more. Therefore, it is true that Section 5.2 showed that, at the margin, pushing university professors to focus more on research may induce them to crowd out time from preparing teaching classes and may worsen their teaching performance. However, Figure 5 shows that sorting effect may potentially be much more effective.

Second, the fact that teaching and research ability are positively correlated entails that if universities are able to attract good researchers, they will also, indirectly, improve teaching quality. One of the most popular proposal to solve the trade-off between teach-ing and research, is to increase specialization of faculty members. This would entail, for example, the creation of two groups of professors, one more research-oriented and one more teaching-oriented. Figure 5 and Table 15 show that these proposals should take into consideration that good researchers are also good teachers and the potential benefit of separating careers may be minimal.

5.5 Sorting

The first way I analyze sorting effects is by evaluating the difference between the OLS and the fixed effect estimates in Tables 6 and 9. OLS estimates are always larger than fixed effects estimates, suggesting that the policy induced some positive sorting effects.

As mentioned in Section 5.5, I also analyze sorting in and out separately using direct estimates of teachers' underlying ability, obtained through equations 7 and 8.-off.

Table 16 shows how teachers' fixed effects change for (internal) professors fired before and after the change in the incentive regime, for research ability and teaching ability respectively. Columns (1) and (2) use fixed effects estimated for the entire period. The dependent variable of columns (3) and (4) are, instead, fixed effects estimated for the pre 2006 period only. Results, in line from the predictions of Section 2, show that the change in incentives induced worse researchers and therefore worse teachers to leave.

Table 17 reports instead results from equation 10 and it shows no effects on Bocconi's ability of attracting good teachers or good researchers. The reason why I don't find any positive sorting-in effect, partly in contrast with what is expected from the results of Section 2, may be due to the fact that it takes time to publish papers and it may be too early to evaluate the research and teaching productivity of very young scholars.

fixed effects, instead, the number of teacher-specific observations used depends on the number of time I observe teacher p teaching undergraduate compulsory courses.

5.6 Heterogeneity by teachers' ability

This Section analyzes how the effect on teaching and research performances changes with respect to teachers' ability. Section 2 shows that the bulk of the effort reallocation should be concentrated on low ability researchers, while the effect on research should be concentrated on very low (because of fair of being fired) or very high (because they benefit more from any unit of effort in research) ability researchers.

Column 1 of Table 14 shows that the negative effect on teaching activity is stronger for low ability researchers than for higher ability ones. The Tables displays the coefficients of the $internal_p * post2006_t$ dummy of equation 5 interacted with research ability tertiles.[33]

For what concerns the heterogeneity of the effect on research performance, columns 2 and 3 of Table 14 show that the positive effect is driven by low ability and middle ability researchers. The difference is more evident in column 3, that looks at the effect on the number of working papers.

6 Alternative Control Groups

One possible concern of using external teachers as control group is that these teachers may react to the policy as well if, for instance, their final objective is to be hired by Bocconi. This would spoil my identification strategy because it implies that the effect of the policy would spill over my control group. Moreover, one may think that external teachers are a natural control group for evaluating the effect on teaching performance but may not be as good as a control group for research activity, because they may have very different research productivity and may be on very different trends in any case. To tackle these issues I propose two alternative control groups.

The first one refers to the analysis on research. In Table 19, I use all professors belonging to Bologna University faculty in 2005 as alternative control group. Bologna University is another Italian University, whose department of management and economics is quite similar to Bocconi University in terms of quality of the economics/management department. Bologna university's economics and management department is indeed ranked as the best[34] department among Italian public institutions. Table 18 shows the productivity of Bologna University faculty members in terms of research, compared to Bocconi's faculty members. Again, I obtain data on their publications from the Web of Science website and data on the faculty composition in 2005 from the website of the Italian Ministry of Education.[35]

The second alternative control group are professors who became tenured before the policy. Given that the change in the incentive structure acts mainly in terms of promotions and tenure decisions, full professors should only be marginally affected. Since they are already fully hired by Bocconi, they should not react to changes in hiring/promotion strategies. If we assume that the effect of monetary incentives is the same on full and junior professors, than what I estimate using full professors as control group is the effect

[33]Tertiles are calculated using θ_p^r of equation 8, and are estimated only for the years before the change in the incentive regime. This is to avoid that the way ability is measured is affected by the change in incentive regime itself.

[34]or one of the top three departments, depending on the ranking considered

[35]www.miur.it

of the change in career requirements only. However, it is very likely that trends for junior and senior professors are very different, after they get tenured, especially for publications, since tenure decisions are, indeed, based research productivity or potential productivity. I will therefore use this alternative control group only for the analysis on teaching.

Moreover, both for the analysis on teaching and on research, I estimate my difference-in-difference separately on two subsamples of professors with similar age. In particular I split both the sample of internal and external teachers between those older than 43 (the mean age) and those younger than 41. This allows me to use as control group for young researchers, young external researchers since, especially for research, junior and senior faculty members may be on very different trends.

Table 19 reports results for research activity. In columns (1) and (2) I run equation 4 on the subsample of teachers younger and older than 43, respectively. As for the main results, the effect is larger for junior professors. Columns (3), (4) and (5) use, instead, Bologna faculty members as control group. Column (3) looks at the aggregate effect, column (4) looks at junior professors and column (5) at full professors. The effect is remarkably similar to my baseline estimates. The introduction of incentives led to an increase in the number of publication of 0.17 for Bocconi faculty members. The increase is stronger for young faculty members.

Table 20 reports, instead, results for teaching quality. Columns (1) and (2) split again the sample by age. The effect is similar to what found in my baseline estimates and is more negative for junior professors. Columns (3) and (4) use full professors as alternative control group. Columns (3) does not include teacher fixed effects, without and with specific trends for junior and full professors respectively. Columns (4) shows results including teachers fixed effects. Again, results are remakably similar to what found in Table 9. The introduction of research incentives worsened teaching performance by 0.04, about 7% of a standard deviation. I can't use Bologna faculty members as control group for the analysis on teaching, because information on teaching performance of Bologna faculty members is not publicly available.

Figure 3 checks the presence of parallel trends.

7 Conclusions

This paper exploits a natural experiment to test predictions of models of incentives in a multitask environment. I use administrative data from Bocconi University to analyze faculty reaction to a sharp increase in research incentives. The heterogeneity in the teaching faculty type of contracts allows me to find a control group for my Difference-in-Difference estimation. The randomization of teachers to students within the same course, in a context where the syllabus and the exams are fixed, allows me to build a credible measure of teaching performance. In particular, the specific Bocconi setting allows me to overcome two of the reasons why analyses of teachers' effectiveness are rarely done at the post secondary level: the lack of standardized tests and the endogeneity in students selection of courses (and professors).

I find evidence that the introduction of research incentives affects the allocation of effort across tasks. Results show that professors' teaching performance gets worse while their research performance significantly improves. In line with the predictions of Holmstrom and Milgrom [1991, 1994], I find that the effect is stronger for young faculty members, more exposed to career concerns. This provides evidence of the importance of implicit and explicit incentives in an organization. The number of working papers and published papers of internal Bocconi faculty increases after the introduction of incentives on research. The magnitude is in line with the literature on provision of incentives (see Prendergast [1999], Lazear [2000], Bandiera et al. [2009], Checchi et al. [2014], for example). I observe that the effect on quantity of publications does not go against the quality of publications. This may be due to the way research incentives are structured by Bocconi. On the other hand, teaching quality of faculty members more exposed to research incentives is 7% of a standard deviation lower after the change in the incentive regime. The effect is nonproportionally borne by lower ability students. My estimates suggest that encouraging one more paper has an implicit cost of 0.3 standard deviation on teaching quality. Moreover, I find evidence of positive sorting effects. After the change in incentives, lower quality researchers left Bocconi faculty and, since teaching and research ability are positively correlated, the policy attracted also good teachers.

My results suggest that it is beneficial to evaluate new policies not in isolation but as part of a coherent incentive system. I believe this paper delivers two important policy-relevant messages. First, since the negative effect on teaching is not homogenously borne by the entire students population, there is room for systems of allocation of tasks and courses to teachers that match successful scholars with those students who benefit more from their knowledge and that minimize the consequences of possible distortions. Second, I show that, while at the margin there is a trade-off between teaching and research, the overall effect is ambiguous: universities are also able to keep only good researchers under the new incentive regime and, since good researchers are also good teachers, teaching quality will improve. Finally, i provide the first piece of evidence on the correlation between research and teaching ability. This has important implications for the design of professors' incentives and hiring schemes. Policies aimed at increasing teachers' specialization that propose to dedicate part of the faculty exclusively to teaching and part of it exclusively to research, should take into consideration that good researchers are also good teachers on average.

References

Daniel Aaronson, Lisa Barrow, and William Sander. Teachers and student achievement in the chicago public high schools. *Journal of Labor Economics*, 25(1):95–135, 2007.

Joshua D. Angrist and Victor Lavy. Using Maimonides' Rule To Estimate The Effect Of Class Size On Scholastic Achievement. *The Quarterly Journal of Economics, MIT Press*, 114(2):533–575, May 1999.

Oriana Bandiera, Iwan Barankay, and Imran Rasul. Incentives for managers and inequality among workers: evidence from a firm-level experiment. *The Quarterly Journal of Economics*, 122(2):729–773, 2007.

Oriana Bandiera, Iwan Barankay, and Imran Rasul. Social connections and incentives in the workplace: Evidence from personnel data. *Econometrica*, 77(4):1047–1094, 2009.

Oriana Bandiera, Iwan Barankay, and Imran Rasul. Social incentives in the workplace. *The Review of Economic Studies*, 77(2):417–458, 2010.

Roland Benabou and Jean Tirole. Intrinsic and extrinsic motivation. *The Review of Economic Studies*, 70(3):489–520, 2003.

Eric Bettinger and Long Bridget Terry. Does cheaper mean better? the impact of using adjunt instructors on student outcomes. *Review of Economics and Statistics*, 92(3): 598–613, 08 2010.

Ernest L. Boyer. Scholarship reconsidered priorities of the professoriate. Technical report, Princeton, N.J. : Carnegie Foundation for the Advancement of Teaching, 1990.

Michela Braga, Marco Paccagnella, and Michele Pellizzari. Evaluating students evaluations of professors. Working papers, IGIER (Innocenzo Gasparini Institute for Economic Research), Bocconi University, 2011.

James A. Brickley and Jerold L. Zimmerman. Changing incentives in a multitask environment: evidence from a top-tier business school. *Journal of Corporate Finance*, 7 (4):367–396, December 2001.

Alexander Brüggen and Frank Moers. The role of financial incentives and social incentives in multi-task settings. *Journal of Management Accounting Research*, 19(1):25–50, 2007.

Scott E. Carrell and James E. West. Does professor quality matter? evidence from random assignment of students to professors. *Journal of Political Economy*, 118(3): 409–432, 06 2010.

Daniele Checchi, Gianni De Fraja, and Stefano Verzillo. Publish or Perish? Incentives and Careers in Italian Academia. IZA Discussion Papers 8345, Institute for the Study of Labor (IZA), July 2014.

Raj Chetty, John N Friedman, and Jonah E Rockoff. Measuring the impacts of teachers i: Evaluating bias in teacher value-added estimates. Technical report, National Bureau of Economic Research, 2013.

Iain Cockburn, Rebecca Henderson, and Scott Stern. Balancing incentives: The tension between basic and applied research. NBER Working Papers 6882, January 1999.

Esther Duflo, Pascaline Dupas, and Michael Kremer. Peer effects, teacher incentives, and the impact of tracking: Evidence from a randomized evaluation in kenya. *American Economic Review*, 101(5):1739–74, 2011.

Etienne Dumont, Bernard Fortin, Nicolas Jacquemet, and Bruce Shearer. Physicians' Multitasking and Incentives: Empirical Evidence from a Natural Experiment. *Journal of Health Economics*, 27(6):1436–1450, 2008.

Glenn Ellison. Evolving standards for academic publishing: A q-r theory. *Journal of Political Economy*, 110(5):994–1034, 2002a.

Glenn Ellison. The slowdown of the economics publishing process. *Journal of Political Economy*, 110(5):947–993, 2002b.

Ernst Fehr and Urs Fischbacher. Why social preferences matter–the impact of non-selfish motives on competition, cooperation and incentives. *The economic journal*, 112(478): C1–C33, 2002.

Susan Feng Lu. Multitasking, information disclosure, and product quality: Evidence from nursing homes. *Journal of Economics Management Strategy*, 21(3):673–705, 2012.

David N. Figlio and Lawrence W. Kenny. Individual teacher incentives and student performance. *Journal of Public Economics*, 91(5-6):901–914, 2007.

David N. Figlio, Morton O. Schapiro, and Kevin B. Soter. Are tenure track professors better teachers? Working Paper 19406, National Bureau of Economic Research, September 2013.

Jr. Fryer, Roland G. and Richard T. Holden. Multitasking, dynamic complementarities, and incentives: A cautionary tale. Technical report, Cambridge, Massachusetts: Harvard University, 2013.

Michael Gibbs, Kenneth Merchant, Wim Van der Stede, and Mark Vargus. Performance measure properties and incentives. 2004.

Giacomo De Giorgi, Michele Pellizzari, and William Gui Woolston. Class size and class heterogeneity. *Journal of European Economic Association*, 10(4):795–830, 08 2012.

Bengt Holmstrom and Paul Milgrom. Multitask principal-agent analyses: Incentive contracts, asset ownership, and job design. *Journal of Law, Economics and Organization*, 7(0):24–52, Special I 1991.

Bengt Holmstrom and Paul Milgrom. The firm as an incentive system. *American Economic Review*, 84(4):972–91, September 1994.

Fuhai Hong, Tanjim Hossain, John A. List, and Migiwa Tanaka. Testing the theory of multitasking: Evidence from a natural field experiment in chinese factories. Working paper, National Bureau of Economic Research, 2013.

Jun Ishii and Steven G Rivkin. Impediments to the estimation of teacher value added. *Education Finance and Policy*, 4(4):520–536, 2009.

Brian A. Jacob. Accountability, incentives and behavior: the impact of high-stakes testing in the chicago public schools. *Journal of Public Economics*, 89(5-6):761–796, 2005.

Carlos Munoz Johnson, Ryan and David H. Reiley. The war for the fare': How driver compensation affects bus system performance. *Economic Inquiry*, forthcoming.

Thomas J. Kane and Douglas O. Staiger. Estimating teacher impacts on student achievement: An experimental evaluation. NBER Working Papers 14607, National Bureau of Economic Research, Inc, December 2008.

Edward P. Lazear. Performance pay and productivity. *American Economic Review*, 90 (5):1346–1361, December 2000.

Daniel McCaffrey and J.R. Lockwood. Models for value-added modeling of teacher effects. *Journal of Educational and Behavioral Statistics*, 29(1), 2004.

Abigail Payne and Joanne Roberts. Government oversight of public universities: Are centralized performance schemes related to increased quantity or quality? *The Review of Economics and Statistics*, 92(1):207–212, 2010.

Canice Prendergast. The provision of incentives in firms. *Journal of Economic Literature*, 37(1):7–63, March 1999.

Canice Prendergast. Intrinsic motivation and incentives. *The American economic review*, 98(2):201–205, 2008.

Steven G Rivkin, Eric A Hanushek, and John F Kain. Teachers, schools, and academic achievement. *Econometrica*, 73(2):417–458, 2005.

Jonah E Rockoff. The impact of individual teachers on student achievement: Evidence from panel data. *The American Economic Review*, 94(2):247–252, 2004.

Jesse Rothstein. Student sorting and bias in value added estimation: Selection on observables and unobservables. Nber working papers, National Bureau of Economic Research, Inc, January 2009.

Jesse Rothstein. Teacher quality in educational production: Tracking, decay, and student achievement. *The Quarterly Journal of Economics*, 125(1):175–214, February 2010.

Margaret E. Slade. Multitask agency and contract choice: An empirical exploration. *International Economic Review*, 37(2):465–486, 1996.

A Appendix A

Given the exponential utility function and normality of ϵ_i, the agent receives certainty equivalent

$$CE = b_r\alpha_r e_r + b_t\alpha_t e_t + s - \frac{1}{2}(e_r^2 + e_t^2) - \delta e_r e_t - \frac{\eta}{2}(b_t^2\sigma_t + b_r^2\sigma_r) \tag{11}$$

The first order conditions obtained from maximizing the expected utility of the agent with respect to e_r and e_t are:

$$\alpha_r b_r = e_r + \delta e_t; \quad \alpha_t b_t = e_t + \delta e_r \tag{12}$$

and the optimal (internal) solutions are:

$$e_r^* = \frac{b_r\alpha_r - \delta b_t\alpha_t}{1-\delta^2}; \quad e_t^* = \frac{b_t\alpha_t - \delta b_r\alpha_r}{1-\delta^2} \tag{13}$$

Therefore, taking the partial derivatives with respect to b_r, I get:

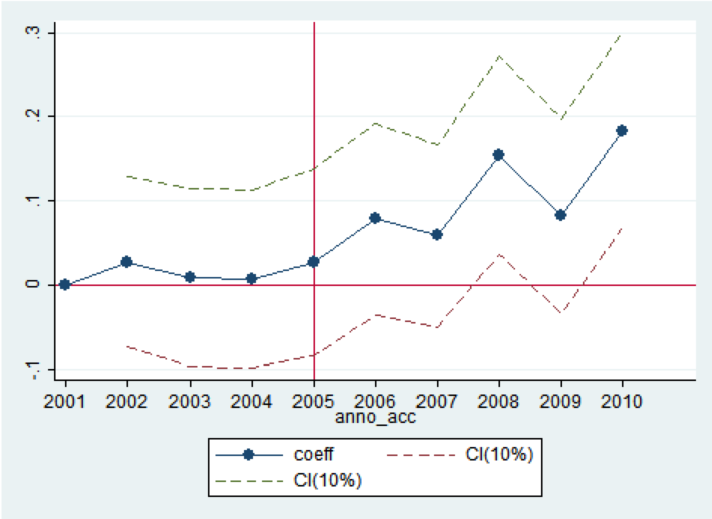
$$\frac{\partial e_r^*}{\partial b_r} = \frac{\alpha_r}{1-\delta^2} > 0; \qquad \frac{\partial e_t^*}{\partial b_r} = -\frac{\delta\alpha_r}{1-\delta^2} = \begin{cases} > 0 & if \delta < 0 \\ < 0 & if \delta > 0 \end{cases} \tag{14}$$

To show the results stated in Proposition 2, I take the derivatives also with respect to ability:

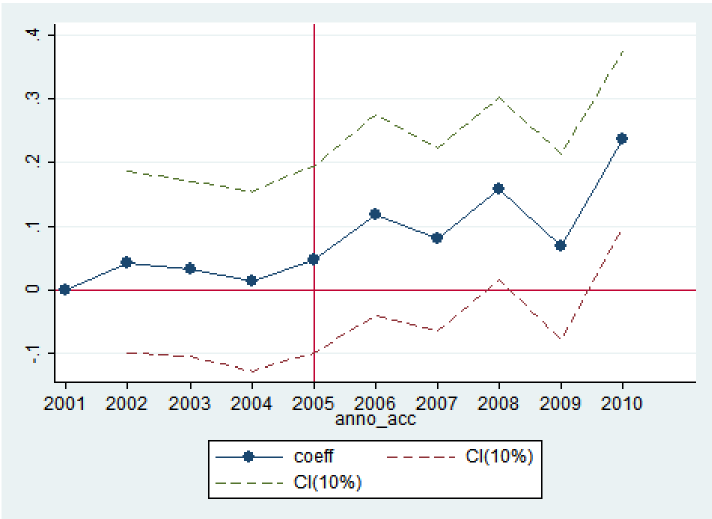
$$\frac{\partial^2 e_r^*}{\partial b_r \partial\alpha_r} = \frac{1}{1-\delta^2} > 0; \qquad \frac{\partial e_t^*}{\partial b_r \partial\alpha_r} = -\frac{\delta}{1-\delta^2} = \begin{cases} > 0 & if \delta < 0 \\ < 0 & if \delta > 0 \end{cases} \tag{15}$$

Figure 1: Research Difference in Difference graphs

N pub (WoK)



N pub weighted (WoK)



Working papers (google)

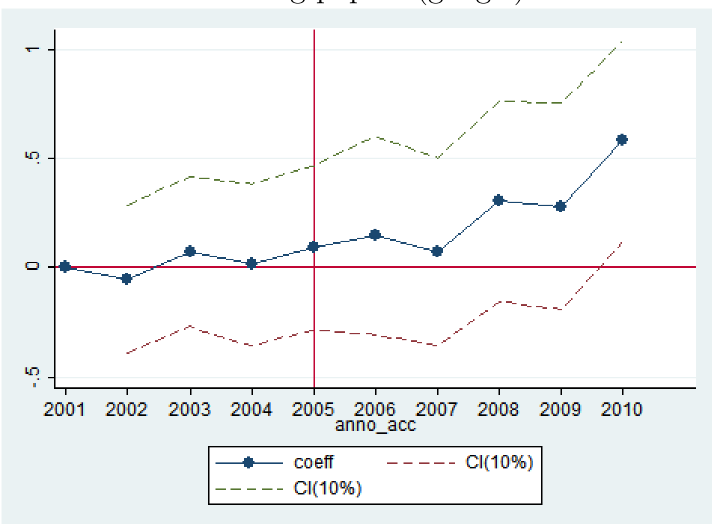
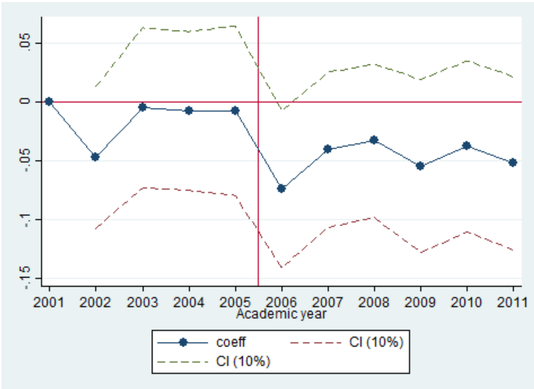


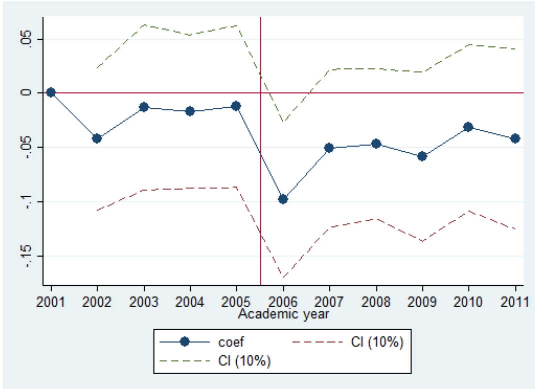
Figure 2: Teaching Difference in Difference graphs

Internal professors
teaching performance (α_{pct})



a

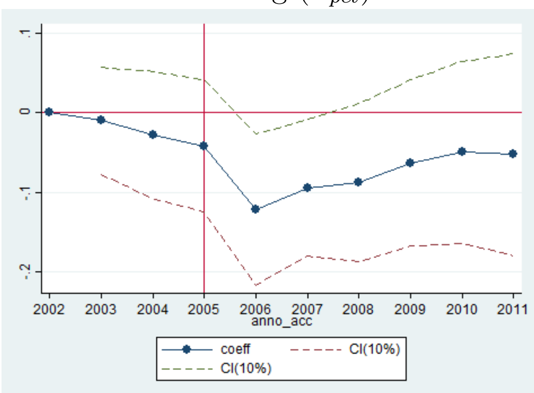
Assistant professors
teaching performance (α_{pct})



b

Figure 3: Alternative identification strategies graphs

control: full prof
Teaching (α_{pct})



control: Bologna prof
Research (N pub)

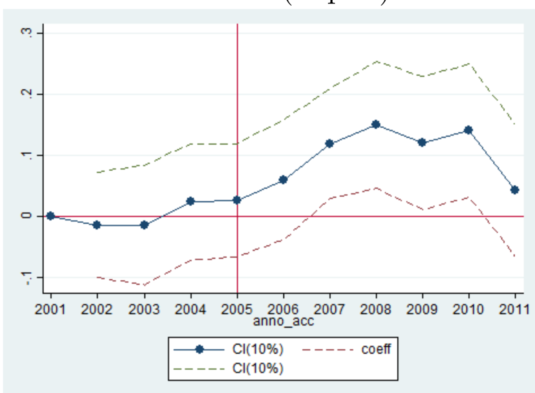


Figure 4: Robustness of teachers fixed effects

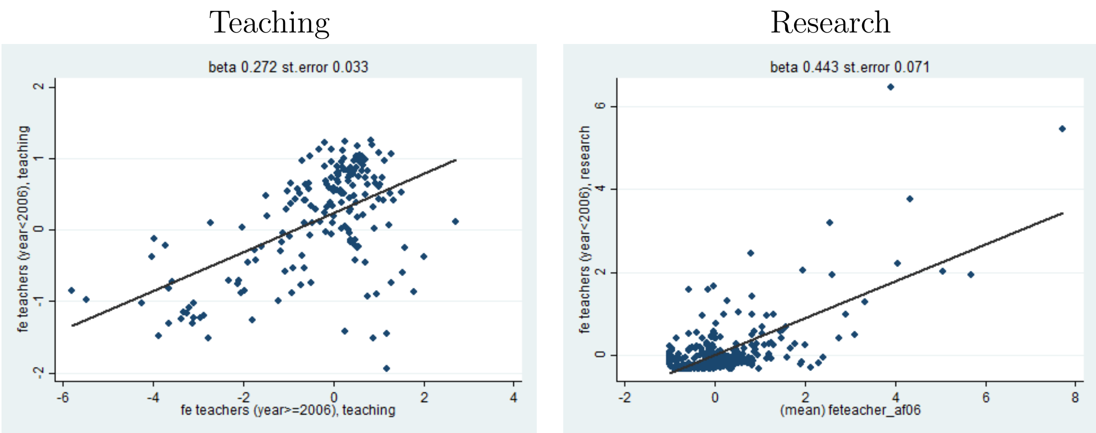
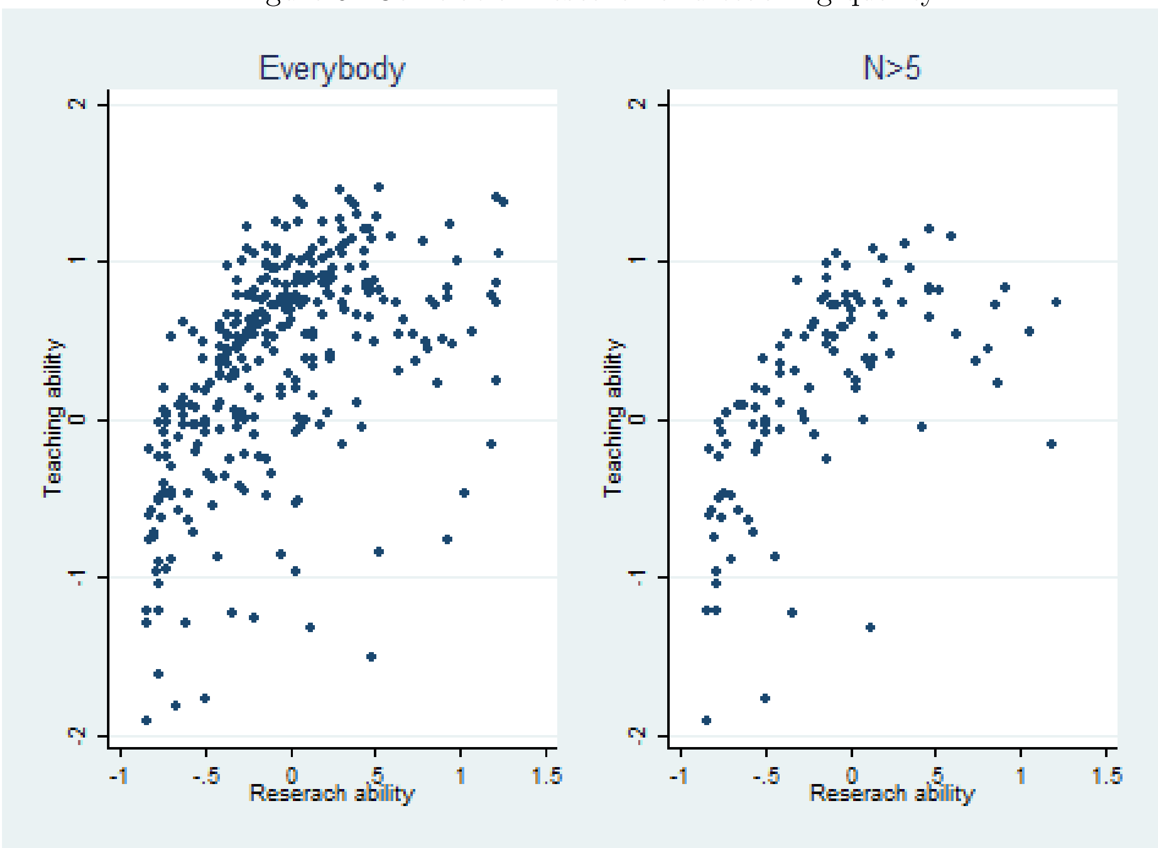


Figure 5: Correlation research and teaching quality



Tables

Table 1: Descriptive statistics Students

Variable	Mean	Std. Dev.	Min.	Max.
	[1]	[2]	[3]	[4]
1=female	0.469	0.499	0	1
year birth	1985	3.249	1954	1993
1=italian	0.973	0.163	0	1
1=from Milan	0.246	0.431	0	1
hs grade	0.899	0.103	0.6	1
exam grades	25.532	3.532	18	31
N		501189		

Table 2: Types of Teacher contracts

Description	category
Adjunct Professor	assistant
Researcher Bocconi	assistant
Assistant professor Bocconi	assistant
Assistant Professor (Job Market) Bocconi	assistant
Assistant Professor (Young Foreigners) Bocconi	assistant
1 year scholar Bocconi	assistant
2 year scholar Bocconi	assistant
3 year contract researcher Bocconi	assistant
Phd Student Bocconi	assistant
Assistant professor Bocconi senior	assistant
Researcher Bocconi	assistant
Full contract researcher Bocconi	assistant
Researcher Bocconi on leave	assistant
Associate professor Bocconi	associate
Full Professor Bocconi	full
Extraordinary professor Bocconi	full
Non academics (expert in the subject)	non academics
Associate professor other university	other univ
Associate professor Bocconi on leave	other univ
Temporary contract collaborator SDA[a]	other univ
Collaborator SDA	other univ
permanent contract collaborator Research centers	other univ
Full contract researcher SDA	other univ
Lecturer SDA	other univ
Lecturer SDA Senior	other univ
Full Professor other university	other univ
Full Professor Bocconi on leave	other univ
Associate professor other university	other univ
Full Professor other university	other univ
Researcher other university	other univ
Extraordinary professor other university	other univ
Visiting Professor Long Term	other univ
Visiting Professor Short Term	other univ

the big amount of contracts is due to the fact that identical contracts were having different names over the years.

[a] SDA is the Bocconi School of Managers. It offers MBAs and master course only. Faculty is hired and promoted according to different and independent standards.

Table 3: Descriptive statistics - Teachers

	Internal	External	Diff
Teachers' descriptives			
N teaching hours per class	38.91	33.91	5.47***
	(16.60)	(17.44)	(1.34)
Age	43.18	41.29	1.89**
	(9.45)	(7.80)	(0.77)
% female	32.27	34.25	-0.20
	(0.47)	(0.47)	(0.045)
Teachers' Department			
Accounting	14.8 %	20.8%	
Math/Stat	13.3%	24.6%	
Economics	20.2%	13.8%	
Finance	16.7%	7.4%	
Management	39.0%	33.5%	
Tot	100%	100%	
Teachers' Position			
% Assistant prof	27.03%		
% Associate prof	15.52%		
% Full prof	15.33%		
% Non academic		17.36%	
% Other univ prof		13.43%	
% Lecturers	2.45%		

in parenthesis standard deviation (columns 1 and 2) and standard errors (column 3).

Table 4: Random Allocation

	Av. final hs grade[a]	Av. female	Av. from Mi	Sd final hs grade
internal	0.001	0.000	0.001	-0.002
	(0.001)	(0.000)	(0.003)	(0.002)
age	0.000	-0.000	-0.000	-0.000*
	(0.000)	(0.000)	(0.000)	(0.000)
female	0.001	0.000	0.000	-0.001
	(0.001)	(0.000)	(0.002)	(0.002)
coordinator	0.000	-0.001	0.003	-0.000
	(0.001)	(0.001)	(0.003)	(0.002)
N	3889	3889	3889	3889
course*year fe	Yes	Yes	Yes	Yes
F stat joint sign	0.75	0.95	0.39	1.58

Robust standard errors clustered by course in parentheses.

Table 5: Summary statistics-Research

	Overall	Post 2006	Pre 2006	Diff
		N publications		
Internal 2005	0.539	0.680	0.302	0.381***
sd	1.561			(0.061)
External 2005	0.416	0.481	0.264	0.239***
sd	1.318			(0.042)
Diff	0.199	0.199**	0.037***	0.143**
	(0.199)	(0.100)	(0.071)	(0.074)
		N publications (Bocconi index)		
Internal 2005	0.814	0.927	0.625	0.336***
sd	2.328			(0.084)
External 2005	0.575	0.634	0.437	0.251***
sd	1.944			(0.080)
Diff	0.293	0.293**	0.187*	0.085
		N working papers (Google Scholar)		
Internal 2005	1.506	1.692	1.193	0.526***
sd	2.583			(0.126)
External 2005	1.052	1.159	0.809	0.343***
sd	2.278			(0.105)
Diff	0.533	0.533**	0.385***	0.182*
	(0.533)	(0.172)	(0.191)	(0.164)

sd (or se) in parenthesis

Table 6: Effect on Research

	[1]	[2]	[3]	[4]	[5]	[6]
	Dependent variable: N Pub				N pub$^{(1/2)}$	
1=internal*post2006	0.206**		0.142**		0.081**	
	(0.100)		(0.070)		(0.034)	
1=junior pr * post2006		0.224*		0.157*		0.087**
		(0.114)		(0.091)		(0.039)
1=full pr* post2006		0.137		0.099		0.065
		(0.186)		(0.123)		(0.050)
N	5230	5230	5230	5230	5230	5230
	Dependent variable: N Pub (weighted by Bocconi)				N pub w$^{(1/2)}$	
=internal*post2006	0.315***		0.130		0.082**	
	(0.103)		(0.098)		(0.041)	
1=junior pr * post2006		0.266**		0.154		0.094**
		(0.110)		(0.109)		(0.047)
1=full pr* post2006		0.496**		0.064		0.050
		(0.231)		(0.174)		(0.060)
N	5209	5209	5209	5209	5209	5209
	Dependent variable: N wp (Google Scholar)				N wp$^{(1/2)}$	
1=internal*post2006	0.711***		0.148		0.091*	
	(0.166)		(0.139)		(0.052)	
1=junior pr * post2006		0.492***		0.212*		0.120**
		(0.166)		(0.136)		(0.059)
1=full pr* post2006		1.572***		-0.035		0.008
		(0.404)		(0.203)		(0.073)
N	5113	5113	5113	5113	5113	5113
Teacher fe	No	No	Yes	Yes	Yes	Yes

Robust standard errors clustered by teacher in parentheses. Additional controls: age, age squared, academic year fixed effects. Years between 2001 and 2010. Only professors included in the analysis on teaching. Junior professors are assistant and associate professors.

[a] Publications are weighted in the same way Bocconi University assigns monetary incentives. I give weight=15 if articles are in journals considered by Bocconi as belonging to band "A+", weight=7 if journals are considered as belonging to band "A", weight=3 if belonging to band "B" and weight=1 if not belonging to any band. The index is computed as $\Sigma_i(wight_i * pub_i)/Nauthors_i$ where i is a publication published by professor p in year t.

[b] This is the internal status in 2005

Table 7: Step 1: regression on students micro data.

Dependent variable: exam grade	All [1]
hs grade	-3.704***
	(0.225)
hs grade^2	4.159***
	(0.131)
1=female	-0.051***
	(0.003)
1=italian	0.142***
	(0.013)
1=from Milan	0.074***
	(0.003)
N	501132

Robust standard errors clustered by class-year in parentheses. Additional controls: dummies for type of high school, dummies for the full interaction of classes and years (α_{pct}).

Table 8: Descriptives Teaching quality

		α_{pct}			
		Overall	Post 2006	Pre 2006	Diff
Internal 2005	*mean*	-0.020	0.146	-0.197	0.343***
	sd	0.632			0.024
External 2005	*mean*	0.074	0.239	-0.192	0.431***
	sd	0.645			0.026
Diff			-0.093***	-0.005	-0.088***
			0.033	0.015	0.036

Table 9: Step 2: regression at teacher level - students' grades

	[1]	[2]	[3]	[4]
int*post06	-0.011		-0.037**	
	(0.012)		(0.018)	
jun pr*post06		-0.014		-0.042**
		(0.013)		(0.020)
full pr*post06		-0.001		-0.023
		(0.016)		(0.022)
N	3889	3889	3889	3889
Teachers fe	No	No	Yes	Yes
Year*course*degree pr fe	Yes	Yes	Yes	Yes

Robust standard errors clustered by teacher in parentheses. Regressions are weighted by number of teaching hours per class. Additional controls: age and age squared of teachers, class size, class average final high school grade. Junior professors are assistant and associate professors.

[a] Status as it was before 2006

[b] The number of observations is lower because Bocconi collected students evaluations in only a subsample of courses for the years 2004/2005 and 2005/2006.

Table 10: regression at student level - students' grades

	[1]	[2]	[3]	[4]
	Dependent variable: stud grade (std)			
	[1]	[2]	[3]	[4]
int^a*post06	-0.037***		0.002	
	(0.014)		(0.016)	
jun^a pr*post06		-0.045***		-0.005
		(0.016)		(0.017)
full^a pr*post06		-0.009		0.028
		(0.020)		(0.022)
int^a*post06*mid ability stud			-0.079***	
			(0.014)	
int^a*post06*low ability stud			-0.097***	
			(0.020)	
jun^a*post06*mid ability stud				-0.077***
				(0.015)
jun^a*post06*low ability stud				-0.100***
				(0.021)
ord^a*post06*mid ability stud				-0.086***
				(0.022)
ord^a*post06*low ability stud				-0.086**
				(0.036)
N	346628	346628	346628	346628
Teachers fe	Yes	Yes	Yes	Yes
Year*course*degree pr fe	Yes	Yes	Yes	Yes

Control set: teacher age, age sq, student gender, hs, whether Italian, whether from Milano. Se clustered by teacher. Ability based on final high school grade of students: High ability (omitted)=between 1 and 0.9; middle ability = between 0.8 and 0.9; low ability: below 0.8

[a] Status as it was before 2006

[b] The number of observations is lower because Bocconi collected students evaluations in only a subsample of courses for the years 2004/2005 and 2005/2006.

Table 11: Robustness checks Teaching

	no 09-10-11	also lecturers	include switches	weight by h. taught
	[1]	[2]	[3]	[4]
int05[a]*post06	-0.037*	-0.034*		-0.035*
	(0.020)	(0.018)		(0.021)
lecturer[a]*post06		-0.047		
		(0.042)		
int[b]*post06			-0.027*	
			(0.016)	
N	2848	4201	3889	3889
Teachers fe	Yes	Yes	Yes	Yes

Robust standard errors clustered by teacher in parentheses. Additional controls: age and age squared of teachers, teacher experience in Bocconi class size. Column (1) excludes the years when teaching incentives were also in place; column (2) includes lecturers and specifies a different treatment effect for lecturers; column (3) includes switchers and teachers fixed effects; column (4) weights professors by number of teaching hours.

[a] Status as it was before 2006

[b] contemporaneous status

Table 12: Robustness checks Teaching 2

Dep var:	Grading α_{ptc}		1=course coordin[a]	1=Num of taught h[b]
	[1]	[2]	[3]	[4]
int*post 06	-0.045**	-0.042**	0.025	0.671
	(0.020)	(0.020)	(0.037)	(1.084)
int*post 06*obj[c]	0.024			
	(0.047)			
int*post 06*math dep[d]		0.017		
		(0.046)		
N	3889	3889	3889	2989[e]
Teachers fe	Yes	Yes	Yes	Yes

Robust standard errors clustered by teacher in parentheses. Additional controls: age and age squared of teachers, dummies for teacher experience in Bocconi.

[a] 1=whether professor p in year t was the course coordinator

[b] Tot n of teaching hours in year t by professor p

[c] Objective if the name of the course includes the words *"math"*, *"stat"*, *"quantit"*

[d] Math if the teacher belongs to the math and statistics departments

[e] N of observations at the teacher-year level (if a teacher teaches more than one courses n of teaching hours are summed)

Table 13: Robustness checks- teaching 3: average class grades

	sd av. class gr internal	sd av. class gr external
2001	0.390	0.434
2002	0.283	0.379
2003	0.390	0.453
2004	0.415	0.375
2005	0.423	0.431
2006	0.407	0.477
2007	0.375	0.400
2008	0.450	0.349
2009	0.406	0.442
2010	0.428	0.425
2011	0.468	0.404

[a] This is the standard deviation of average class grades within courses (of classes that sit the same exam).

Table 14: Heterogeneity by teachers' ability

Dep. var	α_{pct}	n pub	n pub weight	n wp (google)
	[1]	[2]	[3]	[4]
int*post 06* ability q 1	-0.087**	0.146***	0.185**	0.296***
	(0.037)	(0.042)	(0.073)	(0.113)
int*post 06* ability q 2	-0.036	0.200***	0.351***	0.166
	(0.041)	(0.070)	(0.102)	(0.172)
int*post 06* ability q 3	-0.035	0.184	0.197	0.095
	(0.042)	(0.140)	(0.257)	(0.254)
N	3770	6281	6264	6082

Additional controls: age, age squared, all double interactions, teacher fixed effects, year fixed effects.

Table 15: Teaching and research ability

	Dep. var= Teaching Fe			
	everybody	N>5[a]	everybody	N>5[a]
	[1]	[2]	[3]	[4]
research FE	0.715***	0.795***	0.542***	0.640***
	(0.067)	(0.102)	(0.062)	(0.094)
N	313	109	313	109
Controls	No	No	Yes	Yes

Additional controls: age at entry (linear and squared), gender.
[a] N>5 is referred to the n of observations over which is estimated the teacher fixed effect in the teaching quality regression (for the researh quality regression N=10 for every teacher)

Table 16: Sorting out

	Fixed Effects all		pre 06 Fixed Effects	
Dep Variable:	Research Fe	Fe Teaching	Fe Research	Fe Teaching
	(θ^r_p)	(θ^t_p)	(θ^r_p)	(θ^t_p)
	[1]	[2]	[3]	[4]
1=exit after 2006	-0.133**	-0.113**	-0.099*	-0.100
	(0.063)	(0.054)	(0.055)	(0.074)
1=exit pre 2006	-0.044	0.023	-0.008	-0.040
	(0.054)	(0.034)	(0.046)	(0.036)
N	345	352	232	232

[a] Excluding those exiting because retiring, omitted category=those staying. additional controls: dummies for year of entry, gender, age at entry, age at entry squared.

Table 17: Sorting in

Dep Variable:	Research Fe	Fe Teaching
	(θ^{r_p})	(θ^{t_p})
	[1]	[2]
1=entry after 2006	-0.051	0.049
	(0.091)	(0.056)
tr y entry	0.009	0.099***
	(0.011)	(0.013)
tr y entry sq	0.001*	-0.000
	(0.000)	(0.000)
N	350	352

Excluding those exiting because retiring, omitted category=those staying. additional controls: dummies for year of entry, gender, age at entry, age at entry squared. Columns 3 and 4: omitted category= entry before 2006, additional controls=time trend of year of entry (linear and squared), age at entry (linear and squared and triple), gender. Only for teachers eneterd after 2000.

Table 18: Summary statistics on number of publications

	Bologna	Bocconi	diff
	Junior prof		
N pub	0.201	0.417	-0.217***
	(0.018)	(0.022)	0.033
	1221	2197	
	Senior prof		
N pub	0.280	0.481	-0.201***
	(0.030)	(0.043)	(0.051)
	792	709	

Table 19: Alternative identification strategies - Research

	age groups		Bologna prof		
	< m age (43)	> m age (43)	All	Jun	Full
	[1]	[2]	[3]	[4]	[5]
internal*post06	0.170*	0.119			
	(0.096)	(0.124)			
bocconi*post06			0.162**		
			(0.064)		
jun bocc*post06				0.221***	
				(0.080)	
ord bocc*post06					0.051
					(0.107)
N	3119	2111	4497	3063	1434

Robust standard errors clustered by teacher in parentheses. Additional controls: age and age squared of teachers, year fixed effects. column (1) and (2) use as control group external teachers teachers in the same age group (< or > meanage), columns (3) and (4) use as control group use as control group professors from Bologna University.

[a] Status as it was before 2006

Table 20: Alternative Identification Strategy - Teaching

	Dep Var: α_{pct}			
	only	only	control=professors just	
	(< 43)	(> 43)	became tenured	
	[1]	[2]	[3]	[4]
internal*post06	-0.061*	-0.034		
	(0.032)	(0.029)		
no full pre05*post06			-0.221***	-0.042*
			(0.052)	(0.025)
N	1958	1931	2068	2068
Teachers fe	Yes	Yes	No	Yes
Year*course*deg fe	Yes	Yes	Yes	Yes

Robust standard errors clustered by teacher in parentheses. Additional controls: age and age squared of teachers, dummies for year of arrival in Bocconi. Only internal teachers (in 2005).

CENTRE FOR ECONOMIC PERFORMANCE
Recent Discussion Papers

1385	Ferdinando Monte Stephen J. Redding Esteban Rossi-Hansberg	Commuting, Migration and Local Employment Elasticities
1384	Tito Boeri Juan Francisco Jimeno	The Unbearable Divergence of Unemployment in Europe
1383	Sarah Flèche	Distaste for Centralization: Evidence from a Quasi-Natural Experiment in Switzerland
1382	Johannes Boehm	The Impact of Contract Enforcement Costs on Outsourcing and Aggregate Productivity
1381	Andrew B. Bernard Swati Dhingra	Contracting and the Division of the Gains from Trade
1380	Warn N. Lekfuangfu Nattavudh Powdthavee Andrew E. Clark George Ward	Early Maternal Employment and Non-cognitive Outcomes in Early Childhood and Adolescence: Evidence from British Birth Cohort Data
1379	Scott R. Baker Nicholas Bloom Steven J. Davis	Measuring Economic Policy Uncertainty
1378	Holger Breinlich Volker Nocke Nicolas Schutz	Merger Policy in a Quantitative Model of International Trade
1377	Kalina Manova Zhihong Yu	How Firms Export: Processing vs. Ordinary Trade With Financial Frictions
1376	Jordi Blanes i Vidal Tom Kirchmaier	The Effect of Police Response Time on Crime Detection
1375	Fabrice Defever Christian Fischer Jens Suedekum	Relational Contracts and Supplier Turnover in the Global Economy

1374	Brian Bell Rui Costa Stephen Machin	Crime, Compulsory Schooling Laws and Education
1373	Christos Genakos Costas Roumanias Tommaso Valletti	Loss Aversion on the Phone
1372	Shaun Larcom Ferdinand Rauch Tim Willems	The Benefits of Forced Experimentation: Striking Evidence from the London Underground Network
1371	Natalia Ramondo Veronica Rappoport Kim J. Ruhl	Intrafirm Trade and Vertical Fragmentation in U.S. Multinational Corporations
1370	Andrew Eyles Stephen Machin Olmo Silva	Academies 2: The New Batch
1369	Yonas Alem Jonathan Colmer	Consumption Smoothing and the Welfare Cost of Uncertainty
1368	Andrew Eyles Stephen Machin	The Introduction of Academy Schools to England's Education
1367	Jeremiah Dittmar Skipper Seabold	Media, Markets and Institutional Change: Evidence from the Protestant Reformation
1366	Matthew D. Adler Paul Dolan Georgios Kavetsos	Would you Choose to be Happy? Tradeoffs Between Happiness and the Other Dimensions of Life in a Large Population Survey
1365	Jeremiah Dittmar	New Media, Competition, and Growth: European Cities After Gutenberg

The Centre for Economic Performance Publications Unit
Tel 020 7955 7673 Fax 020 7404 0612
Email info@cep.lse.ac.uk Web site http://cep.lse.ac.uk