Effort and Selection Effects of Performance Pay in Knowledge Creation *

Erina Ytsma,[†] Carnegie Mellon University, Tepper School of Business

August 2019

Abstract

I study the effect of performance pay on research productivity through effort and selection effects using the introduction of performance pay in German academia as a natural experiment. To this end, I consolidated information from various, unstructured data sources to construct a data set that encompasses the affiliation history and publication records of the universe of academics in Germany. The performance pay reform introduced attraction and retention bonuses, as well as relatively weaker on-the-job performance bonuses that take effect at a later point in time. I estimate the pure effort effect of these performance pay incentives in a difference-in-differences framework, comparing changes in research productivity of a treated cohort of academics, who receive performance pay because they started their first tenured position after the reform, with a control cohort that receives flat wages because they started their first tenured position just before the reform. I find a positive effort effect of performance pay that is economically large; amounting to a 12 to 16% average increase in research productivity. This increase manifests itself most robustly as an increase in research quantity and persists for a number of years. The effort response is strongest and most robust for less productive academics, with increases in pure quantity as well as quality-adjusted research output, while the average impact of the work of top quartile academics decreases. Performing textual analysis on paper abstracts to construct novelty and impact metrics, I find that the least novel work of top quartile academics becomes more novel, but the novelty of their most novel work declines. This work however does find more follow-on research in subsequent papers in the same field and is thus more impactful. I estimate the selection effect by analyzing the rate at which academics of different productivity levels switch to the performance pay scheme. I use the fact that the old and new wage schemes compare differently for academics at different ages, which gives rise to selection incentives that are inversely related to age. Exploiting this variation in a difference-in-differences framework, I find that more productive academics are more likely to select into performance pay. Hence, performance pay, increases research output in academia through both effort and selection effects. However, because the effort effect is strongest for relatively less productive academics, while relatively more productive academics select into performance pay, the selection effect partially counteracts the impact of the effort effect.

PRELIMINARY AND INCOMPLETE; PLEASE DO NOT CITE OR DISTRUBUTE

JEL classification: J33, M52, O31

*I would like to thank Pierre Azoulay, Oriana Bandiera, Tim Besley, Jordi Blanes-i-Vidal, Kenneth Corts, Baran Duzce, Florian Englmaier, Jeff Furman, Maitreesh Ghatak, Bob Gibbons, De Gruyter, Rosario Macera, Bob Miller, Steve Pischke, Andrea Prat, Carol Propper, John van Reenen, Mark Schankerman, Axel Schniederjuergen, Ananya Sen, Chris Stanton, Scott Stern, Neil Thompson, the ministries of education of the German states and seminar and conference participants at DRUID 2016, the 31st EEA Congres, ESNASM 2017, ESEM 2017, ESEWM 17, IOEA 18, SIOE 18, BePE 2018, AEA 2019, GEIRC 2019, EAA 2019, GSE-OE 2019, MIT TIES, MIT IDE, MIT OE Lunch, Universidad Carlos III, Copenhagen Business School, Carnegie Mellon Tepper and UANDES for helpful comments, information or data. A previous version of this paper was titled "Career Concerns in Knowledge Creation".

†eytsma@andrew.cmu.edu

1 Introduction

Knowledge creation is an important driver of economic growth (Romer 1990, 1986, Lucas 1990, 1988), yet much is still unknown about the drivers of knowledge creation itself. For instance, although implicit performance incentives in the form of career concerns, and explicit incentives such as (publication) bonuses are widespread in academia and other knowledge-intensive industries (Lazear and Rosen 1981, Holmström 1982, Waldman 1990, Gibbons and Murphy 1992, DeVaro and Waldman 2012), we have a limited understanding of how effective these incentives are for knowledge production. Studies by Backes-Gellner and Schlinghoff (2004), Coupé et al. (2005), Rauber and Ursprung (2008), using observational data of research output over the academic life cycle, find a hump-shaped relationship between research productivity and time elapsed since PhD conferral, with a peak around the time of tenure or promotion to full professor (in the US). These patterns are in line with up-or-out and promotion incentives driving productivity in the run-up to tenure and promotions, but the data do not allow for a causal estimation of the size of the incentive effect. The decline in productivity after tenure, respectively promotion to full professor, also begs the question whether (stronger) explicit performance incentives could potentially keep research productivity up for longer (Gibbons and Murphy 1992).

This paper aims to shed light on the effect of performance pay on knowledge creation by causally identifying the effort and selection effects of performance incentives in academia. To this end, I exploit the introduction of performance pay in German academia in 2005 as a natural experiment. The reform introduced a new professorial salary scheme comprising a basic wage plus performance related bonuses which increased both non-contractible performance incentives and explicit pay-for-performance incentives. The bonuses take the form of potentially very large wage premiums determined in contract negotiations, and relatively weaker, explicit on-the-job performance tournament bonuses. The performance related pay scheme replaced a pay scheme in which professorial salaries increased with age and wages were thus effectively flat (Handel 2005). Combined with the fact that especially the negotiated wage premiums in the new performance pay scheme can more than double a professor's monthly wage, this means that the reform substantially increased performance incentives. The specifics of the roll-out of the performance related pay scheme give rise to a differential incidence of performance incentives across academic cohorts, which allows me to causally and separately identify the effort and selection effects. I consolidated information from various, unstructured data sources to construct a data set comprising affiliations, research productivity measures and related information of the universe of academics in Germany[1] for the empirical analyses.

I present a theoretical model to analyze the effects of performance pay on the effort and selection of academics, designed to reflect the particular features of the performance pay reform in German academia, much like Lazear (2000a). The model combines tournament pay as in Lazear and Rosen (1981) with career concerns as in Holmström (1982, 1999) and Gibbons and Murphy (1992) to capture, respectively, the on-the-job performance tournaments and attraction and retention bonuses of the performance pay reform. Ederer (2010), Hansen (2012), Miklós-Thal and Ullrich (2015, 2016), Brehm et al. (2017) also combine learning (career concerns) with tournament pay. Ederer (2010) and Hansen (2012) focus on the effects of provision of interim feedback on effort, while Brehm et al. (2017) study the productivity effects of value-added merit pay tournaments for agents closer and further from an uncertain award threshold. Miklós-Thal and Ullrich (2015) and Miklós-Thal and Ullrich (2016) study the effort incentives of players with differential prospects in a model in which career concerns and tournament are rolled into one in a one-period tournament model in which the player with the higher expected ability wins. This paper builds on this literature and extends it by introducing a risk choice, as in Hvide (2002), Kräkel and Sliwka (2004), Kräkel (2008). This allows for the analysis of not just effort but also the risk of the work undertaken, and thus, the novelty of output. Furthermore, the theoretical analysis yields hypotheses about heterogeneous effects by (expected) ability type; of effort and risk choice, as well as selection into performance pay.

[1] This is the same data set as used in Ytsma (2017).

In order to empirically estimate the effort effect of performance pay in academia, I use the fact that any appointment after the implementation of the reform necessarily falls under the new performance pay scheme, while any existing contract continues to fall under the old age-related pay scheme. Academics who start their first tenured affiliation just before the reform therefore fall under the age-related pay scheme, while those who start their first tenured afiliation just after the reform are paid according to the performance pay scheme. If the timing of the start of the first tenured affiliation is exogenous, any difference in productivity from before to after the reform between academics who start their first tenured position just before the reform and those who start a first tenured position directly after the reform can be interpreted as the causal effect of performance pay on effort. I estimate this effort effect in a difference-in-differences framework and find evidence of a positive effort effect of at least 12 to 16% on average. This effort effect manifests itself as of the announcement of the reform, most robustly as an increase in research quantity (number of publications), and persists for several years.

I find no evidence of pre-existing trends, which lends support to the identifying parallel trends assumption of the difference-in-differences estimation. Furthermore, I perform a number of tests to assess the validity of the assumption that the timing of the tenure decision is exogenous. I find no effect of the performance pay reform on academic effort in a placebo difference-in-differences, where I use two cohorts of academics who make tenure before the performance pay reform as placebo treatment and control. Since both these cohorts fall under the age-related pay scheme, the pay reform should not have any differential effect on their productivity. If, however, either lower productivity academics or higher productivity academics were able to speed up the tenure process, and if their productivity growth would slow down, respectively speed up after tenure, the placebo difference-in-differences would return a negative, respectively positive effort effect. I find no evidence of this.

The average effort effect estimate hides a considerable amount of heterogeneity. Less productive academics show the strongest and most robust effort response to performance pay incentives compared to more productive academics. When measured in terms of individual-level productivity measures, the effort response of academics in the bottom three quartiles of the productivity distribution takes the form of a 18 to 27% increase in both pure quantity and quality-adjusted research output. In contrast, top quartile academics increase their research output quantity, by about 15%, but the average impact of their work decreases by around 17%.

In order to analyze whether performance pay incentivizes valuable knowledge work, I use textual analysis techniques to construct metrics that capture the novelty and impact of papers published. These metrics are based on cosine similarities of vector representations of the abstracts of papers, similar to Kelly et al. (2018). I find that the novelty and impact of published academic work does not change in response to performance pay, on average. However, when looking at the novelty and impact responses by productivity quartile, I find that the most novel work of top quartile academics becomes less novel, while their least novel work becomes more novel. The most impactful papers published by treated top quartile academics become even more impactful, as they give rise to about 18% more follow-on research.

The second part of the paper is devoted to estimation of the selection effect of performance pay in academia. Academics who hold a tenured position when the reform is implemented switch from the age-related pay system to the performance pay system when they change their affiliation or position. Because pay no longer increases with age in the performance pay system but only with productivity, I expect more productive professors to be more likely to select into the performance pay scheme by changing position or affiliation. Hazard rate and survival function analyses confirm that more productive academics are indeed more likely to switch to the performance pay scheme. Moreover, an academic's productivity has a greater effect on his probability of selecting into performance pay if he is younger, in line with the performance pay scheme being relatively less attractive for older academics because of a larger difference in basic wage between the age-related and performance pay scheme.

To distinguish between general incentives to change affiliation or position and incentives to select into performance pay specifically, I estimate differential changes in switching hazard rates by age and cohort for academics with a different average productivity. Here I use the fact that only academics who already hold a tenured affilitation before the reform switch to the performance pay scheme when they change their affiliation, position or contract, while academics tenure after the reform fall under the performance pay scheme from the start. Accordingly, I designate those academics as treated cohort in a difference-in-differences framework and compare their differential hazard rates by age with those of academics who make tenure after the reform. I find that older academics in the treated cohort are less likely to switch, but this differential age effect only occurs for relatively less productive academics. Accordingly, more productive academics are more likely to select into performance pay.

While the effort and selection effect of performance pay in academia both contribute to an increase in research output, they also partially counteract each other. The effort effect is strongest for relatively less productive academics, yet it is the more productive academics who select into performance pay. Incentive systems that trigger effort and selection effects that would reinforce each other, by attracting more productive academics and inspiring the greatest effort response from those same academics would be more efficient.

This paper builds on the literature on university governance (cf. McCormack et al. (2014), Haeck and Verboven (2012), Aghion et al. (2010), Belenzon and Schankerman (2009), Lach and Schankerman (2004)) and incentives for innovators (Hvide and Jones 2016, Borjas and Doran 2015, Ederer and Manso 2013, Azoulay et al. 2011, Manso 2011, Lach and Schankerman 2008) in particular, and the literature on the economics of science (Stephan 1996) and the organisation of knowledge creation (Grigoriou and Rothaermel 2014, Phelps et al. 2012, Jones 2009, Wuchty et al. 2007, Singh 2005, Audretsch and Feldman 2004, 1996, Jaffe et al. 1993)) more generally. By studying the effort and selection effects of performance pay in academia, this paper studies one of the key components of human resource management, incentive systems, (cf. Lazear and Oyer (2012), Bloom and Reenen (2011), Oyer and Schaefer (2011), Lazear and Shaw (2007)) in relation to university governance, thus building on and adding to the vast literature on incentives in organizations as well as that on university management and the organisation of knowledge creation.

Within the body of literature on incentives, numerous papers have examined how incentives affect worker productivity empirically, and many of those report significant positive effort effects of higher-powered incentive schemes, both in the field (e.g. Muralidharan and Sundararaman (2011), Shi (2010), Lavy (2009), Bandiera et al. (2005), Shearer (2004), Lazear (2000a)) and in the lab (i.a. Hossain and List (2012), Boly (2011), Dohmen and Falk (2011), Carpenter et al. (2010), Freeman and Gelber (2010), Bellemare et al. (2010), Ariely et al. (2009), Dickinson and Villeval (2008), Dickinson (1999)). The kinds of higher-powered incentive schemes that have been shown to have positive effort effects are several; from piece-rate pay (cf. Dohmen and Falk (2011), Shi (2010), Bellemare et al. (2010), Ariely et al. (2009), Bandiera et al. (2005), Shearer (2004), Lazear (2000a), Dickinson (1999)) or bonus pay (i.a. Hossain and List (2012), Muralidharan and Sundararaman (2011), Lavy (2009)) to tournament schemes (cf. Carpenter et al. (2010), Freeman and Gelber (2010), Harbring et al. (2004)) and monitoring regimes (e.g. Boly (2011), Dickinson and Villeval (2008)).

Most of the aforementioned papers study the effect of performance incentives in relatively routine tasks, such as car window replacement (Lazear 2000a), tree planting (Shearer 2004) or fruit picking (Bandiera et al. 2005). Though some papers study the effort effect of performance pay in education, either by estimating the effect of teacher incentives (cf. Muralidharan and Sundararaman (2011), Glewwe et al. (2010)) or student incentives (i.a. Bettinger (2012), Leuven et al. (2011), Angrist and Lavy (2009), Angrist et al. (2009)), only few study the effect of performance pay on academics, particularly with respect to research productivity. It is not obvious that performance pay schemes would increase academic research output. Academics are thought to be intrinsically motivated (McCormack et al. 2014, Sauermann et al. 2010, Stephan 1996), and extrinsic incentives might crowd out this intrinsic motivation (Dickinson and Villeval 2008, Besley and Ghatak 2005). Furthermore, innovation and scientific creativity might require particular incentive structures; such as rewarding long-term success and tolerating early failure (Ederer and Manso 2013, Azoulay et al. 2011, Manso 2011) .

Of the papers that study the effect of incentives on innovation and knowledge creation, a great number look at incentives for commercializable knowledge (patenting) (Hall and Harhoff 2012). Hvide and Jones (2016) for instance study the abolition of the so-called "professors' privilege" in Norway, which granted professors the full rights to their innovations. They find that patenting rates and entrepreneurship declined after the abolition. Lach and Schankerman (2008, 2004) report a positive relationship between royalty rates and university licensing income in the US, and Azoulay et al. (2009) show that patenting is associated with higher rates of publication and weakly with higher quality of publications. Most of the papers that study incentives for academic research focus on either funding or awards. Azoulay et al. (2011) find that recipients of grants that reward longer-term success and are lenient towards early failures produce more high-impact papers. And while Borjas and Doran (2015) find a decline in productivity of Fields Medal winners after the award, Chan et al. (2014) report that receiving the John Bates Clark Medal or Fellowship of the Econometric Society is associated with a higher subsequent research productivity. These papers cannot distinguish between the effect of funding itself and the incentive effect of funding or awards on output. Furthermore, they cannot distinguish between the effort effect of performance incentives and selection into the funding or award schemes. Yet it is likely that there are significant positive selection effects. Dohmen and Falk (2011) and Lazear (2000a) for instance study the selection effect of piece-rate schemes for workers in a lab experiment and in the field for windshield installers, respectively, and find that higher productivity workers self-select into the higher-powered pay scheme. Leuven et al. (2011) report a similar finding for selection in tournament schemes for students. The contribution of this paper is therefore to seperately and causally identify the effort and selection effects of implicit and explicit performance incentives for scientific knowledge production.

The paper is structured as follows: the next section provides an institutional background regarding the academic pay reform and the academic sector in Germany and section 3 presents a simple theoretical framework. The empirical analysis and core of the paper make up section 4, with the first part focusing on the effort effect and the second part on the selection effect of performance pay. Section 5 concludes.

2 Institutional Background

The German academic pay reform that I exploit as a natural experiment in this paper introduced a new pay scheme ("W-pay") under which professors can earn performance related bonuses on top of a basic wage[2] (BMBF 2002). These performance related bonuses can be substantial; potentially more than doubling a professor's monthly wage. Before the reform, professors were paid according to an age-related pay scheme ("C-pay") in which pay increased every two years until the age of 49 (Hochschullehrerbund 2009, Oeffentlicher-Dienst 2004, Expertenkommission 2000). For all but the youngest professors, the basic wage of the performance pay scheme is lower than the age-related wages (Hochschullehrerbund 2009, Oeffentlicher-Dienst 2004), but the total pay under the performance pay scheme can exceed that under the age-related pay scheme if an academic is paid large performance bonuses (Detmer and Preissler 2006).

2.1 Performance Pay (W-Pay)

There are three basic pay levels in the performance pay scheme: W1 pays a basic monthly wage of 3405.34 euro, W2 3890.03 euro and W3 4723.60[3] (Detmer and Preissler 2006, Oeffentlicher-Dienst 2004). As mentioned above,

[2] I exploit the same pay reform in Ytsma (2017).

[3] These were the basic wage levels under the performance pay scheme in former West-German states as of 1 August 2004. The corresponding basic wage levels in former East-German states were 92.5% of the West-German rates (Detmer and Preissler 2006).

professors can earn performance bonuses on top of these basic wages in the W-pay scheme (BMBF 2002). It is important to note that universities themselves decide how to award these performance bonuses [4], though the federal law governing the pay reform provides ground rules. Performance bonuses can be awarded on the basis of three grounds: as attraction or retention bonus, for on-the-job performance, and for taking on management roles or tasks (BMBF 2002).

The first kind of bonus, the attraction or retention bonus, is determined as part of contract (re)negotiations and generally awarded on the basis of a professor's qualifications and past achievements and performance, taking into account applicant pool quality and labour market tightness (Detmer and Preissler 2005). These bonuses therefore constitute implicit performance pay incentives, and they take effect from the moment the reform is announced. Many states give universities the option to award attraction or retention bonuses on a permanent basis, for a fixed term (initially) or even as a one-off payment (Detmer and Preissler 2004, 2005). If retention or attraction bonuses are awarded for a fixed term with the option of renewal, universities generally enter into a target agreement with the respective professor. The target agreement specifies the achievements, such as the number and type of publications and grants, that are expected of the professor in a 3- or 5-year period. If these targets are met, the attraction or retention bonus continues to be paid, either for another 3- to 5-year period, or permanently. The target agreements constitute explicit performance incentives that take effect only after the reform is implemented, for professors that fall under the performance pay scheme. State laws and university statutes generally do not impose a maximum on the amount paid as attraction or retention bonus.

Secondly, the pay reform introduced bonuses for on-the-job performance. These bonuses can be awarded for performance in research, art, teaching, mentoring and supervision (BMBF 2002). To assess research performance for instance, universities take into account the number and quality of publications, research prizes, patents and external research grants, while for instance exceptional teaching evaluations can serve as evidence of special teaching achievements (Detmer and Preissler 2005). Most state laws stipulate that the performance for which on-the-job bonuses are awarded must have been effected over multiple years, and allow universities to award these bonuses either for a fixed term (with the option to renew) or permanently (Handel 2005). The on-the-job performance bonuses are explicit performance pay incentives that take effect after the implementation of the reform, and only for professors who fall under the performance pay scheme.

Most universities pay on-the-job bonuses through what is essentially a promotion tournament (the so-called "Stufenmodell") (Harbring et al. 2004, Kräkel 2006, Lünstroth 2011). This system comprises a number of performance or hierarchy levels, each of which is associated with a bonus (Lünstroth 2011). Generally speaking, universities announce at the beginning of a year either both the number of levels and associated bonus pay or only the number of promotions to higher levels (if the associated bonus pay is specified in the university's statutes) to be awarded in that year (Lünstroth 2011). There is subtantial variation in the number of pay levels across universities; the number of levels ranges from 2 (e.g. Augsburg and Erfurt University) to 10 (University of Trier), and the associated pay from 90 (Technical University of Berlin) to 2500 euro per month (e.g. Bielefeld and Bremen University) (Lünstroth 2011).

Some universities pay on-the-job bonuses through a relative performance pay system which is essentially a Japanese style (J-type) tournament ("Leistungspunkte Modell") (Kräkel 2003, Lünstroth 2011). In this system, a university announces the total amount available for on-the-job performance bonuses in a given year and awards professors points for achievements in research and education. Each academic then receives as performance bonus the share of the total "prize pot" that is equal to his relative performance that year.

The third kind of bonus in the performance pay scheme comprises pay supplements for taking on management tasks or roles (BMBF 2002). These bonuses are paid for the duration of the task or role only and generally range

[4] See Handel (2005) for a comprehensive overview of how much discretion higher education institutes have regarding hiring and pay decision after the reform in the different German states.

somewhere between 200 and 600 euro (for the dean) per month.

Finally, the reform also introduced the option for professors to extract a pay supplement from third-party awarded funds for research or teaching projects for the duration of such projects (BMBF 2002)[5]. Some states restrict the amount of these supplements to not exceed the basic wage of the professor (Detmer and Preissler 2005).

Only tenured professors[6] can earn bonuses in the performance pay scheme. Junior Professors, the German equivalent of assistant professors, can only earn a (non-pensionable) supplement of 260 euro per month upon positive evaluation[7] (Detmer and Preissler 2005). For comparison, tenured professors can earn performance bonuses up to a total amount of 5241,48 euro per month [8] - more than the highest basic wage in the performance pay scheme - or more in special circumstances[9] (BMBF 2002, Detmer and Preissler 2005). Consequently, I restrict attention to tenured professors when analyzing effort and selection effects of the pay reform in this paper.

2.2 Comparison With Age-Related Pay (C-Pay)

There are four pay levels in the age-related pay scheme (C1-C4); university professors are generally awarded C3 or C4 pay. The monthly salary in these levels increases every two years by roughly 170 euro[10], from the age of 21 to the age of 49 (Hochschullehrerbund 2009, Oeffentlicher-Dienst 2004, Expertenkommission 2000). In contrast, the basic wage in the W-pay scheme does not vary with age, and the level is such that professors earn a higher before-bonus wage in the performance pay scheme at first, but once they get older, they would earn a higher basic wage in the C-pay system. Figure 1 depicts the monthly wage by age for the several pay levels in the performance pay and age-related pay schemes. As is clear from this figure, the crossing point of these basic wage schedules depends on the specific (offered) pay level of an academic[11]; the age-related wage starts to exceed the basic wage in the performance pay scheme either at age 33[12] or at age 43[13] (Oeffentlicher-Dienst 2004, Handel 2005).

Before the pay reform, professors in the highest pay level of the age-related pay scheme (C4) could earn bonuses when they received offers after their first appointment as C4-professor. These bonuses were standardised to be around 650 euro for the second C4-offer, and about 730 euro for the third C4-offer from another university, and roughly 75% of this if a counter-offer of the home university was accepted (Detmer and Preissler 2006, Preissler 2006, Dilger 2013). By comparison, the average attraction and retention bonus in the W-pay system had already grown to 1187 euro in 2006, and the average on-the-job performance bonus to 1649 euro (BMI 2007). Furthermore, only a small fraction of professors qualified for and received bonuses under the age-related pay system. Handel (2005) for instance calculates, using data from the Ministry of Science and Culture in Niedersachsen, that only about 16.5% of professors received attraction or retention bonuses in the age-related pay system. In contrast, any tenured professor in the performance pay system can receive bonuses, and already in 2006 about 77% of professors in the performance pay scheme did receive bonuses (BMI 2007). Consequently, only about 3.55% of the total professorial pay volume was spent on attraction and retention bonuses in the age-related system, before the reform (Handel (2005), using data from Expertenkommission

[5] These supplements were intended to motivate professors to take on activities as part of their academic job that they may have otherwise performed on the side (Handel 2005)

[6] There are two tenured professorial ranks in Germany; the equivalent of an associate professorship (“ausserordentliche (or a.o.) Professur”), and the equivalent of a full professorship (“ordentliche (o.) Professur”).

[7] In special cases, Junior Professors can earn an additional supplement per month, which is not to exceed 10% of a junior professor’s basic wage (Detmer and Preissler 2005).

[8] This limit is set at the difference between the basic wage of W3 and B10 (another, non-professorial pay scheme), which was 5241.48 on 1 August 2004 (Detmer and Preissler 2005)

[9] If the academic already earns bonuses that exceed this limit and a higher bonus is necessary to attract the academic to another German university or prevent him from wandering off to another German university (BMBF 2002).

[10] Using pay tables valid as of August 2004 (Hochschullehrerbund 2009)

[11] C3 or C4 in the age-related system; W2 or W3 in the performance pay system.

[12] When comparing W2 with C3 or W3 with C4

[13] When comparing W3 with C3, a transition which is allowed in most states (Handel 2005).

(2000)), while an estimated 26% of the professorial pay volume was available for performance bonuses under the performance pay scheme immediately after the reform(Handel 2005). Combined with the fact that, at most ages, the basic wage is lower in the performance pay system than in the age-related system, this means that a larger portion of professorial pay depends on performance and there is a greater spread in professorial pay in the W-pay system. The W-pay system therefore offers higher-powered performance incentives than the old, age-related pay system.

2.3 Implementation

The federal law introducing the new professorial pay scheme was passed by Germany's parliament in February 2002 and applies to all public institutions of higher education.. The law required all states to implement the reform within their respective jurisdiction latest 1 January 2005 and only three states (Bremen, Niedersachsen and Rheinland-Pfalz) did so before the end of 2004 (Detmer and Preissler 2005). Hence any new or renegotiated professorial contract entered into as of 1 January 2005 falls under the performance pay scheme. Importantly, this means that, once a professor switches to performance pay by renegotiating his current contract or changing position or affiliation after 2004, he can never go back to age-related pay (Detmer and Preissler 2004). Professors do not have to wait for outside offers in order to be able to renegotiate their contract and switch to performance pay; they can opt into the performance pay scheme at any time after the reform's implementation. Detmer and Preissler (2005) however report that few professors have exercised this option.

The academic pay reform was mandated to be cost-neutral. In particular, the average professorial pay at the federal and state level was to remain at the respective pre-reform levels (benchmark year 2001) (BMBF 2002)[14]. In many states, the state's ministry of education implements the cost-neutrality requirement by calculating university-specific professorial pay averages that are used as benchmark professorial pay average for the respective university (Handel 2005). The law does allow for the benchmark professorial pay average to be exceeded by, on average, 2% per year, though not exceeding 10% in total[15] (BMBF 2002). The budget-neutrality stipulation was explicitly introduced to prevent the reform leading to cost-cutting or a cost explosion (Detmer and Preissler 2006, Handel 2005). Because the basic wage in the performance pay system is lower than most of the age-related wages, the cost neutrality requirement guarantees that whatever is saved on basic wage payments, is paid as bonuses in the performance pay scheme. Handel (2005) calculates that, with a pre-reform professorial pay average of 71.000 euro at universities, about 26% of total professorial pay for university professors is available for performance pay bonuses[16].

The next section presents a model that reflects the particular features of the German academic pay reform. It is used to derive testable hypotheses for the effects on effort and selection that the reform's performance pay incentives can be expected to have.

3 Theoretical Framework

Consider an academic who works for two periods; as untenured scholar in the first period, and as tenured professor in the second. In each period, an academic's output depends on their effort e, ability θ and a random noise draw ε, according to:

$$y_{i,t}\left(e_{i,t},\sigma_{\varepsilon,i,t};\theta_i\right)=\theta_i+e_{i,t}+\varepsilon_{i,t} \tag{1}$$

[14] The law does allow states to raise their target average professorial pay level to - at most - the highest average professorial pay at the state or federal level.

[15] As long as the state's budget allows.

[16] For this calculation, Handel (2005) uses 2001/2002 data and assumes that the ratio of W2 to W3 professors at universities will be about the same as that of C3 to C4, namely 46:54.

Ability is not know either by the academic or the market, but there is common knowledge about the ability distribution. In particular, an academic's ability θ_i is an iid draw from a distribution $f(.)$ with support $[\theta_{min}, \theta_{max}], \theta_{min} \geqq 0$. The noise terms are assumed to be iid draws from a normal distribution $g(.)$ with mean zero and variance σ_ε^2.

I assume that each period consists of two phases. In the first phase, academics decide the risk level for their production process by choosing the variance $\sigma^2_{\varepsilon,i,t} \in \left[\sigma^2_{min}, \sigma^2_{max}\right], \sigma^2_{min} > 0$ for the noise distribution $g(.)$. In what follows, I will refer to this variance as the 'risk level' that an academic chooses. This risk level operationalizes the notion of novelty in this model. A higher risk level increases the likelihood of both very high output values (novel, seminal papers), and very low output values (failed projects, derivative papers). In the second phase, academics decide how much effort to exert. Effort is costly and is restricted to be non-negative. The cost of effort is given by a convex cost function $C(e_i)$, with $C', C" > 0$, and $C(0) = 0$.[17] Risk and effort choices are private knowledge. Academics are risk-neutral, expected pay-off maximizers who discount future pay-offs by a factor $\delta, 0 < \delta < 1$. [18]

In what follows I first solve the model for the case of age-related pay. This will serve as a benchmark for the performance pay case, which I present next.

Age-Related Pay

The wages in the age-related pay scheme increase with age, so $0 < w_t^a < w_{t+1}^a$[19]. Age-related wages do not vary with an academic's performance. The expected life-time utility of academic i from the point of view of the first period is then:

$$E[U_i] = w_1^a - C(e_{i,1}) + \delta\left[w_2^a - C(e_{i,2})\right] \tag{2}$$

Because wages depend neither on output (and hence, effort) nor risk in either, academics exert no effort and are indifferent between any risk level in both periods. Even though all academics exert the same effort level - namely, no effort, their output will differ across individuals because of differences in ability and noise draws, and across periods because of differences in noise draws. This means that we would expect to find differences in measures of research output both across and within individuals empirically, with more able individuals producing systematically more output than less able individuals.

Performance Pay

In the performance pay scheme, an academic's basic wage no longer depends on their age, but any tenured professor can earn performance pay bonuses. As discussed in the previous section, the performance pay bonuses come in two forms: as attraction and retention bonuses and on-the-job performance tournament bonuses. I model the attraction and retention bonuses as career concerns incentives as in Holmström (1982, 1999) and Gibbons and Murphy (1992). The attraction and retention bonuses are determined as part of contract (re)negotiations, and are based on an academic's qualifications and past performance. Their structure thus aligns with market-based incentives that arise because an academic's past performance is used to update beliefs about the academic's ability and the academic's wage is based

[17] Many papers model heterogeneity in agent ability through different cost functions, with higher ability agents having a lower marginal cost of effort (Lazear 2000b, Schotter and Weigelt 1992, Bull et al. 1987). Alternatively, one could model ability as a higher marginal productivity, as in Gürtler and Kräkel (2010), Kräkel and Sliwka (2004), Chen (2003). Here, I follow Holmström (1982, 1999) and Gibbons and Murphy (1992) and model output as additively separable in effort and ability.

[18] Given that the risk and effort choices are private knowledge, the two phases in each period are purely for expositional clarity, as the risk and effort choices are effectively simultaneous choices.

[19] Strictly speaking, the age-related wages increase until the age of 49 only, so this should be a weak inequality (Hochschullehrerbund 2009, Oeffentlicher-Dienst 2004, Expertenkommission 2000). For the purposes of the analysis here however, all that matters is that age-related wages are lower, including relative to the basic wages under performance pay, in earlier periods.

on these beliefs. I model the on-the-job performance pay bonuses as a tournament as in Lazear and Rosen (1981), but with a risk choice as well as an effort choice as in Hvide (2002), Kräkel and Sliwka (2004), Kräkel (2008). Taken together, academics in the performance pay scheme then face implicit incentives from the career concerns, and explicit performance incentives from the tournament[20].

I assume that each university hires two tenured professors, who compete against each other in the on-the-job performance tournament. In a simple two-player tournament, academic i wins the tournament if $y_{i,t} > y_{j,t}$. This happens with probability

$$Pr\left(\theta_i + e_{i,t} + \varepsilon_{i,t} > \theta_j + e_{j,t} + \varepsilon_{j,t}\right) = G_{\varepsilon_j - \varepsilon_i}\left(\theta_i + e_{i,t} - \left(\theta_j + e_{j,t}\right)\right) \tag{3}$$

where $G_{\varepsilon_j - \varepsilon_i}(.)$ is the cdf of $\varepsilon_j - \varepsilon_i$ and $\varepsilon_j - \varepsilon_i \sim g\left(\varepsilon_j - \varepsilon_i\right)$.[21] The distribution $g\left(\varepsilon_j - \varepsilon_i\right)$ is symmetric around zero and has variance $\sigma^2_{\varepsilon,ij,t} = \sigma^2_{\varepsilon,i,t} + \sigma^2_{\varepsilon,j,t}$ because ε_i and ε_j are iid[22]. Denote the on-the-job performance bonus that can be won in the tournament by $b > 0$ and let the loser prize be 0. Universities then offer tenured professors performance pay contracts of the following form: $w^p_{i,t} = c_{i,t} + b * G\left(\theta_i + e_{i,t} - \left(\theta_j + e_{j,t}\right)\right)$. Recall that only tenured professors can earn performance pay bonuses, so this contract determines only second period pay for academics who fall under the performance pay scheme. In the first period, universities offer untenured scholars contracts of the following form: $w^p_{i,1} = c^p_{i,1}$.

The timing of the game is as follows: at the beginning of the first period, universities simultaneously offer first period (untenured) contracts of the form outlined above. Academics choose the contract that maximizes their utility and then produce knowledge. At the end of the period, everyone's output becomes common knowledge, and universities offer second period (tenured) contracts. Academics again choose the contract that mazimizes their utility and then produce knowledge. I assume that it is not possible to write long-term contracts, so every contract ends at the end of a period. I solve for the Bayesian Nash equilbrium in pure strategies by backward induction.

The expected utility in the performance pay scheme of academic i, from the point of view of the first period, is then given by:

$$E\left[U_i\right] = E_{\theta_i,\theta_j}\left[c^p_1 - C\left(e_{i,1}\right) + \delta\left[w_{i,2} + b * G\left(\theta_i + e_{i,2} - \left(\theta_j + e_{j,2}\right)\right) - C\left(e_{i,2}\right)\right]\right] \tag{4}$$

where E_{θ_i,θ_j} denotes the expectation over ability types for i and j. In period two, an academic maximizes the following expected utility with respect to second period effort and risk:

$$E\left[U_i\right] = E_{\theta_i,\theta_j}\left[w_{i,2} + b * G\left(\theta_i + e_{i,2} - \left(\theta_j + \hat{e}_{j,2}\right)\right) - C\left(e_{i,2}\right) | y_1\right] \tag{5}$$

Here $\hat{e}_{j,2}$ is the effort level that i conjectures j will exert.

The first order condition for optimal second period effort for academic i is then

$$E_{\theta_i,\theta_j}\left[b * g\left(\theta_i - \theta_j + \left(e_{i,2} - \hat{e}_{j,2}\right)\right) | y_1\right] = C'\left(e_{i,2}\left(y_1\right)\right) \tag{6}$$

while for their competitor j, it is

$$E_{\theta_i,\theta_j}\left[b * g\left(\theta_j - \theta_i + \left(e_{j,2} - \hat{e}_{i,2}\right)\right) | y_1\right] = C'\left(e_{j,2}\left(y_1\right)\right)$$

[20] The explicit performance incentives of the performance pay scheme also comprise target agreements that govern the continuation of bonus payments. Because these target agreements generally take the form of standard-setting agreements, where a bonus is paid if performance meets a certain standard, the incentives effects of these agreements are very similar to that of the tournament (Cf. Gibbs (1994)).

[21] To improve legibility, I will leave out the subscript for distribution functions in what follows.

[22] This follows the specification of the winning probabilities in a rank order tournament in Lazear and Rosen (1981).

Noting that $g(x) = g(-x)$ because $g(.)$ is symmetric around 0, it follows that, for every pair of competitors, there is a unique set of equilibrium efforts; namely, for each academic to exert the same effort as their opponent. Formally, for every pair of competitors i, j: $e^*_{i,2}(y_1) = e^*_{j,2}(y_1) = e^*_2(y_1), s.t.:$

$$E_{\theta_i,\theta_j}[b * g(\triangle\theta)|y_1] = C'(e^*_2(y_1)) \tag{7}$$

where I am writing $\triangle\theta$ for $\theta_i - \theta_j$[23].

The first order condition for optimal second period risk for academic i is

$$E_{\theta_i,\theta_j}\left[b * \frac{\partial G}{\partial \sigma_{\varepsilon,i,2}}\left(\theta_i + e_{i,2} - \left(\theta_j + \hat{e}_{j,2}\right)\right)|y_1\right] \tag{8}$$

The first order condition for colleague/competitor j is again symmetric. Using that $e^*_{i,2}(y_1) = e^*_{j,2}(y_1) = e^*_2(y_1)$ in equilibrium, this first order condition becomes:

$$E_{\theta_i,\theta_j}\left[b * \frac{\partial G}{\partial \sigma_{\varepsilon,i,2}}(\triangle\theta)|y_1\right] \tag{9}$$

I assume that there is perfect competition in the labor market. This implies that universities earn no profits, and hence that academics must be paid the value of their expected output. Setting the unit price for research output to 1, the non-contingent[24] portion of second period pay of an academic (the sum of the basic wage and attraction and retention bonus) must then be equal to

$$w_{i,2} = E_{\theta_i,\theta_j}[\theta_i|y_1] + e^*_{i,2}(y_1) - E_{\theta_i,\theta_j}[b * G(\triangle\theta)|y_1] \tag{10}$$

where I have used that $E_{\theta_i,\theta_j}[y_{i,2}|y_1] = E_{\theta_i,\theta_j}[\theta_i|y_1] + e^*_{i,2}$ in equilibrium. The expectation of i's ability, conditional on first period output is given by

$$E_{\theta_i,\theta_j}[\theta_i|y_1] = \int_\theta \theta f(\theta|y)\, d\theta \tag{11}$$

This conditional expectation differs from the one used in Holmström (1982, 1999), Gibbons and Murphy (1992), which is based on formulas from DeGroot (1970), because I assume that ability types are non-negative and distributed according to a general function $f(.)$ with support $[\theta_{min}, \theta_{max}]$. Holmström (1982, 1999), Gibbons and Murphy (1992) assume that ability is normally distributed, and hence that ability types can take on any value on the real line and are symmetrically distributed around a mean m_0. As I will show in the empirical section, the distribution of productivity types of academics is highly skewed, so the normal distribution does not appear to be appropriate here. I restrict ability types to be non-zero, because this aligns with the measures of productivity that I observe and use in the empirical section, and it allows for a more intuitive interpretation of ability types.

Using this, the expected second-period utility of an academic, conditional on first period output can then be written as

$$E[U_{i,2}|y_1] = E_{\theta_i,\theta_j}[\theta_i|y_1] + e^*_{i,2}(b, y_1) - C\left(e^*_{i,2}(b, y_1)\right) \tag{12}$$

When there is perfect competition in the labor market, universities will set the winner prize b such that it maximizes the above expected utility. That is, $b^*$is such that

$$\frac{\partial e^*_{i,2}(b, y_1)}{\partial b} = C'\left(e^*_{i,2}(b, y_1)\right)\frac{\partial e^*_{i,2}(b, y_1)}{\partial b}$$

[23] Here I assume that the second order condition, $b * g'\left(\theta_i - \theta_j + \left(e_{i,2} - \hat{e}_{j,2}\right)\right) < C''(e_{i,2})$, is satisfied for all possible ability types and equilibrium effort and risk choices.

[24] Not contingent on second period output, though, as is shown below, this part of second period pay does depend on first period output.

or

$$\frac{\partial e^*_{i,2}(b,y_1)}{\partial b}\left[1-C'\left(e^*_{i,2}(b,y_1)\right)\right]=0 \tag{13}$$

Substituting $E_{\theta_i,\theta_j}[b*g(\triangle\theta)|y_1]$ for $C'\left(e^*_{i,2}(b)\right)$ from the first order condition for equilibrium second-period effort (7), and implicitly differentiating (7) to get

$$\frac{\partial e^*_{i,2}(b,y_1)}{\partial b}=\frac{E_{\theta_i,\theta_j}[g(\triangle\theta)|y_1]}{C''\left(e^*_{i,2}(b,y_1)\right)}>0 \tag{14}$$

we can write the equilibrium winner prize as

$$b^*=\frac{1}{E_{\theta_i,\theta_j}[g(\triangle\theta)|y_1]} \tag{15}$$

Importantly, and in contrast to e.g. Gibbons and Murphy (1992), the optimal explicit incentives in the second period do depend on first period output here, through the expectation of the difference in abilities of a pair of colleagues/contestants. Substituting (15) for b in 7 and using the law of iterated expectations, we see that, in equilibrium, second period effort is such that

$$1=C'\left(e^*_2(y_1)\right) \tag{16}$$

That is, the optimal winner prize induces efficient second period effort[25].

Substituting (15) for b^* and (10) and (11) for $w_{i,2}$ in (4) and using the law of iterated expectations, we can rewrite academic i's expected utility from the point of view of the first period as

$$E[U_i]=w^p_1-C(e_{i,1})+\delta E\left[E[\theta_i|y_i]+e^*_{i,2}(b,y_1)-C\left(e^*_{i,2}(b,y_1)\right)\right] \tag{17}$$

The first order condition for optimal first period effort is then given by

$$C'(e_{i,1})=\delta E\left[E[\theta_i|y_i]+\left(1-C'\left(e^*_{i,2}(b,y_1)\right)\right)\left(\frac{\partial e^*_{i,2}(b,y_1)}{\partial b}\frac{\partial b}{\partial e_{i,1}}+\frac{\partial e^*_{i,2}(b,y_1)}{\partial y_1}\frac{\partial y_1}{\partial e_{i,1}}\right)\right] \tag{18}$$

Using (16), the first order condition for first period effort reduces to

$$C'(e_{i,1})=\delta\frac{\partial}{\partial e_{i,1}}E[\theta_i|y_i] \tag{19}$$

That is, optimal first period effort is driven by career concerns only, as in Gibbons and Murphy (1992). Even though the optimal winner prize in the second period tournament depends on first period output, it is set such that second period effort is efficient in equilibrium, and hence such that marginal cost equals marginal benefit. (Expected) marginal cost and marginal benefit are independent of first period output because of the additive separability of effort and ability in the production function. First period output thus affects second period utility only through its direct effect on the non-contingent pay portion $w_{i,2}$ (the sum of the basic wage and attraction and retention bonus), and hence through the direct effect of expected ability on second period utility only. Note also that, because the right-hand side of the first-order condition is a constant defined by the model's primitives only, optimal first period effort is the same for every academic.

Finally, the first order condition for optimal first period risk is

[25] This arises because academics are assumed to be risk-neutral here.

$$\delta E\left[\frac{\partial}{\partial\sigma_{\varepsilon,i,1}}\left(E\left[\theta_i|y_i\right]\right)+\left(1-C'\left(e^*_{i,2}\left(b,y_1\right)\right)\right)\left(\frac{\partial e^*_{i,2}\left(b,y_1\right)}{\partial b}\frac{\partial b}{\partial\sigma_{\varepsilon,i,1}}+\frac{\partial e^*_{i,2}\left(b,y_1\right)}{\partial y_1}\frac{\partial y_1}{\partial\sigma_{\varepsilon,i,1}}\right)\right] \tag{20}$$

which, after some rearranging can be simplified to

$$\delta E\left[\frac{\partial}{\partial\sigma_{\varepsilon,i,1}}\left(E\left[\theta_i|y_i\right]\right)+\left(1-C'\left(e^*_{i,2}\left(b,y_1\right)\right)\right)\left(\frac{\partial e^*_{i,2}\left(b,y_1\right)}{\partial b}\frac{\partial b}{\partial\sigma_{\varepsilon,i,1}}+\frac{\partial e^*_{i,2}\left(b,y_1\right)}{\partial y_1}\frac{\partial y_1}{\partial\sigma_{\varepsilon,i,1}}\right)\right]=\delta E\left[\frac{\partial}{\partial\sigma_{\varepsilon,i,1}}\left(E\left[\theta_i|y_i\right]\right)\right] \tag{21}$$

The second term in the expectation on the left-hand side is 0 because in equilbrium, the explicit performance incentives in the second period are efficient, i.e. $1=C'\left(e^*_2(y_1)\right)$. The first order condition for optimal first period risk therefore depends on the direct effect of first period risk on utility through its effect on the conditional expectation of ability in the second period, and hence second period pay.

Testable Implications

Comparing the optimal effort and risk levels under the age-related (flat wage) pay scheme and the performance pay scheme yields a number of testable implications.

Proposition 1 - Effort Effect: *Equilibrium effort levels are larger in both periods under the performance pay scheme compared to the age-related pay scheme.* (Proof in Appendix A)

Corollary 1: *Academics in the performance pay scheme produce more output, both while tenured and untenured.*

In the age-related pay scheme, pay does not depend on output - either explicitly or implicitly. There is no reward for exerting effort, but there is a cost to providing effort. The optimal action is then not to exert any effort[26]. In the performance pay scheme on the other hand, untenured professors have an incentive to exert effort because they want to influence the beliefs about their ability so as to increase the attraction bonus they will be able to negotiate once they make tenure. In equilibrium this is of course ineffective, because all untenured academics exert effort for this reason - the same level no less - and the market correctly updates beliefs about ability from the outputs produced. Once tenured, academics have an incentive to exert effort to increase their chance of winning the on-the-job tournament. This exertion of effort is ultimately not swaying winning chances either, because within each pair of colleagues/competitors, the optimal second period effort level is the same. Across pairs, the equilibrium effort level however differs.

Proposition 2 - Heterogeneous Effort Effect: *The equilibrium effort level in response to explicit, tournament pay is lower for higher ability academics than for lower ability academics .* (Proof in Appendix A)

The optimal effort levels in the second period decrease as the on-the-job tournaments become more uneven, as the tournament will be determined more by the existing ability difference than by effort or noise. Accordingly, expected utility increases in the ability difference of competitors, though only for the more able of the competitors. Expected utility in the second period decreases in the ability difference for the less able of a pair of competitors. In equilibrium therefore, we should see positive assortative matching into tournaments, as the least able academic can do no better than match with the second least able academic, and so on. A companion paper, Ytsma (2017), provides empirical evidence of this. With a right-skewed distribution of abilities, the ability difference between academics who compete against each other in a tournament at the lower end of the ability distribution is smaller than the ability difference between positively assortatively matched academics at the higher end of the ability distribution. But this means that the equilibrium effort of more able academics will be smaller in response to the explicit tournament incentives than the effort of less productive academics.

[26] It is straightforward to moderate this result and derive an equilibrium with positive effort in the age-related pay scheme by e.g. assuming that academics are intrinsically motivated (derive private benefit from the output they create), or their outside option has a small, positive utility value.

Proposition 3 - Risk Choice - Novelty: *The equilibrium risk level in response to explicit, tournament pay is lower for higher ability academics than for lower ability academics. Untenured academics choose a minimum risk level if they fall under performance pay.* (Proof in Appendix A)

In response to explicit, tournament pay incentives, the lower ability academic of a pair of competitors prefers a high risk level. While a higher risk increases the chances of a very low output, it also increases the chances of a very high output. The disadvantaged academic needs such a high output to be in for a chance of winning the tournament. In contrast, the higher ability academics stand to lose relatively more by choosing a risky strategy, as a risky strategy increases the chance of a very low outcome, which could cost them the tournament.

In response to implicit, career concerns incentives academics opt for a lower risk level. Increasing risk reduces the informativeness of output as a signal of ability. A high output level is then more likely to be the result of a lucky draw rather than a more able academic. Unilaterally increasing risk reduces the market's expectation of own ability, while unilaterally decreasing risk increases the market's expectation of ability. Of course in equilibrium, everyone reduces risk and the market correctly updates beliefs about ability.

The chosen risk level could manifest itself both in the impact of research produced as well as the novelty of the research. As derived above, academics are expected to be indifferent between risk levels under age-related pay. If academics opt for an intermediate level of risk under age-related pay, we should see a shift towards lower impact or less novel work in response to implicit, career concerns incentives for academics that end up in the performance pay scheme.

Proposition 4 - Selection into Performance Pay: *Higher ability academics are more likely to select into performance pay.* (Proof in Appendix A)

As shown above, with performance pay, academics are paid the value of their conditional, expected output. If this expected value is greater than the wage they would have earned under the age-related pay scheme, an academic is better off in the performance pay scheme and will switch when given the chance.

The next section provides empirical tests of these propositions.

4 Empirical Analysis

4.1 Data Description

In order to study effort and selection effects of performance pay in knowledge creation, I constructed an individual level panel data set that encompasses the affiliations of the universe of academics in German academia for the years 1999-2013[27], as well as measures of their research productivity from 1993 onwards. The panel contains information about the academic position of academics in a given year, whether their position is tenured, the specific institution they are affiliated with, whether the institution is a university and if so, whether it is public or private. The data set also provides the year in which an academic completed their PhD, obtained their postdoctoral qualification, and started working in academia, as well as an academic's gender, birth year and, if applicable, year of passing. The productivity variables in the data set are based on an academic's publication record. All in all, the panel encompasses 55132 academics who held a tenured position at a German public university at some point between 1999 and 2013.

In this paper, I restrict attention to public universities only, of which there are 89 in Germany[28] (HRK 2014). I dis-

[27] This is the same data set used in Ytsma (2017)

[28] Number of public universities reported by HRK (2014) on 31 August 2014. The set of public universities that I consider for the empirical part of the paper is slightly different, because some institutions became universities only recently (the Hochschule Geisenheim University became a university on 1 January 2013 and the Hochschule für Film und Fernsehen Potsdam only in July 2014 for instance (HS-Geisenheim 2014, Filmuniversitaet 2014)) and are therefore not included in this set.

card higher education institutions other than universities, because I focus on research productivity as outcome variable, and the research output of universities is incomparable to that of other higher education institutions BMBF (2002). I further restrict attention to tenured professors, because performance bonuses can be earned in tenured positions only. .

I draw from three main, mostly unstructured, input data sets to construct the individual panel data set that I use in this paper; Kuerschners Deutscher Gelehrten Kalender, Forschung & Lehre Magazine and ISI Web of Science. Kuerschners Deutscher Gelehrten Kalender (hereafter: DGK) is a comprehensive encyclopedia of academics affiliated with German universities (Gruyter 2013, 2006, 2008). I use DGK as a register of the universe of academics affiliated with German universities and extract academics' personal information (full name, birth date, year of passing, gender) as well as professional information (academic affiliation at different points in time, start year of academic career in Germany, end year of academic career in Germany, self-reported information on career history) from it. I supplement the information in DGK regarding the timing of affiliation changes and the obtainment of postdoctoral qualifications with information from Forschung & Lehre Magazine (hereafter: FuL) (DHV 1999-2013). FuL is Germany's largest academic professional magazine. Every month, it publishes an overview of scholars that obtained their post-doctoral qualification (habilitation), as well as professorial offers that were extended, accepted or rejected. Finally, I extract publication records for the academics in my data set for the years 1993-2012 from the ISI Web of Science database to construct measures of research productivity. Some of the measures of productivity I use in this paper are based on weighted publication records, where I weigh each publication by the two-year impact factor of the journal in which it is published. I use the impact factors from the ISI journal citation report (JCR) of the year of publication[29].

I match academics across the three input data sets on the basis of last name, initials and field[30]. Doing so yields an 83% match rate of academics who FuL reports as having a tenured affiliation with a German university to academics listed in DGK. Differences in the spelling of names, typos and erroneous information regarding affiliation changes in FuL mostly explain the 17% that I cannot match. I have direct information on the timing of half the affiliation changes[31] in my panel data set from FuL. For the other half of affiliation changes, DGK provides the year of change in 23% of the cases[32] and I infer the timing of the remaining affiliation changes from academics' affiliations listed in DGK at different points in time, the year they passed their habilitation as well as the start and end year of their academic career in Germany recorded in DGK. A detailed description of the construction of the individual level panel data set can be found in (Ytsma 2017).

4.2 Effort Effect

In order to identify the pure effort effect of the introduction of performance pay in German academia on knowledge creation, I use the fact that any contract for a professorial position at a public university in Germany signed or renegotiated as of 1 January 2005 necessarily falls under the performance pay scheme, whereas any contract signed before this date falls under the old, age-related pay scheme[33]. Accordingly, academics who start their first tenured affiliation before

[29] Because I have ISI JCR data for the years 2000-2013 only, I use the average of the impact factors from JCR 2000 through JCR 2004 to weigh publications before 2000.

[30] FuL distinguishes 12 broad fields in which academics can be active, and categorises the information regarding affiliation changes and postdoctoral qualifications under these fields. I define departments along the same lines as FuL's fields and classify academics registered in DGK under these same categories for data matching purposes. The fields used for classification in FuL, and hence the departments that I distinguish, are: theology; philosophy and history; social sciences; philology and cultural studies; law; economics; mathematics, physics and computer science; biology, chemistry, earth sciences and pharmaceutics; engineering; agricultural sciences, nutrition and veterinary medicine; medicine (human); dentistry.

[31] Where at least one of the affiliations concerns a tenured position at a German university.

[32] This is self-reported career information and hence may introduce bias in my data set. I therefore use the information regarding affiliation changes provided in FuL whenever available. Reassuringly though, a consistency check revealed that the information in DGK regarding the timing of affiliation changes differs from that in FuL for only 5% of the individuals who change a (tenured) affiliation at least once.

[33] With the exception of Bremen, Niedersachsen and Rheinland-Pfalz, who introduced performance pay before this deadline (in 2003 and 2004, respectively) (Detmer and Preissler 2005). Note that using 2005 as uniform before-after cut-off yields a conservative measure of the effort effect, since some of the control group is, in fact, already treated before this time.

2005 continue to fall under the age-related pay scheme[34], whereas academics who start their first tenured affiliation after 2004 switch to the performance pay scheme upon starting their first tenured position. If the timing of the start of the first tenured position is exogenous, the performance incentives that first-time tenured affiliates face are exogenous as well. I can then identify the causal effort effect of performance pay on knowledge creation by comparing the change in research productivity from before to after the pay reform of academics who start their first tenured affiliation before 2005 (the control group) with the change in research productivity of academics who start their first tenured affiliation as of 1 January 2005 (the treatment group).

Unless indicated otherwise, I use a three-year window before and after the reform to define the treatment and control group for the analyses below in order to abstract from seniority effects. Thus the treatment group consists of academics who start their first tenured position at a public university in 2005, 2006 or 2007, while the control group consists of academics who start their first tenured affiliation at a public university in 2002, 2003 or 2004. I exclude academics who hold a foreign affiliation before their first tenured affiliation in Germany to avoid confounding the effort effect with selection effects of performance pay. The treated cohort comprises 2,844 academics, the control cohort 3,197.

I use a three-year window for my preferred cohort specification to abstract from any possible confounding effect of the introduction in 2002 of a new road to tenured professorships: the Junior Professorship. Traditionally, aspiring academics had to complete a post-doctoral qualification (the "habilitation") in order to qualify for a professorship. The habilitation phase generally lasts six years, during which time habilitands work as part of the research group of a full professor and write a postdoctoral thesis (Fitzenberger and Schulze 2014, Pritchard 2006). In 2002, together with the announcement of the pay reform, the German federal government introduced the equivalent of assistant professorships ("Juniorprofessur"), to supersede the habilitation[35](Pritchard 2006). Junior Professors enjoy greater academic independence compared to habilitands. Junior Professorships are equivalent to the Anglo-Saxon Assistant Professorship in this sense, though Junior Professorships are generally not tenure-track positions (Fitzenberger and Schulze 2014). Junior Professorships last 6 years, so the first academics to start a Junior Professorship in 2002/3 were eligible for their first tenured affiliation as of 2008. The three-year cohort window defined above is therefore the most comprehensive window definition that does not include any first-time tenured professors who completed a Junior Professorship instead of a habilitation, thus avoiding possible confounding effects of the Junior Professorships. Results are however robust to extending or reducing the cohort window (Table 9 panels A and B).

4.2.1 Descriptive Statistics

For the effort effect analysis I focus on measures of research productivity that are based on the publications of academic i in field f in year $t+x_f$, where x_f denotes the average publication lag in field f, rounded up to the nearest year. The average publication lags are taken from Björk and Solomon (2013) and range from 8 months for Chemistry to 18 months for Economics and Business. Correspondingly, I backdate publications by one to two years. After correcting for average publication lags I have productivity measures for 16 years, from 1993 through 2010; from 9 years before the announcement of the reform until 9 years after, and from 12 years before the implementation of the reform until 6 years after.

I use five measures of research productivity: number of publications, both raw counts and weighted by the impact factor of the journal in which the publication appeared[36][37]; sum of citations to publications; average impact-factor rating of publications; and average number of citations to publications. The citations data were last extracted in January 2019, so there are at least 6 years between the publication date and the time at which citations were counted

[34] They would be promoted to a higher pay grade in the age-related pay scheme when starting a tenured position (Detmer and Preissler 2004).
[35] The complete abolition of the habilitation was however declared unconstitutional and retracted (Pritchard 2006).
[36] The impact factors are taken from Thomson Reuters' Journal Citation Reports.
[37] These are the same measures used in Levin and Stephan (1991)

for each publication. The number of publications is used as a measure of research productivity quantity, the impact factor-rated number of publications is used as a quality-adjusted measure of total output quantity, and the number of citations as an indicator of the total impact or research quality. The average impact factor rating and average number of citations on the other hand are used to shed light on the quality or impact of the research output on average.

Table 1 Panel A reports summary statistics for these variables for the treatment and control cohort. On average, academics in the treatment and control cohort publish almost 3 papers per year. Accounting for impact factor of the outlet in which a publication appears brings this sum to almost 10. To put this in perspective, the latest two-year impact factor ratings, available for 2017, put the top five general interest journals in economics at an impact factor rating between 3.750 (Econometrica) and 7.863 (Quarterly Journal of Economics), while a top field journal like the Journal of Labor Economics had an impact factor rating of 3.607 (Journal Citation Report 2017). Impact-factor ratings do get considerably greater than this. Science had an impact factor rating of 41.058 and Nature of 41.577, for instance (Journal Citation Report 2017). The average total number of citations to publications from a given year is 102. The distributions of these variables are highly skewed: the median academic does not have any publication in any given year, while the most productive academics produce orders of magnitude more work by any of these measures.

The average impact factor rating of publications is defined for academics who have at least one publication in a given year only. Consequently, this variable is less skewed, with a mean of 2.7 and a median of 2.1, though there is still a substantial right tail. The average number of citations measure is similarly defined for academics with at least one publication in a year, but the distribution is more skewed. While the average publication has garnered almost 33 citations by the beginning of 2019, the median publication has only 19 citations. Furthermore, the maximum number of citations to a publication from a given year is 99 on average, with a median of 42, while the average of the minimum number of citations approaches 10, with a median of 1. That is, the majority of publications are referenced only modestly, but a few seminal papers are cited extensively - some thousands of times.

4.2.2 Baseline Difference-in-Differences

I estimate the effort effect in a parametric difference-in-differences model given by the following equation:

$$E\left[Y_{i,f,t-x_f}|X_{i,f,t}\right] = exp[\alpha_i + \beta_3 post'02 * Treatment_i \tag{22}$$

$$+\beta_5 tenured_{i,j} * Treatment_i + \sum_{j=-7}^{7} tenure_{i,j} + \gamma_t]$$

The dependent variable, $Y_{i,f,t-x_f}$, denotes any of the productivity measures of academic i in field f in year $t-x_f$, where x_f denotes the average publication lag in field f as defined above. The $Treatment$ variable is 1 for academics who start their first tenured affiliation at a public university in 2005, 2006 or 2007, and 0 otherwise. I restrict the sample to include only those academics who start their first tenured affiliation at a public university between 2002 and 2007 (i.e. the treatment and control cohort). Furthermore, I only include those academics who did not hold a foreign affiliation before their first tenured affiliation at a German public university, to abstract from any sorting effects here. The variable $post'02$ is 1 as of 2002 and 0 beforehand, and the variable $tenured$ is 1 as of the year in which an academic starts their first tenured affiliation at a public university and 0 beforehand. The $tenure_{i,j}$ variables are year-to-tenure dummies. They control flexibly for the seven years before and after academics start their first tenured position, as well as the tenure year itself. I control for 7 year-to-tenure dummies because the median number of years between the end of the PhD and the completion of the habilitation is 7[38] and academics are, traditionally, required to have completed

[38] Although the duration of the habilitation period is generally 6 years, the average age at which academics who later obtain a tenured affiliation

their habilitation before they become eligible for a tenured affiliation. Results are however robust to including more or fewer year-to-tenure dummies (Table 9). I also include individual fixed effects, α_i, and calendar year fixed effects, γ_t. Finally, because the dependent variables are highly skewed with a large mass at zero and long right tail, I estimate the model as a conditional quasi-maximum likelihood fixed-effect Poisson model[39]. The corresponding estimation results are shown in Table 2. Robust standard errors are reported throughout.

The $post'02$ variable is included to pick up on the effect of the announcement of the reform. The moment the reform is announced, the lure of future attraction and retention bonuses, and hence the implicit incentives of the performance pay scheme, takes effect. Given that the attraction and retention bonuses are awarded for i.a. past achievements in research, academics have an incentive to increase their research output from the moment the reform is announced so as to increase their chances of earning a (higher) attraction bonus once the reform takes effect. This incentive should be stronger for academics who anticipate starting their first tenured affiliation after the reform (the treated cohort) since the basic wage of the performance pay scheme is lower at most ages than the wage an academic would have earned in the age-related pay scheme. The incentive effect of the explicit on-the-job performance bonuses takes effect only after academics enter into the performance pay scheme. This entry into the performance pay scheme coincides with the start of the first tenured affiliation for academics in the treated cohort. By including both a $post'02$ and $tenured$ variable and their interactions with the $Treatment$ variable, the above specification allows for two separate trend breaks; at the moment the reform is announced and the associated implicit incentives take effect, and at the moment of entry into the performance pay scheme, when the scheme's explicit incentives start to play a role (for the treated cohort, at least). Thus the $post'02 * Treatment_i$ and $tenured * Treatment_i$ interaction terms taken together provide a difference-in-differences estimate of the total effort effect of implicit and explicit performance incentives in knowledge creation. Note that, because of the way the $post'02$ and the $tenured$ variable are defined, the interactions of these variables with the $Treatment$ variable capture *persistent* differential changes in the research productivity of the treated cohort relative to the control cohort after the announcement of the reform and entry into the performance pay scheme, respectively.

Baseline Results

Column 1 in Panel A of Table 2 reports the results from the baseline regression with number of publications as dependent variable. The positive and significant interaction of $post'02$ with $Treatment$ implies that the number of publications of the treated cohort increases by 18.3% after the announcement of the reform relative to the control cohort[40]. Because the $post'02$ variable is not a simple dummy variable that is 1 in the announcement year only, but remains 1 for the years after the announcement, this increase does not reflect a temporary jump in the number of publications in 2002 only, but a persistent increase relative to the control cohort. There is no additional increase in the number of publications from tenure onwards, when the treated academics enter into the performance pay scheme. To allow for a more easily interpretable result, I also estimate a linear fixed effects version of the baseline regression, the results of which are reported in Table 3. The results in column 1 suggest that academics in the treated cohort produce almost one extra publication every three years after the announcement of the reform compared to the control cohort[41]. Moving from a pure quantity measure of productivity to measures of quality-adjusted quantity and total impact, I find that the results are comparable (cf. columns 2 and 3 in Table 2). The number of publications of the treated cohort,

receive their PhD degree is 29.6, and the average age at which they complete their habilitation - and become eligible for a tenured position - is 37.2.

[39] This is the same model as used in, for instance, Azoulay et al. (2015). Even though the dependent variables here are not all integers, Silva and Tenreyro (2006) show, using a result from Gourieroux et al. (1984), that the estimator based on the Poisson likelihood function is consistent even for non-integer dependent variables, as long as the conditional mean is correctly specified.

[40] The exponentiated coefficients of the Poisson QML, minus one, can be interpreted as elasticities.

[41] The estimates of the linear FE model are biased because of the censored and highly skewed distribution of the dependent variables, so they should be interpreted with caution.

weighted by impact-factor rating, increases by 14.2% after the announcement of the reform, while the total number of citations increases by 13.8% relative to the control cohort, with no further increase upon entry into the performance pay scheme.

Moving from measures of total quantity and quality of research output to measures of average quality or impact, the results in columns 4 and 5 of panel A in Table 2 show that the average quality or impact of publications of the treated cohort decreases in response to higher-powered incentives. The average impact factor rating of publications produced by the treated cohort decreases by 9% after the announcement of the reform, and the average number of citations decreases by 10% after the announcement of the reform relative to the control cohort[42]. The coefficient estimates of the equivalent linear fixed effects estimation in columns 4 and 5 of panel A in Table 3 allow for slightly easier interpretation. The decrease in the average impact factor rating of publications of the treated academics after the announcement of the reform is equivalent to a change from the average publication appearing in the Journal of Political Economy, which had a 2-year impact factor rating of 5.247 in 2017, to the average publication appearing in the American Economic Journal: Applied Economics, which had an impact-factor rating of 5.028 in 2017 (Journal Citation Report 2017). The 10% decrease in the average number of citations in 2019 to publications by treated academics published after the reform announcement is equivalent to these publications having received on average 3.6 fewer citations.

The decrease in average quality or impact could be the result of substitution of efforts geared towards producing quality research to efforts geared towards producing more research. Although the observed differential increases in the quality-adjusted number of publications and the total number of citations rule out one-for-one substitution of efforts towards quality for efforts towards quantity, we may still worry that the reduction in the average quality and impact is the manifestation of a move away from time-consuming high-impact projects in a bid to produce more publications. To investigate whether this concern is warranted, I report the results of the baseline regression with the maximum and minimum number of citations to publications in columns 1 and 2 of Panel A.II in Table 2. Column 1 shows that there is no significant decrease in the maximum number of cites to the publications of treated academics published after the announcement of the reform or entry into performance pay, relative to citations to publications of the control cohort. Thus there is no evidence that the higher-powered incentives of the reform crowd out the production of high impact research. I also find no significant increase in the minimum number of citations to publications of the treated cohort in response to higher-powered incentives (cf. column 2, Panel A.II, Table 2). The coefficients in the regression with minimum number of citations as dependent variable are, however, rather imprecisely estimated. The coefficient of especially the first interaction term in this regression is quite large negative, so I cannot rule out that the minimum number of citations does not decrease in response to higher-powered incentives. Taken together, the results are consistent with the following illustrative example. Suppose an academic produces one high impact and one low impact paper every year. If this academic is treated, they still produce one high impact and one low impact paper every year (of the same impact levels as before the imposition of the high powered incentives), but the academic now also publishes a second low impact paper every year. The aforementioned results are not consistent with the academic producing one medium impact and two low impact papers after the reform. More generally, the results suggest that academics do not produce less high impact work in response to higher-powered incentives, but they do produce more lower-impact work.

Robustness

The changes in productivity metrics could be the result of strategic behaviors. Perhaps academics add their name to the papers of befriended academics, and return the favor for their friends, to boost publication records. The results in panel B in Table 2, for the baseline regression with dependent variables weighted by number of authors however do

[42] The latter estimate is significant at the 10% level only.

not show evidence of this behavior, as the results are very similar to the baseline results. The author-weighted number of publications of the treated cohort increases by 19.4% (column 1); the impact factor-rated number of publications, weighted by number of authors, increases by 11.6% (column 2); while the average impact factor rating, weighted by number of authors, decreases by 9.4% (column 4) relative to the control cohort after the announcement of the reform. The fact that accounting for the number of authors on a paper does not decrease the estimate of the increase in the (quality-adjusted) knowledge output or the decrease in its average quality or impact in response to higher-powered incentives, suggests that treated academics do not beef up their publication record by strategically adding their name to the papers of befriended colleagues.

The baseline specification includes 15 year-to-tenure dummies to flexibly control for changes in productivity in the run up to and after the start of the first tenured affiliation. Given that I have 20 years of data for these regressions, and considering the first tenure year occurs at different times for the treated and control cohorts, I am not controlling for a full set of year-to-tenure fixed effects for these cohorts. I would not be able to control for a full set of time-to-tenure fixed effects, while also controlling for year fixed effects and individual effects, as these effects are collinear (Cf. Macchiavello and Morjaria (2015)). However, to evaluate the robustness of the baseline specification, I estimate an alternative specification that includes the absolute value of the time until tenure, a non-linear function of years-to-tenure[43], instead of year-to-tenure fixed effects. Panel B of Table 3 shows that the main findings of the baseline estimation still go through in this specification.

Academics who are paid according to the age-related pay scheme can switch to the performance pay scheme after its implementation by changing affiliation or position, or by opting into the pay scheme while retaining the same position[44]. Academics in the control cohort may therefore end up being treated as well. Any effort response of the control cohort would lead me to *underestimate* the effort effect of the treated cohort, so, if anything, the baseline results provide a conservative estimate of the effort effect. To test whether this is the case indeed, I re-estimate the baseline specification with a control group without any switchers. where I label any academic who changes affiliation, position or contract[45] after implementation of the pay reform (as of 2005) as a switcher. As expected, Panel C in Table 9 shows that the estimates of the effort effect are, if anything, larger in this specification than in the baseline. Compared to control academics who remain in the age-related pay scheme, treated academics produce 23% more unweighted publications, and 19.5% more impact factor-rated publications, and their publications garner 17% more citations. This suggests that not accounting for contamination of the control group leads me to underestimate the effect of performance pay on (quality-adjusted) output and total impact by around 3 to 5 percentage points. The estimates of the effects on average quality or impact are, at a 9% and 10% decrease respectively, qualitatively the same.

The effort effect results are also robust to restricting attention to articles only(Table 8); widening or narrowing the treatment and control cohort windows (Panels A and B in Table 9); including more or fewer year-to-tenure dummies (Panels C and D in Table 9); including a $post'05 * Treatment_i$ interaction instead of the $tenure * Treatment_i$ interaction to control for implementation instead of entry into the performance pay scheme (Panel E in Table 9); or transforming the dependent variables using the inverse hyperbolic sine transform (Panel F in Table 9).

In short, I find evidence of a positive and significant average effect of performance incentives on the total raw and quality-adjusted quantity of knowledge output. The effect size ranges from 14 to 23% and arises as of the moment (stronger) implicit incentives take effect. This effect is of the same order of magnitude as previous estimates of the effort response to performance incentives, albeit mostly *explicit* performance incentives for routine tasks, in the literature

[43] This variable is equal to 1 in the year before tenure, 0 in the first tenure year, and 1 in the first year after the start of the first tenured affiliation, etc.

[44] Detmer and Preissler (2005) report that only a small number of professors chose to opt into the W-pay scheme in their current position.

[45] I assume that, whenever academics receive an outside offer, they either accept and change affiliation, or reject and renegotiate their current contract. If there are academics who do not at least renegotiate their contract when they receive an outside offer, this overestimates switching and leads to a conservative estimate here.

(e.g. Lazear (2000a), Shearer (2004)). I also find evidence of a significant decrease in the average quality or impact of knowledge output in response to higher-powered incentives of around 9 to 10%.

4.2.3 Validity of Identifying Assumption

The 14% to 23% increase in total output and impact, and the 9% to 10% decrease in average quality or impact can be interpreted as the *causal effect* of performance pay on knowledge creation if, absent the reform, the productivity of the treatment and control cohort would have evolved along parallel paths. There are a number of potential threats to the identification of this causal effort effect. It could be that other events that occurred around the same time are driving the result. Prime concerns are the effect of tenure itself, as well as other reforms occuring around the same time, especially because the introduction of performance pay was part of a comprehensive and continued effort to modernize German academia.

The control cohort start their first tenured affiliation shortly after the announcement of the reform. A drop in productivity of the control cohort post-tenure would therefore invalidate the parallel trends assumption and yield a spurious positive effort effect. In order to rule out that this is driving the results, I test for pre-existing trends, for the treatment and control cohort as well as a placebo treatment and control, below. Recall also that the time-to-tenure dummies in the baseline specification control for any common productivity changes in the run-up to and following the start of first-time tenured positions.

As for concurrent reforms, of particular concern for identification are the introduction of large grants for universities and research centres awarded through the "Excellence Initiative" as of late 2006/early 2007, the introduction of the German equivalent of assistant professorships ("Junior Professors") in 2002, and the abolition of the "professors' privilege" in 2002 (Enders 2001, Teichler and Bracht 2006, Von Proff et al. 2012, Fitzenberger and Schulze 2014, Lutter and Schröder 2016, Forschungsgemeinschaft 2016)[46]. Any events that occur around the time of the pay reform but which do not affect the pre- and post-reform cohorts differentially are, however, not a threat to identification. For this reason, the start of the "Excellenz Initiative" as of late 2006/early 2007 or the abolition of the professors' privilege in 2002 do not invalidate the identifying assumption. The introduction of the "Junior Professorship" in 2002 cannot be driving the earlier results either, because the first Junior Professors complete their Junior Professorship and become eligible for a tenured position by 2008/9 at the earliest[47] (Lutter and Schröder 2016). The three-year window of the treatment and control cohorts therefore does not include first-time tenured professors who completed a Junior Professorship. Nonetheless, I test for pre-existing trends below, to rule out the possibility that other events are driving the baseline results.

Finally, a third concern for the causal interpretation of the baseline results is the validity of the exogeneity assumption of the timing of tenure. In particular, academics could try to get a tenured affiliation sooner after they learn of the impending reform in order to avoid the performance pay system[48]. Tests for pre-existing trends for the treatment and control cohort as well as a placebo treatment and control rule out this alternative explanation. Furthermore, I show that the main results go through with synthetic treatment and control cohorts, defined by the average age at which academics start their first tenured affiliation rather than the actual timing of the first tenured affiliation.

Pre-Existing Trends and Effect Dynamics

I test for pre-existing trends by estimating the following model as a conditional QML Poisson fixed effects model:

[46] Under the professors' privilege regime, professors own the IPR of their inventions (Hvide and Jones 2016, Von Proff et al. 2012). The abolition of this privilege should reduce, rather than increase, incentives to produce (patentable) knowledge.

[47] Junior Professorships last six years, and are generally not tenure-track positions

[48] Note that attempts to delay the tenure decision would not be rational and therefore not much of a concern, since an academic would delay earning the higher pay associated with a tenured position, while they can always opt into the performance pay scheme after 2005 if so preferred.

$$E\left[Y_{i,f,t-x_f}|X_{i,f,t}\right] = exp[\alpha_i + \sum_{k=1}^{15}\beta_k \sum_{j=-7}^{7} tenure_{i,j} * Treatment_i + \sum_{j=-7}^{7} tenure_{i,j} + \gamma_t] \quad (23)$$

Here, $tenure_{i,j} * Treatment_i$ denote interactions of year-to-tenure dummies with the treatment variable. All other variables are as before. The coefficient estimates and 95% confidence intervals of the interaction terms are depicted in Figure 2a with number of publications as dependent variable, in Figure 2b with impact factor-rated number of publications as dependent variable, in Figure 2c with total number of citations as dependent variable, in Figure 2d with average impact factor-rating as dependent variable, and in Figure 2e with average number of citations as dependent variable. The green vertical dashed line at $t-5$ indicates where in the tenure trajectory of the youngest academics of the treated cohort, who start their first tenured affiliation in 2007, the announcement of the pay reform occurs. The orange vertical dashed line at $tenure_{i,j}$ marks the time at which treated academics enter into the performance pay scheme.

All five figures display a similar pattern: the interaction terms, which capture the difference in the respective output measures between the treated and control cohort, show that these cohorts start to diverge after the announcement of the reform, when the treated cohort faces higher-powered incentives. None of the figures show evidence of pre-existing trends. In the case of the first three figures, for dependent variables that measure total (quality-adjusted) output and impact, the differential becomes positive and significant; while for the last two figures, which display the results for average quality and impact metrics, the differential becomes larger negative. Figure 2a shows that the number of publications starts to differ significantly (at the 1% level) from 4 years before the start of the first tenured affiliation, which is the year after the announcement of the pay reform for the youngest treatment cohort. The coefficient estimate implies that treated academics produce about 19.5% more publications than academics in the control cohort at that point. This difference goes up to 27.5% the following year, and stays between these difference levels thereafter, through entry into the performance pay system and beyond. Recall that the publications have been backdated by the average publication lags in the respective academic fields (rounded up to the nearest year), so the differential increase in number of publications directly after the announcement of the reform is consistent with an immediate effort response to the higher-powered incentives it heralds. The pattern is very similar for the impact factor-rated number of publications (Figure 2b), though the differential becomes significant (at the 5% level) one year later, at $tenure_{i,-3} * Treatment$. This is to be expected if producing high quality research takes more time. Figure 2c shows that the differentials for the total number of citations echo the pattern for the total impact factor-rated number of publications, though they are less precisely estimated and become significant only after tenure, when explicit performance incentives take effects for treated academics. The differential between the average impact factor rating of the treated cohort and the control cohort becomes negative and significant at the 5% level at $tenure_{i,-4} * Treatment$, after the announcement of the reform, and remains this way for much of the post-announcement period (Cf. Figure 2d). The differentials for the total number of citations in Figure 2e are imprecisely estimated, though the coefficient estimates echo the pattern for average impact factor rating differentials with a lag of about two years.

The absence of pre-existing trends and clear alignment of the productivity response with the timing of the announcement of the pay reform, and hence the moment when the treated cohort starts to face higher-powered incentives, lends support to the interpretation of the differential productivity reponse as the causal effort effect of the performance pay reform and not the effect of another event. The analysis also underlines that the effort response to the higher-powered incentives is not just a temporary response, but one that persists.

Placebo Experiment

As discussed above, another threat to identification would be the endogeneity of the timing of the start of the first tenured affiliation. The first piece of evidence that suggests the identifying assumption is met, is coming from the baseline results in Table 2. Academics in both the treatment and control cohort start their first tenured affiliation after the announcement of the reform, so academics in both cohorts may respond to the announcement by increasing their research output in an attempt to get a tenured position before the reform is implemented. If it were the case that the control cohort comprises academics who managed to get a tenured position before the reform takes effect by increasing their research output relatively more than the treated cohort, the $post'02 * Treatment_i$ interaction would be negative instead of positive If, on the other hand, academics in the treated cohort increased their research output relatively more than the control cohort after the announcement of the reform and before its implementation in an unsuccessful attempt to get a tenured position before the reform, the $tenured * Treatment_i$ interaction should be negative. Instead, the baseline results show that the research output of the treated cohort increased significantly more than that of the control cohort after the announcement of the reform, and did not drop back down after the implementation of the reform. The baseline results therefore show no evidence that the timing of tenure is not endogenous.

To further test the validity of the identifying assumption, I estimate a placebo pre-trends regression as in equation 23. I restrict the sample to academics who start their first tenured position in 2001 to 2004 and use the cohort that starts their first tenured position at a public university in 2001 or 2002 as placebo control group and those who start their first tenured position in 2003 or 2004 as placebo treatment group. All other specifications are as before. The coefficient estimates and 95% confidence intervals of the interactions of time-to-tenure dummy variables with a placebo treatment dummy are presented in Figures 3a through 3e for the same dependent variables as in Figures 2a through 2e. Similar to Figure 2, the green vertical line in Figure 3 indicates when the reform is announced for the youngest placebo-treatment cohort - those that start their first tenured position in 2004, while the orange vertical line indicates the earliest point in time when the placebo-treatment cohort would have entered into the performance pay scheme, had they been treated. In contrast to the corresponding graphs for the experiment of interest, the placebo treatment group does not experience a differential increase in research output or decrease in average quality or impact relative to the placebo control cohort after the announcement of the reform or entry into the performance pay scheme. The interaction terms are generally close to 0 and not significant and they do not show evidence of any consistents differential trends. If academics in the placebo treatment group had been able to avoid entry into the performance pay scheme by temporarily stepping up their efforts to obtain a tenured affiliation sooner, the productivity differential between the placebo treatment and control would have been positive between the announcement of the reform and the start of the first tenured affiliation, and possibly negative thereafter. The absence of such a pattern thus lends support to the identifying assumption of the exogeneity of the timing of the start of the first tenured affiliation.

Synthetic Cohorts

As a final test of the validity of the identifying assumption, and the causal interpretation of the productivity differentials between the treated and control cohort as the effort effect of the higher-powered incentives of the performance pay reform, I estimate the baseline regression with synthetic cohorts. The synthetic treatment and control cohorts are defined using the average age at which professors at German public universities start their first tenured affiliation. This average tenure age is 44. The synthetic treatment cohort then comprises academics whose synthetic first tenure year falls between 2005 and 2007, while the synthetic control cohort is made up of academics whose synthetic first tenure year falls between 2002 and 2004. I calculate the synthetic first tenure year by adding 44 to an academic's synthetic age. An academic's synthetic age is derived from an academic's birth year, if this is known, and from a synthetic birth year otherwise. I derive the synthetic birth year from the year in which academics pass their habilitation or, if I don't have this information, the year in which academics receive their PhD, by adding the average age at which academics who become tenured professors at public universities pass their habilitation or receive their PhD, respectively[49]. Panel

[49] These average ages are 37 and 30, respectively

D in Table 3 shows that the results are qualitatively similar with synthetic cohorts; with increases in raw and impact-factor weighted number of publications of 10.4 to 15.7% and a decrease in average impact of 12.4% as of the moment the higher-powered implicit incentives of the performance pay reform take effect.

4.2.4 Heterogeneous Treatment Effects

The baseline results show how performance pay affects research productivity on average. This section delves into the anatomy of the effort response, and tests if and how the effort response differs by productivity quantile and academic field.

Heterogeneous responses by productivity quantile

I first estimate the baseline specification separately for the top quartile and the bottom three quartiles. I determine productivity quantiles on the basis of the averages of the various dependent variables over the three pre-announcement years 1999, 2000 and 2001[50], separately by academic field and tenure cohort. Panel A of Table 4 shows that the raw research output of top quartile academics in the treated cohort increases by 15.5% relative to the raw research output of top quartile academics in the control cohort in response to the implicit incentives that take effect as of the announcement of the performance pay reform. There is no significant effort response in terms of quality-adjusted number of publications or total impact, but the average impact-factor rating and average number of citations to the publications of treated top quartile academics decreases by 16.9% to 17.4% relative to similarly productive academics in the control group. In contrast, the average impact of the work of treated academics in the bottom three quartiles does not decline significantly relative to the same quartiles of the control cohorts in response to performance pay incentives, while total research output, quality-adjusted research output and total impact all increase, by 17.9 to 27.4% (Cf. Panel B of Table 4).

To test whether the effort responses of the different quartiles are significantly different from each other, I estimate the baseline regression model augmented with interaction terms with an indicator for the top quartile. Panel C in Table 4) shows that the differences in effort responses to performance pay are, in fact, significant. Here the bottom three quartiles are the reference group, so the coefficient of the $post'02$ interaction with the $Treatment$ variable in column 1 implies that academics in the bottom three quartiles produce 18.2% more publications after the announcement of the reform, relative to the same quartile in the control group. The triple interaction of the $post'02$, $Treatment$ and Q4 indicator in the row below is close to zero and not significant, meaning that the effort response of the top quartile of the treatment group to the implicit incentives of performance pay is not significantly different from the effort response of the bottom three quartiles of the treatment group as far as number of publications is concerned. The same pattern emerges for total number of citations, average impact-factor rating and average number of citations. The implicit performance pay incentives introduced by the announcement of the reform increases the total number of citations of the bottom three quartiles by 15.5% and decreases the average impact factor rating and citations by 9.2% and 11% relative to the same quartiles in the control group, while the effort response of the top quartile does not significantly differ from this. The only productivity measure for which the effort response of the top quartile to the implicit incentives differs from that of the bottom three quartiles, though only marginally, is the total impact factor-rated number of publications. While the bottom three quartiles in the treatment group show an effort response of 23.9% relative to the bottom three quartiles of the control group, the top quartile's effort response is 11.8% less than this.

The effort responses of the quartiles start to differ significantly across all measures of output when the explicit performance pay incentives take effect upon entry into the performance pay scheme. The effort response of the bottom

[50] I base the quantiles on pre-announcement averages so as to avoid sumultaneity bias. Results are however robust to using contemporaneous productvity measures.

three quartiles of the treated cohort, relative to the same quartiles in the control group, ranges from a - marginal - increase in the average impact factor rating of 6.7%, to a very significant increase in the total impact factor-rated number of publications of 19.8%. In contrast, the effort response of the top quartile of the treated cohort is significantly smaller than that of the bottom three quartiles, with the difference ranging from 9.6% for the average impact factor-rating to 20.3% for the total impact factor-rated number of publications. The average number of citations of top quartile academics finally decreases by 19.5% in response to the explicit performance incentives, while the average number of citations of the bottom three quartile does not change.

A potential concern for the estimation of heterogeneous treatment effects is differential selection into performance pay. If more productive academics are more likely to select into performance pay (which the section on selection below shows to be the case) the top quartile of the control group contains more academics who end up being treated than the bottom quartiles of the control group. If performance pay incentives have a positive effort effect for highly productive academics as well, this would create a larger downward bias of the effort effect estimate for highly productive academics compared to less productive academics. I therefore re-estimate the heterogeneous response regression excluding switchers[51] from the control group (Cf. Panel A in Table 10). Reassuringly, the estimates of the effort responses and differences in effort responses across quartiles are robust to excluding switchcers and, if anything, the results become stronger.

Taken together, the analysis of heterogeneous effort effects by productivity quantiles shows that less productive academics respond strongest to performance pay. They increase total research output as well as the total impact of that output in response to both implicit and explicit performance pay incentives. The effort response of the most productive academics on the other hand is less positive. Their effort response in terms of total research output as well as total impact is less positive and the average impact of their work decreases, especially once explicit performance incentives take effect.

Heterogeneous responses by academic field

Effort responses may also differ by field. Some fields, such as physics, medicine and chemistry rely heavily on teamwork, while others, such as theology and history, feature mostly solo research[52]. The response to higher-powered incentives may be larger in smaller teams, if the likelihood that all or most team members are highly incentivized is larger in smaller teams. Fields also differ in the types of publications that are most commonly used to disseminate research, with books the most common outlet for philosophers, articles in scientific journals the publication type of choice for economists, and conference proceedings the most common publication type for engineers. The specific publication type may affect the speed with which academics can respond to changes in incentives over and above any variation in publication lags (which I account for in the construction of the dependent variables). Books take longer to write than conference proceedings, in general, so fields that disseminate research findings predominantly through more agile channels, such as conference proceedings, may show a faster response to changes in incentives than fields in which scholars mostly write books (e.g. humanities).

To study heterogeneous effort responses by academic field, I estimate the baseline regression separately by broad academic field: science, applied science, social science and humanities. I classify mathematics, physics and informatics, biology, chemistry, earth sciences and pharmacology as *science*; engineering, medicine, dentistry, veterinary, agricultural science and nutrition science as *applied science*; law, economics and other social sciences as *social science*; and theology, philosphy and history, philology and anthropology as *humanities*. Panels A through D of Table 11 show the estimation results of the baseline regression for each of these broad academic disciplines. These results echo

[51] Academics who switch to the performance pay scheme by changing position or affiliation or renegotiating their contract, as defined above in the Robustness section.

[52] As gauged by the average number of authors on a paper in a field (calculated over the pre-reform years 1996-2000)

the baseline estimation of the average effort response, with total research output and total impact increasing in response to implicit performance incentives, while the average impact of the knowledge created decreases. I do however run into power problems by dividing the sample in this way, which leads to imprecise estimates for the smaller disciplines (humanities, social science) especially.

In an attempt to improve power, I next estimate the heterogeneous responses by productivity quantile for each broad discipline separately, using the specification from the previous section. The results are shown in panels A through D of Table 5. These results show that the heterogeneous responses by productivity quantile documented above persist across academic disciplines. In general, the total research output and total impact of less productive academics increases in response to both implicit and explicit incentives, while this response is smaller for the most productive academics, especially in response to the explicit performance incentives. There is also some suggestive evidence that this response materializes sooner, or - which is equivalent in this setting - predominantly in response to implicit performance incentives, in fields with more agile dissemination routes (applied science and science). Furthermore, the size of the response, and in particular the percentage change in total impact, as well as the difference in the response size between different productivity quantiles appears to be larger in academic fields that rely less on teamwork (compare e.g. social science to science).

4.2.5 Novelty

Ultimately, I am interested in how novel the knowledge is that is produced, and if and how the novelty changes in response to performance pay incentives. To this end, I calculate cosine similarity measures of the Term Frequency Inverse Document Frequency vectors of focal papers and comparison papers, as in Kelly et al. (2018). The Term Frequency Inverse Document Frequency (TFIDF) index is defined in the following way:

$$TFIDF_{w,d,t} = \left(\frac{c_{w,d}}{\sum_l c_{l,d}}\right)\left(log\left(\frac{c_{d,s<t}}{1+c_{d,s<t}\, with\, c_{w,d}>0}\right)\right)$$

where $c_{w,d}$ denotes the count of a word w in document d and $c_{d,\tau}$ denotes the count of documents that contain the word w and that were published in period t. The first term is then the frequency of word w in document d (the “Term Frequency” part), while the second term is the log of the inverse of the frequency of documents in which word w appears (the “Inverse Document Frequency” part). A word that occurs more frequently in a document has a higher TFIDF, while more commonly used words have a smaller TFIDF, because the second term is smaller. As in Kelly et al. (2018), I use the “backward” TFIDF; that is, I use only publications published in the years before publication of the focal document to calculate the inverse document frequency. I calculate TFIDFs for the abstracts of all publications in the data set at the bigram (word pair) basis, rather than word by word.[53]

In order to the measure the similarity of the bi-grams used in the focal publication, relative to a set of comparison publications, I calculate the cosine similarity of the normalized vector of TFIDFs of the focal document and a comparison publication. Formally, for focal document d published in t, the cosine similarity with comparison publication $\tilde{d}$ published in $\tilde{t}$ is:

$$\rho_{d,t;\tilde{d},\tilde{t}} = \left(\frac{TFIDF_{d,t}}{|TFIDF_{d,t}|}\right)\cdot\left(\frac{TFIDF_{\tilde{d},\tilde{t}}}{|TFIDF_{\tilde{d},\tilde{t}}|}\right)$$

If the focal publication and comparison publication have abstracts that have no bigrams in common, the cosine similarity is 0. The more common bigrams in the focal publication and comparison publication, the closer the cosine

[53] As an illustration, the following sentence, ’Paul walks home’, has two bigrams: ’Paul walks’ and ’walks home’.

similarity is to 1.

As in Kelly et al. (2018), I calculate two different metrics derived from these cosine similarities: backward similarity and forward similarity. The backward similarity of focal publication d published in t is calculated as the sum of the cosine similarities between the focal document and comparison publications published in a three year window *before* the focal document was published, while the forward similarity is calculated as the sum of the cosine similarities between the focal publication and comparison publications published in a three year window *after* the focal document was published. Formally:

- Backward Similarity: $BS^{\tau}_{d,t} = \sum_{S_{t,\tau}} \rho_{d,t;\tilde{d},\tilde{t}}$ where $S_{t,\tau}$ is the set of papers published in $\tilde{t} : t > \tilde{t} \geq t - \tau, \tau = 3$
- Forward Similarity: $FS^{\tau}_{d,t} = \sum_{S_{t,\tau}} \rho_{d,t;\tilde{d},\tilde{t}}$ where $S_{t,\tau}$ is the set of papers published in $\tilde{t} : t < \tilde{t} \leq t + \tau, \tau = 3$

The backward similarity metric quantifies the novelty of the focal publication. The more bigrams in the abstract of the focal publication that have not been used in abstracts of publications published earlier in the same field, the smaller the backward similarity metric. I consider such publications to be more novel. The foward similarity metric on the other hand captures follow-on research, or impact. The higher this measure, the more publications published after the focal publication use the same bigrams in their abstract.

I calculate two sets of backward and forward similarity measures for each focal publication; a "within-field" set of similarity measures, for which I restrict the set of comparison publications to the same field as the focal publication, and a "within-author" set, for which I restrict the set of comparison publications to publications authored by the same scholar as the focal publication. For each similarity measure, I calculate the minimum, maximum and average similarity (all calculated over the number of publications of author i in year t). Panels A and B in Table 12 show the results of the estimation of the baseline specification in equation 22 with backward and forward similarity measures defined based on within field and within author comparisons, respectively. There is some evidence that research produced by treated academics after the implicit incentives of the reform take effect finds more follow-on research, as the maximum within-field forward similarity of the post-announcement work of treated academics increases by a marginally significant 9.1% though the minimum forward similarity decreases 18.7% after the explicit performance incentives take effect.

Estimating the same regressions for the top quartile and bottom three quartiles separately shows substantial heterogeneity in the changes in research novelty and impact in response to performance pay[54]. The most impactful work of the top quartile of treated academics finds even more follow-on research, 18.5% to be exact, in response to the implicit performance incentives (Cf. column 5 in Panel A.I of Table 6). This is a persistent increase in impact, as there is no subsequent differential change in maximum forward similarity upon entry into the performance pay scheme. The increase in follow-on research to the work of the most productive academics does not appear to be driven by the academics themselves revisiting the same topics more often. While there is some marginally significant evidence that the treated top quartile academics are more likely to revisit one-off themes initially after the announcement of the reform, relative to top quartile academics in the control group, this result reverses upon entry into the performance pay scheme (Cf. column 4 in Panel B.I of Table 6). At the same time, there is no evidence that the most productive academics take regurgitation to a new level or that they revisit the same themes more on average (Cf. columns 5 and 6 in Panel B.I of Table 6). Yet the most novel work of the most productive academics becomes less novel upon entry into the performance pay scheme (Cf. column 1 in Panel A of Table 6). This decrease in novelty is both statistically and, at 40.5%, economically significant. For the bottom three quartiles not much changes in terms of novelty or impact of the work of treated academics, other than that their least impactful work finds even less follow-on research, relative to the bottom three quartiles of the control group, once explicit performance incentives take effect (Cf. column 4 in Panel B.II of Table 6).

[54] I determine productivity quantiles here on the basis of the averages of impact factor-rated number of publications over the three pre-announcement years 1999, 2000 and 2001, separately by academic field and tenure cohort.

Panel B in Table 12 shows that the differences in the novelty and impact response between productivity quantiles are, in fact, statistically significant. Columns 1 and 2 of panel B.1 show that, while the treated bottom three quartiles do not significantly change the novelty of their least novel or most novel work in response to the implicit or explicit incentives of the performance pay reform, the treated top quartile decreases the novelty of their most novel work by 17.2% but increases the novelty of their least novel work by 15.6% upon entry into the performance pay scheme. The maximum and average within-author backward similarity for the treated top quartile also decreases by around 19%, so their research agendas appear to become more divergent. Finally, while the least impactful work of the treated bottom three quartiles becomes 21.5% less impactful upon entry into the peformance pay scheme, relative to the same quartiles in the control cohort, the least impactful work of the treated top quartile becomes 18.5% more impactful relative to the treated bottom three quartiles.

4.3 Selection Effect

The empirical analysis has so far focused on the effort effect of performance pay. This is only one channel through which performance pay can increase output and I now turn attention to another important channel: selection. To this end, I study which academics sort into performance pay by switching from the old age-related pay scheme to the new performance pay scheme.

In the German system, academics who already held a tenured affiliation before 2005 can select into the performance pay scheme by changing affiliation or position, or by opting into the pay scheme while retaining the same position. I do not have information on the latter, though Detmer and Preissler (2005) report that only a small number of professors chose to opt into the W-pay scheme in their current position. I do however observe professors changing affiliation or position or otherwise renegotiating their existing contract and, consequently, changing into the performance pay scheme. I exploit this information to analyze the selection effect of the reform. As derived in the theory section, I expect more productive academics to be more likely to select into the performance pay scheme[55]. I test this hypothesis through hazard rate and survival function analysis in here.

Non-Parametric Analysis

As a first test, I analyze the hazard and survival rates of switches to performance pay of academics with a tenured affiliation at a public university before the reform. These academics have the choice (i.e. are “at risk”) of switching pay scheme because they are intially paid according to the age-related pay scheme. I label any first affiliation, position or contract change[56] after implementation of the pay reform (as of 2005) as a switch to performance pay (the “failure” event) and use the time until such a switch as duration variable. Academics are “at risk” of switching after their most recent affiliation, position or contract change if they are tenured, at a public university and not retired. There are 11237 academics who hold a tenured affiliation at a public university before 2005 and I observe 1231 switches in a total of 85716 periods (years) during which these academics can switch from the age-related to the performance pay system. The average incidence rate of switches is 0.014.

Figure 4 shows the Epanechnikov kernel density estimates of the hazard function for switches from age-related to performance pay for academics whose average productivity falls in the top quartile or bottom three quartiles of the average productivity distribution[57]. I base productivity quantiles on the averages over the three pre-implementation

[55] Cf. proposition 2 in the theoretical model.

[56] I assume that, whenever an academic receives an offer, he either accepts and changes position, or rejects and renegotiates his current contract. In either case, the academic switches to the new performance pay scheme if the change or renegotiation happens after the reform. If there are academics who do not at least renegotiate their contract when they receive an offer, these academics are more likely to be of a lower productivity type, and including them in the pool of switchers would reduce the estimate of the selection effect I find.

[57] Because only academics who already held a tenured affiliation before the reform can switch from age-related to performance pay by changing affiliation or position, only these changes are considered here.

years (2002, 2003 and 2004) of the impact factor-rated number of publications in panel a and the impact factor-rated number of publications weighted by number of authors in panel b of Figure 4. As before, I calculate quantiles separately by academic field and tenure year. In both figures, the hazard rate for switching to the performance pay scheme is clearly greater for top quartile academics throughout, so higher productivity academics are more likely to sort into the performance pay scheme. A log-rank test of the equality of the survival functions of top quartile academics and bottom three quartile academics rejects the equality of the survival functions at the 1% significance level[58].

Difference-in-Differences Estimation

As a next step, I estimate the selection effect of the introduction of performance pay parametrically by estimating the following Weibull proportional hazard model:

$$\lambda_{i,t} = \rho * exp\left[\beta_0 + \beta_1 Treated_i + \beta_1 AvgProd_i + \beta_2 age_{i,t} + X_i + u_{i,t}\right] * t^{\rho-1} \quad (24)$$

I use this model to estimate the hazard ratio of changing affiliation, position or contract after 2004, that is, after implementation of the reform. Instead of restricting attention to academics who start their first tenured affiliation before the implementation of the reform only, I now also include academics who start their first tenured position after the reform. The latter automatically fall under the performance pay scheme from the moment they make tenure. Thus they cannot decide to select into the performance pay scheme and they do not switch performance scheme when they change affiliation, position or contract. In contrast, academics who start their first tenured position before 2005 fall under the age-related pay scheme when the reform is implemented. They switch to performance pay when they change their affiliation, position or contract after the reform takes effect in 2005. Hence the cohort of academics who start their first tenured position before 2005 is the treated cohort here, while those that tenure after 2004 make up the control cohort. Accordingly, *Treated* is a dummy variable that is 1 for academics who started a first tenured affiliation before 2005, and 0 for those whose first tenured affiliation started as of 2005.

AvgProd is an academic's average productivity calculated as three-year pre-implementation averages (2002-2004) of the unweighted impact factor-rated number of publications. The age variable is equal to an author's self-reported age if known, and equal to a synthetic age otherwise. I calculate synthetic ages using the average age at habilitation, promotion or tenure. All models control for academic field fixed effects and are estimated for years $t > 2004$ and for academics who start their first tenured position as of 1999. Standard errors are robust and clustered by individual academic.

The coefficient estimate of the *AvgProd* variable in column 1 in Panel A of Table 7 is positive and significant, suggesting that more productive academics are more likely to change their position or renegotiate their contract in general. The coefficient estimate in column 1a implies that a one standard deviation increase in average productivity increases the likelihood that an academic changes position or contract by 5.2%[59]. Older academics are also less likely to change position or contract. The coefficient estimate of the age variable in column 1a for instance implies that being one year older decreases the probability that an academic will change position or contract by 13.2%. In short, more productive academics are more likely to switch to performance pay, while older academics are less likely to switch, and this is true both for academics who start their first tenured position before 2005 (the treated cohort) and after 2005 (the control cohort).

It is likely that younger and more productive academics change affiliation or renegotiate their contract with a higher probability in general, for instance because they face lower relocation costs and are more likely to receive outside

[58] When using the quantile based on impact factor-rated number of publications, the log-rank test returns a Chi-squared statistic of 8.14 (p-value 0.0043). The log-rank test returns a Chi-squared statistic of 13.19 (p-value 0.0003) when using the quantiles based on impact factor-rated number of publications weighted by number of authors.

[59] Calculated as EXP(0.002*25.53797)-1, where 25.53797 is the standard deviation of the average productivity variable here.

offers, respectively. In order to distinguish between these general selection patterns and specific selection effects of performance pay, I exploit the variation in selection incentives by age as well as productivity in a difference-in-differences framework. Older academics in the treated cohort should have weaker incentives to switch than academics of the same age in the control cohort due to the single-crossing property of the basic wage schedules of the age-related and performance pay schemes (cf. Figure 1). At the same time, more productive academics in the treated cohort have stronger incentives to change their affiliation or contract than academics of similar productivity in the control cohort, because treated academics stand to gain potentially sizeable bonuses by switching to the performance pay scheme.

I test for the selection effect of performance pay in a difference-in-differences framework by estimating the model specified in equation 24, augmented with $age * Treated$ and $AvgProd * Treated$ intereactions. I include $age * Treated$ to prevent a spurious correlation between productivity and switching rates from biasing the estimates of the prime interaction of interest; $age * Treated$. Such a spurious correlation would arise if older academics are less productive and less likely to switch (for reasons other than performance pay per se, such as the single-crossing property). The results of this estimation are reported in column 1b of Panel A in Table 7. The positive and significant coefficient estimate of the interaction $AvgProd * Treated$ implies that a one standard deviation increase in the average productivity of treated academics increases the likelihood that they change affiliation or contract, and hence switch to performance pay, by 10.7% more than academics in the control group. This holds while controlling for the interaction term $age * Treated$, which, as expected, is negative and significant and implies that the likelihood that a treated academic switches, decreases by an extra 2.8% per year of age relative to academics in the control group. These results are robust to estimating the model as a Cox proportional hazard model (Columns 1a and 1b in Panel A of Table 13), basing the *AvgProd* variable on the three year pre-reform average of the impact factor-rated number of publications weighted by number of authors (Columns 1a and 1b in Panel B of Table 13)[60] and estimating the Weibull model with academic field strata (Columns 1a and 1b in Panel C of Table 13). Replacing the *AvgProd* variable by a dummy indicating whether an academic's pre-reform average productivity is above median shows that treated academics of above median productivity are 20.7% more likely to switch than similarly productive academics in the control group (Cf. Columns 1a and 1b in Panel B of Table 7). Performance pay therefore has a positive selection effect, as it is significantly more likely to attract more productive academics.

Validity Checks

To assess the validity of the selection effect estimation, I run placebo estimations of the aforementioned models. Academics who start their first tenured position before 2002 are defined as placebo treatment group here, while academics who start their first tenured affiliation between 2002 and 2005 act as control group. Both groups switch into the performance pay scheme when they change affiliation, position or contract after 2004, so they face the same incentives to switch into performance pay. The estimation results of these placebo estimations are reported in columns 2a and 2b of the various panels in Tables 7 and 13. The $AvgProd * Treated$ interaction is never significant, so there is no evidence that productive academics in the placebo treatment group are more likely to select into performance pay than academics in the placebo control group. Reassuringly, the $Age * Treated$ interaction is never significant either, as it should be, because the placebo treatment and control groups face the same difference in basic wage schemes between the age-related and performance pay scheme.

As a final check, I also estimate any changes in switching rates from before to after the implementation of the reform for academics in the treated cohort (those whose first tenured affiliation started before 2005). Table 14 shows, that, while a one standard deviation increase in average productivity increases the likelihood of a switch by around 5.5% on average, this increase in the likelihood of a switch grows to 8 or 9% after the implementation of the reform.

[60] Results are robust to basing the average productivity variable on other productivity measures as well. Results are available from the author on request.

This consolidates the finding of a positive selection effect of performance pay.

5 Conclusion

This paper studies the effort and selection effect of performance pay in academia and provides empirical evidence that academics significantly increase effort in response to performance pay and that higher productivity academics are more likely to select into performance pay. In order to do so, I use the introduction of performance pay in German academia in 2002 as a natural experiment and employ a newly constructed data set encompassing information regarding research productivity and affiliations of the universe of German academics. Before the reform, academics were all paid according to an age-related pay scheme, in which the effectively flat wage increases with age. In contrast, in the performance pay scheme implemented after the reform, academics earn a basic wage that does not increase with age and is lower than the basic wage in the age-related pay scheme for most ages. On top of this basic wage however, academics can now earn bonuses that are distributed through explicit on-the-job performance tournaments and also take the form of implicit incentives.

In order to estimate the effort effect, the paper exploits the fact that academics who make tenure just before the reform fall under the old, age-related pay scheme, while academics who make tenure directly after the reform are paid according to the performance pay scheme. If the timing of the tenure decision is exogenous, the difference in the change in productivity from before to after the reform between the cohort making tenure just before the reform and the cohort making tenure directly after, can be interpreted as the causal effect of performance pay on academic effort. I estimate this differential change in a difference-in-differences framework and find an effort effect that amounts to a conservatively estimated 12 to 16% increase in productivity on average. Because the pay reform was required to be budget-neutral, this increase in research productivity represents a net gain, as it comes at no extra cost. The average effort effect is driven by the implicit incentives of the performance pay scheme and only robustly manifests itself as an increase in the quantity of research output. A placebo difference-in-differences estimation shows that the identifying assumption of an exogenous tenure decision is plausible, as there is no evidence of academics speeding up the timing of their first tenured position. Furthermore, I find no evidence of pre-existing trends, which lends support to the identifying parallel-trends assumption of the difference-in-differences framework.

The average effort effect hides considerable heterogeneity in the response to performance pay. Relatively less productive academics display the strongest and most robust effort response to performance pay. In terms of individual-level measures of research output, academics in the bottom three quartiles of the productivity distribution increase productivity by 18 to 27% in response to implicit incentives. This increase holds for quality-adjusted measures of research output, as well as more purely quantity based measures. In contrast, top quartile academics increase only research quantity, and only by about 15% and the average quality and impact of their work decreases.

I then estimate the selection effect though hazard rate and survival function analyses, where I use the fact that any tenured professor who changes affiliation or position after the reform automatically switches to the performance pay scheme. Because pay in the performance pay scheme only increases with performance, and no longer with age, I expect to find that more productive academics are more likely to switch to performance pay. This selection effect is borne out by the analysis indeed and, moreover, I find that this effect is stronger for younger academics. The latter finding aligns with the fact that the gap in basic wage between the age-related pay and the performance pay scheme is larger for older academics. To distinguish the selection effect from other factors that might cause younger and more productive academics to be more likely to change affilitiation or position, estimate the differential effect of age on switching rates for academics who make tenure before the reform and those who make tenure after. Only the former cohort of academics actually switch into the performance pay scheme when changing affiliation or position. It is this cohort that experiences a differential decline in switching rates with age, but only for relatively less productive academics. That is, relatively more productive academics are more likely to select into performance pay.

By studying the effort and selection effect of performance pay in academia, the paper aims to contribute to and form a bridge between the literature on university governance and incentives in organizations. Given the economic importance of the academic sector, both in terms of direct economic value as well as for innovation and growth, it is crucial to understand the factors that determine the performance of universities. The literature on incentives in organisations provides ample evidence that incentives can significantly increase performance, and this paper shows that incentives can improve academic performance too. Performance pay increases average effort and attracts relatively more productive academics. Yet the average effort effect only robustly pertains to increases in research quantity and the greatest effort response - both in terms of raw quantity and quality-adjusted quantity - comes from relatively less productive academics. Performance pay does however not attract the relatively less productive academics, so much of the potential effort response might be left untapped because of the selection effect.

There are several steps than can be taken next. The current paper focuses on research productivity of academics, but it would be very interesting and equally relevant to study the effect of performance pay on the educational performance of academics. Furthermore, it would be interesting to see if and when extrinsic motivation starts to crowd out academics' intrinsic motivation. Another important question is how performance pay affects the selection of candidate-academics. Other possible impacts of performance pay, such as on collaboration and network formation would make for exciting research avenues too.

References

Aghion, P., Dewatripont, M., Hoxby, C., Mas-Colell, A. and Sapir, A.: 2010, The governance and performance of universities: evidence from europe and the us, *Economic Policy* **25**(61), 7–59.

Angrist, J., Lang, D. and Oreopoulos, P.: 2009, Incentives and services for college achievement: Evidence from a randomized trial, *American Economic Journal: Applied Economics* pp. 136–163.

Angrist, J. and Lavy, V.: 2009, The effects of high stakes high school achievement awards: Evidence from a randomized trial, *The American Economic Review* pp. 1384–1414.

Ariely, D., Bracha, A. and Meier, S.: 2009, Doing good or doing well? image motivation and monetary incentives in behaving prosocially, *The American Economic Review* pp. 544–555.

Audretsch, D. B. and Feldman, M. P.: 1996, R&d spillovers and the geography of innovation and production, *The American economic review* pp. 630–640.

Audretsch, D. B. and Feldman, M. P.: 2004, Knowledge spillovers and the geography of innovation, *Handbook of regional and urban economics* **4**, 2713–2739.

Azoulay, P., Ding, W. and Stuart, T.: 2009, The impact of academic patenting on the rate, quality and direction of (public) research output, *The Journal of Industrial Economics* **57**(4), 637–676.

Azoulay, P., Furman, J. L., Krieger, J. L. and Murray, F.: 2015, Retractions, *Review of Economics and Statistics* **97**(5), 1118–1136.

Azoulay, P., Graff Zivin, J. S. and Manso, G.: 2011, Incentives and creativity: evidence from the academic life sciences, *The RAND Journal of Economics* **42**(3), 527–554.

Backes-Gellner, U. and Schlinghoff, A.: 2004, Careers, incentives, and publication patterns of us and german (business) economists, **Working paper**.

Bandiera, O., Barankay, I. and Rasul, I.: 2005, Social preferences and the response to incentives: Evidence from personnel data, *The Quarterly Journal of Economics* pp. 917–962.

Belenzon, S. and Schankerman, M.: 2009, University knowledge transfer: private ownership, incentives, and local development objectives, *Journal of Law and Economics* **52**(1), 111–144.

Bellemare, C., Lepage, P. and Shearer, B.: 2010, Peer pressure, incentives, and gender: An experimental analysis of motivation in the workplace, *Labour Economics* **17**(1), 276–283.

Besley, T. and Ghatak, M.: 2005, Competition and incentives with motivated agents, *American economic review* **95**(3), 616–636.

Bettinger, E. P.: 2012, Paying to learn: The effect of financial incentives on elementary school test scores, *Review of Economics and Statistics* **94**(3), 686–698.

Björk, B.-C. and Solomon, D.: 2013, The publishing delay in scholarly peer-reviewed journals, *Journal of Informetrics* **7**(4), 914–923.

Bloom, N. and Reenen, J. V.: 2011, Human resource management and productivity, *Handbook of labor economics* **4**, 1697–1767.

BMBF: 2002, Gesetz zur reform der professorenbesoldung. ProfBesReformG.
URL: *http://www.bmbf.de/pubRD/profbesreformg.pdf*

BMI: 2007, Bericht zum besoldungsrechtlichen vergaberahmen bei der professorenbesoldung nach § 35 abs. 5 bundesbesoldungsgesetz, *Technical report*, Bundesministerium des Innern. Stand 29 February 2008.
URL: *http://www.parlament-berlin.de/ados/16/WissForsch/vorgang/Bericht%20Professorenbesoldung.pdf*

Boly, A.: 2011, On the incentive effects of monitoring: evidence from the lab and the field, *Experimental Economics* **14**(2), 241–253.

Borjas, G. J. and Doran, K. B.: 2015, Prizes and productivity how winning the fields medal affects scientific output, *Journal of human resources* **50**(3), 728–758.

Brehm, M., Imberman, S. A. and Lovenheim, M. F.: 2017, Achievement effects of individual performance incentives in a teacher merit pay tournament, *Labour Economics* **44**, 133–150.

Bull, C., Schotter, A. and Weigelt, K.: 1987, Tournaments and piece rates: An experimental study, *The Journal of Political Economy* pp. 1–33.

Carpenter, J., Matthews, P. H. and Schirm, J.: 2010, Tournaments and office politics: Evidence from a real effort experiment, *The American Economic Review* **100**(1), 504–517.

Chan, H. F., Frey, B. S., Gallus, J. and Torgler, B.: 2014, Academic honors and performance, *Labour Economics* **31**, 188–204.

Chen, K.-P.: 2003, Sabotage in promotion tournaments, *Journal of Law, Economics, and Organization* **19**(1), 119–140.

Coupé, T., Smeets, V. and Warzynski, F.: 2005, Incentives, sorting and productivity along the career: Evidence from a sample of top economists, *Journal of Law, Economics, and Organization* **22**(1), 137–167.

Detmer, H. and Preissler, U.: 2004, Abenteuer w, strategien, risiken und chancen, *Forschung und Lehre* (6), 308–311.

Detmer, H. and Preissler, U.: 2005, Die neue professorenbesoldung, ein ueberblick, *Forschung und Lehre* (5), 256–258.

Detmer, H. and Preissler, U.: 2006, Die w-besoldung und ihre anwendung in den bundeslaendern, *Beitraege zur Hochschulforschung* **28**(2), 50–65.

DeVaro, J. and Waldman, M.: 2012, The signaling role of promotions: Further theory and empirical evidence, *Journal of Labor Economics* **30**(1), 91–147.

DHV, D. H.: 1999-2013, Forschung und lehre.
URL: *http://www.forschung-und-lehre.de/wordpress/?page_id=7*

Dickinson, D. L.: 1999, An experimental examination of labor supply and work intensities, *Journal of Labor Economics* **17**(4), 638–670.

Dickinson, D. and Villeval, M.-C.: 2008, Does monitoring decrease work effort?: The complementarity between agency and crowding-out theories, *Games and Economic behavior* **63**(1), 56–76.

Dilger, A.: 2013, Vor- und nachteile der w-besoldung, *Discussion Paper of the Institute for Organisational Economics* . Westfaelische Wilhelms-Universitaet Muenster.
URL: *http://www.econstor.eu/bitstream/10419/74542/1/747328412.pdf*

Dohmen, T. and Falk, A.: 2011, Performance pay and multidimensional sorting: Productivity, preferences, and gender, *The American Economic Review* pp. 556–590.

Ederer, F.: 2010, Feedback and motivation in dynamic tournaments, *Journal of Economics & Management Strategy* **19**(3), 733–769.

Ederer, F. and Manso, G.: 2013, Is pay for performance detrimental to innovation?, *Management Science* **59**(7), 1496–1513.

Enders, J.: 2001, A chair system in transition: Appointments, promotions, and gate-keeping in german higher education, *Higher Education* **41**(1-2), 3–25.

Expertenkommission: 2000, Bericht der expertenkommission - reform des hochschuldienstrechts, *Technical report*.
URL: *http://www.bmbf.de/pub/Bericht_.pdf*

Filmuniversitaet: 2014, Die filmuniversitaet.
URL: *http://www.filmuniversitaet.de/de/filmuniversitaet.html*

Fitzenberger, B. and Schulze, U.: 2014, Up or out: Research incentives and career prospects of postdocs in germany, *German Economic Review* **15**(2), 287–328.

Forschungsgemeinschaft, D. D.: 2016, Excellence initiative 2005-2017.
URL: *http://www.dfg.de/en/research_funding/programmes/excellence_initiative*

Freeman, R. B. and Gelber, A. M.: 2010, Prize structure and information in tournaments: Experimental evidence, *American Economic Journal: Applied Economics* **2**(1), 149–164.

Gibbons, R. and Murphy, K. J.: 1992, Optimal incentive contracts in the presence of career concerns: Theory and evidence, *Journal of political Economy* **100**(3), 468–505.

Glewwe, P., Ilias, N. and Kremer, M.: 2010, Teacher incentives, *American Economic Journal: Applied Economics* pp. 205–227.

Gourieroux, C., Monfort, A. and Trognon, A.: 1984, Pseudo maximum likelihood methods: applications to poisson models, *Econometrica: Journal of the Econometric Society* pp. 701–720.

Grigoriou, K. and Rothaermel, F. T.: 2014, Structural microfoundations of innovation the role of relational stars, *Journal of Management* **40**(2), 586–615.

Gruyter, D.: 2006, Kuerschners deutscher gelehrten kalender, cd-rom. Data set used for 21st edition but containing more and more comprehensive data.

Gruyter, D.: 2008, Kuerschners deutscher gelehrten kalender, cd-rom. Data set used for 22nd edition but containing more and more comprehensive data.

Gruyter, D.: 2013, Kuerschners deutscher gelehrten kalender online.
URL: *http://www.degruyter.com/databasecontent?dbid=kdgo&dbsource=%2Fdb%2Fkdgo*

Gürtler, O. and Kräkel, M.: 2010, Optimal tournament contracts for heterogeneous workers, *Journal of Economic Behavior & Organization* **75**(2), 180–191.

Haeck, C. and Verboven, F.: 2012, The internal economics of a university: evidence from personnel data, *Journal of Labor Economics* **30**(3), 591–626.

Hall, B. H. and Harhoff, D.: 2012, Recent research on the economics of patents, *Annu. Rev. Econ.* **4**(1), 541–565.

Handel, K. C.: 2005, *Die Umsetzung der Professorenbesoldungsreform in den Bundesländern*, CHE.
URL: *http://www.che.de/downloads/Profbesoldung_Laendervergleich_AP65.pdf*

Hansen, S. E.: 2012, Performance feedback with career concerns, *The Journal of Law, Economics, & Organization* **29**(6), 1279–1316.

Harbring, C., Irlenbusch, B. and Kräkel, M.: 2004, Ökonomische analyse der professorenbesoldungsreform in deutschland, *Bildung–33. Wirtschaftswissenschaftliches Seminar Ottobeuren. Tübingen: Mohr Siebeck*, pp. 197–219.

Hochschullehrerbund: 2009, Die professorengehaelter in der w-besoldung, art und umfang von berufungsverhandlungern, Online.
URL: *http://professur.fh-hannover.de/bewerbung/Hochschullehrerbund_Leistungsbezuege.pdf*

Holmström, B.: 1982, Managerial incentives schemes-a dynamic perspective, *Essays in Economics and Management in Honor of Lars Wahlbeck* .

Holmström, B.: 1999, Managerial incentive problems: A dynamic perspective, *The review of Economic studies* **66**(1), 169–182.

Hossain, T. and List, J. A.: 2012, The behavioralist visits the factory: Increasing productivity using simple framing manipulations, *Management Science* **58**(12), 2151–2167.

HRK, H.: 2014, Hochschulkompass.
URL: *http://www.hochschulkompass.de/hochschulen/hochschulen-in-deutschland-die-hochschulsuche.html*

HS-Geisenheim: 2014, Historie 1872 bis heute.
URL: *http://www.hs-geisenheim.de/hochschule/geschichtezahlen.html*

Hvide, H. K.: 2002, Tournament rewards and risk taking, *Journal of Labor Economics* **20**(4), 877–898.

Hvide, H. K. and Jones, B. F.: 2016, University innovation and the professor's privilege, *NBER working paper* .

Jaffe, A. B., Trajtenberg, M. and Henderson, R.: 1993, Geographic localization of knowledge spillovers as evidenced by patent citations, *The Quarterly Journal of Economics* **108**(3), 577–598.

Jones, B. F.: 2009, The burden of knowledge and the death of the renaissance man: is innovation getting harder?, *The Review of Economic Studies* **76**(1), 283–317.

Journal Citation Report: 2017.

Kelly, B., Papanikolaou, D., Seru, A. and Taddy, M.: 2018, Measuring technological innovation over the long run, *Technical report*, National Bureau of Economic Research.

Kräkel, M.: 2003, U-type versus j-type tournaments as alternative solutions to the unverifiability problem, *Labour Economics* **10**(3), 359–380.

Kräkel, M.: 2006, Zur reform der professorenbesoldung in deutschland, *Perspektiven der Wirtschaftspolitik* **7**(1), 105–126.

Kräkel, M.: 2008, Optimal risk taking in an uneven tournament game with risk averse players, *Journal of Mathematical Economics* **44**(11), 1219–1231.

Kräkel, M. and Sliwka, D.: 2004, Risk taking in asymmetric tournaments, *German Economic Review* **5**(1), 103–116.

Lach, S. and Schankerman, M.: 2004, Royalty sharing and technology licensing in universities, *Journal of the European Economic Association* **2**(2-3), 252–264.

Lach, S. and Schankerman, M.: 2008, Incentives and invention in universities, *The RAND Journal of Economics* **39**(2), 403–433.

Lavy, V.: 2009, Performance pay and teachers' effort, productivity, and grading ethics, *The American Economic Review* pp. 1979–2011.

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

Lazear, E. P.: 2000b, The power of incentives, *American Economic Review* pp. 410–414.

Lazear, E. P. and Oyer, P.: 2012, Personnel economics, *The Handbook of Organizational Economics* p. 479.

Lazear, E. P. and Rosen, S.: 1981, Rank-order tournaments as optimum labor contracts, *The Journal of Political Economy* **89**(5), 841–864.

Lazear, E. P. and Shaw, K. L.: 2007, Personnel economics: The economist's view of human resources, *The Journal of Economic Perspectives* pp. 91–114.

Leuven, E., Oosterbeek, H., Sonnemans, J. and Van Der Klaauw, B.: 2011, Incentives versus sorting in tournaments: Evidence from a field experiment, *Journal of Labor Economics* **29**(3), 637–658.

Levin, S. G. and Stephan, P. E.: 1991, Research productivity over the life cycle: Evidence for academic scientists, *The American Economic Review* pp. 114–132.

Lucas, R. E.: 1988, On the mechanisms of economic development, *Journal of Monetary Economics* **22**(1), 3–32.

Lucas, R. E.: 1990, Why doesn't capital flow from rich to poor countries?, *The American Economic Review* **80**(2), 92–96.

Lünstroth, P.: 2011, *Leistungslohn und Kooptation-Eine ökonomische Analyse der Reform der Professorenbesoldung*, PhD thesis, Universitaet Trier.
URL: *http://ubt.opus.hbz-nrw.de/volltexte/2012/739/pdf/Dissertation_Pia_Luenstroth.pdf*

Lutter, M. and Schröder, M.: 2016, Who becomes a tenured professor, and why? panel data evidence from german sociology, 1980–2013, *Research Policy* **45**(5), 999–1013.

Macchiavello, R. and Morjaria, A.: 2015, The value of relationships: evidence from a supply shock to kenyan rose exports, *American Economic Review* **105**(9), 2911–45.

Manso, G.: 2011, Motivating innovation, *The Journal of Finance* **66**(5), 1823–1860.

McCormack, J., Propper, C. and Smith, S.: 2014, Herding cats? management and university performance, *The Economic Journal* **124**(578), F534–F564.

Miklós-Thal, J. and Ullrich, H.: 2015, Belief precision and effort incentives in promotion contests, *The Economic Journal* **589**(125), 1952–1963.

Miklós-Thal, J. and Ullrich, H.: 2016, Career prospects and effort incentives: Evidence from professional soccer, *Management Science* **62**(6), 1645–1667.

Muralidharan, K. and Sundararaman, V.: 2011, Teacher performance pay: Experimental evidence from india, *The Journal of Political Economy* **119**(1), 39–77.

Oeffentlicher-Dienst: 2004, Gesetz ueber die erhoehung von dienst- und versorgungsbezuegen in bund und laendern 2003/2004. The parent page of the website, http://oeffentlicher-dienst.info/beamte/bund/a/, contains salary tables from 1927 till now (latest 01.03.2015-29.02.2016). Last checked 11-08.2015.
URL: *http://oeffentlicher-dienst.info/c/t/rechner/beamte/bund/a/2004i?id̄beamte-bund-2004i&matrix1̄*

Oyer, P. and Schaefer, S.: 2011, Personnel economics: Hiring and incentives, *Handbook of Labor Economics* **4**, 1769–1823.

Phelps, C., Heidl, R. and Wadhwa, A.: 2012, Knowledge, networks, and knowledge networks a review and research agenda, *Journal of Management* **38**(4), 1115–1166.

Preissler, U.: 2006, Zwischenbilanz professorenbesoldung - sichtweise und beratung des deutschen hochschulverbandes. Arbeitsgruppe Fortbildung im Sprecherkreis der Deutschen Universitätskanzler.
URL: *https://www.uni-muenster.de/.../prei_ler_jena_5.06.doc*

Pritchard, R.: 2006, Trends in the restructuring of german universities, *Comparative Education Review* **50**(1), 90–112.

Rauber, M. and Ursprung, H. W.: 2008, Life cycle and cohort productivity in economic research: The case of germany, *German Economic Review* **9**(4), 431–456.

Romer, P. M.: 1986, Increasing returns and long-run growth, *The journal of political economy* **94**(5), 1002–1037.

Romer, P. M.: 1990, Endogenous technological change, *Journal of political Economy* **95**(8), S71–S102.

Sauermann, H., Cohen, W. and Stephan, P.: 2010, Doing well or doing good? the motives, incentives and commercial activities of academic scientists and engineers, *DRUID summer conference*.

Schotter, A. and Weigelt, K.: 1992, Asymmetric tournaments, equal opportunity laws, and affirmative action: Some experimental results, *The Quarterly Journal of Economics* pp. 511–539.

Shearer, B.: 2004, Piece rates, fixed wages and incentives: Evidence from a field experiment, *The Review of Economic Studies* **71**(2), 513–534.

Shi, L.: 2010, Incentive effect of piece-rate contracts: Evidence from two small field experiments, *The BE Journal of Economic Analysis & Policy* **10**(1).

Silva, J. S. and Tenreyro, S.: 2006, The log of gravity, *The Review of Economics and statistics* **88**(4), 641–658.

Singh, J.: 2005, Collaborative networks as determinants of knowledge diffusion patterns, *Management science* **51**(5), 756–770.

Stephan, P. E.: 1996, The economics of science, *Journal of Economic literature* **34**(3), 1199–1235.

Teichler, U. and Bracht, O.: 2006, The academic profession in germany, *Reports of changing academic profession project workshop on quality, relevance, and governance in the changing academia: International perspectives*, pp. 129–150.

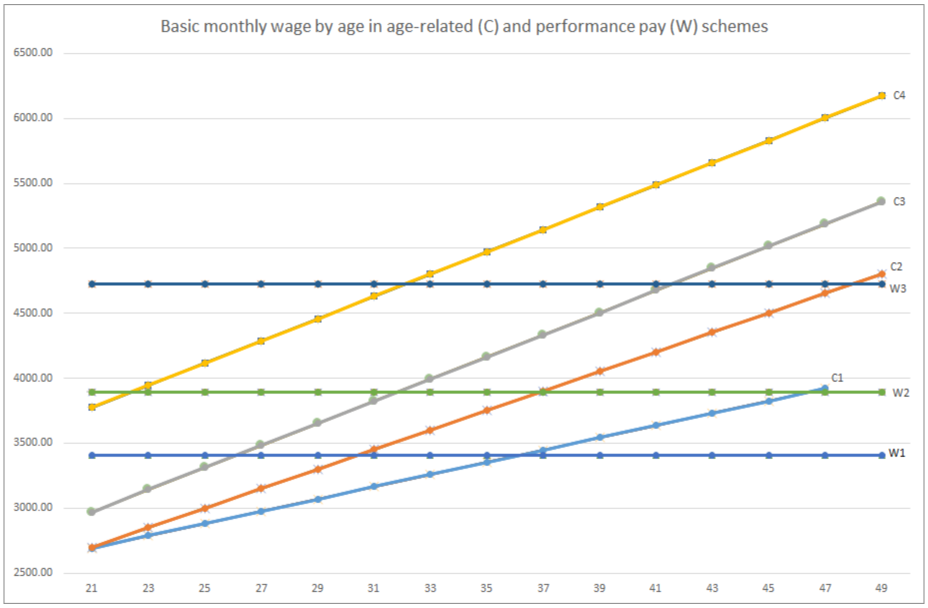
Von Proff, S., Buenstorf, G. and Hummel, M.: 2012, University patenting in germany before and after 2002: what role did the professors' privilege play?, *Industry and Innovation* **19**(1), 23–44.

Waldman, M.: 1990, Up-or-out contracts: A signaling perspective, *Journal of Labor Economics* **8**(2), 230–250.

Wuchty, S., Jones, B. F. and Uzzi, B.: 2007, The increasing dominance of teams in production of knowledge, *Science* **316**(5827), 1036–1039.

Ytsma, E.: 2017, Lone stars or constellations? the impact of performance pay on matching assortativeness in academia. working paper.

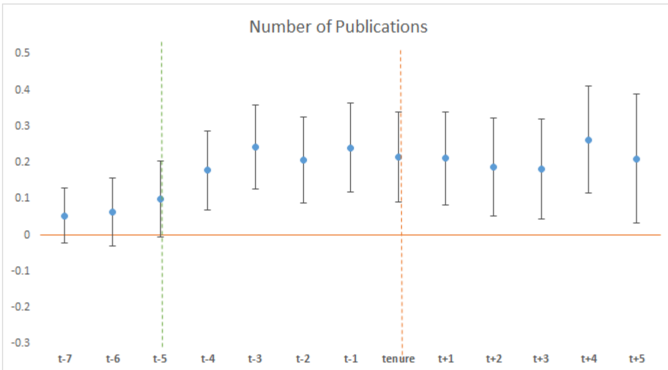
Fig. 1



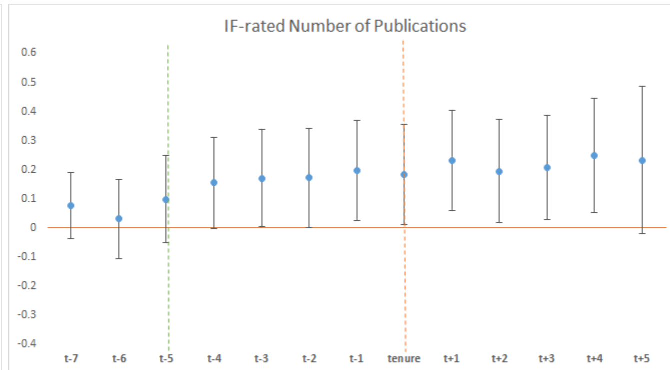
The figure above shows the monthly wages (in euros) by age for the various pay levels in the age-related (C) and performance pay (W) schemes. The depicted wages were valid as of 1 August 2004 in former West-German states; the corresponding monhtly wages in former East-German states were 92.5% of these (Detmer and Preissler 2006). Data source: Oeffentlicher-Dienst (2004).

Fig. 2: Pre-trends and Effect Dynamics - Treatment vs. Control Cohort

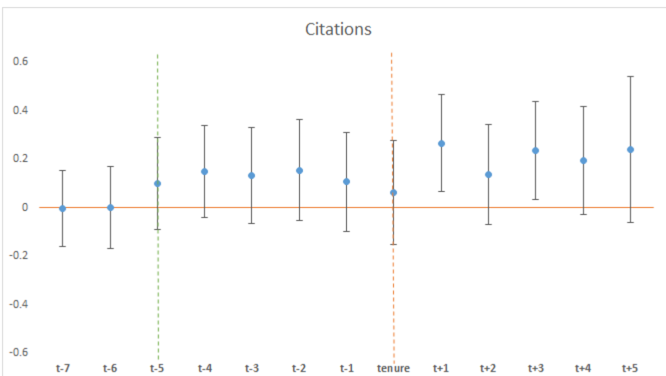
(a) Number of Publications



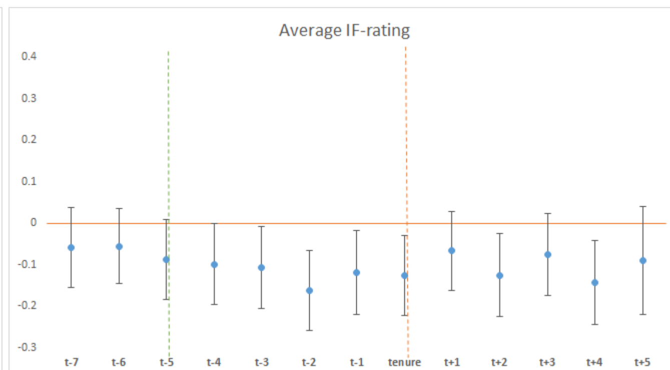
(b) Impact Factor-Rated Number of Publications



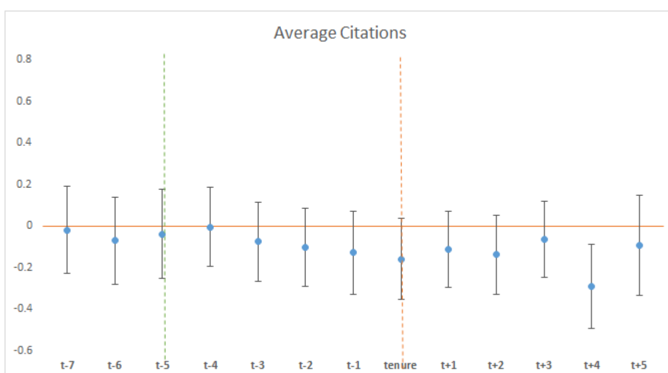
(c) Total Number of Citations



(d) Average Impact Factor Rating



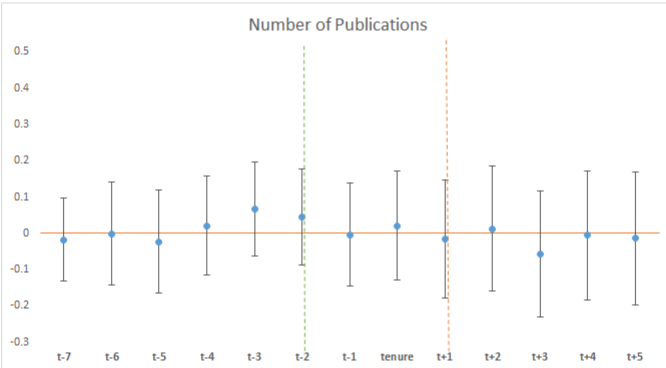
(e) Average Number of Citations



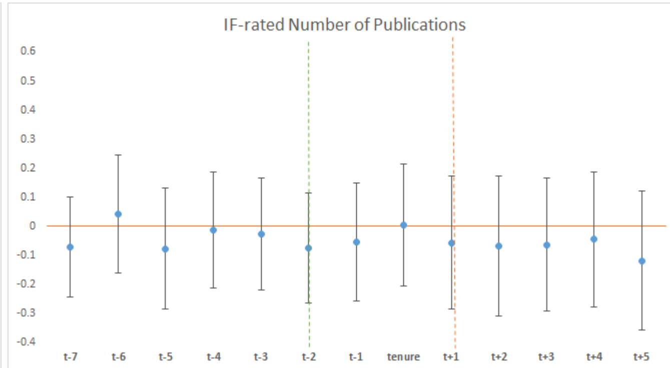
The figures depict the coefficient estimates and corresponding 95% confidence intervals of the interactions of a treatment dummy and time-to-tenure fixed effects in regressions with the following dependent variables: number of publications (panel a), impact factor-rated number of publications (panel b), total number of citations (panel c), average impact factor rating (panel d), and average number of citations. I control for 15 time-to-tenure fixed effects and a full set of interactions with the treatment variable as well as individual and time fixed effects. The dependent variables are calculated for academic i in field f and year t, backdated by average publication lag in field f as reported n Björk and Solomon (2013). The unit of observation is academic i. The sample is restricted to academics who started their first tenured affiliation at a German public university in 2002, 2003, 2004, 2005, 2006 or 2007 and did not hold a foreign affiliation immediately prior. *Treatment* is 1 if an academic makes tenure at a public university in 2005, 2006 or 2007 and 0 if (s)he becomes tenured in 2002, 2003 or 2004 (the control group). Estimation as conditional quasi-maximum likelihood fixed effects Poisson model with robust standard errors.

Fig. 3: Pre-trends and Effect Dynamics - Placebo Experiment

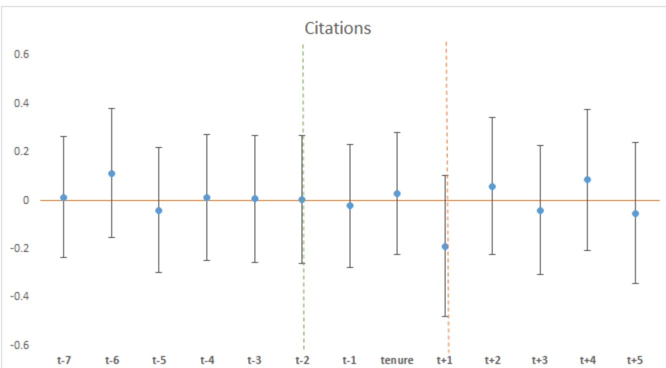
(a) Number of Publications



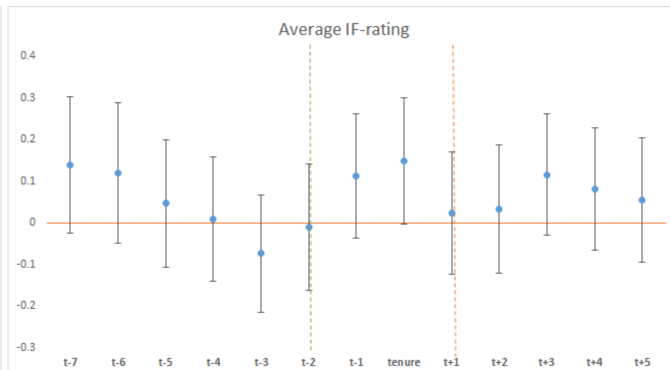
(b) Impact Factor-Rated Number of Publications



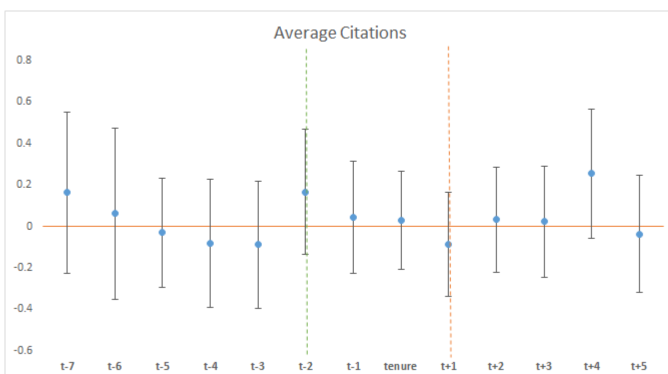
(c) Total Number of Citations



(d) Average Impact Factor Rating



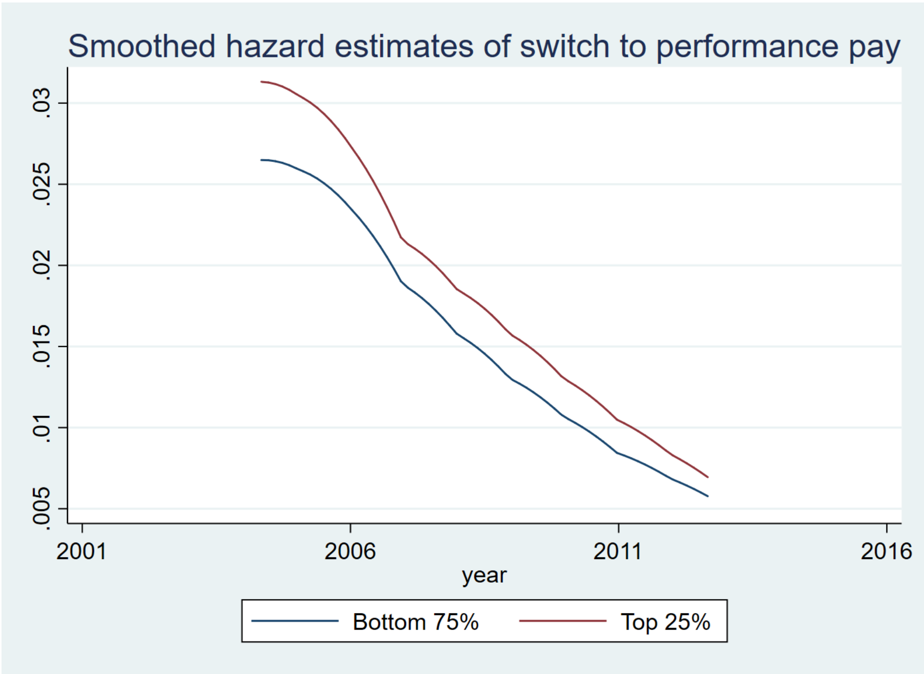
(e) Average Number of Citations



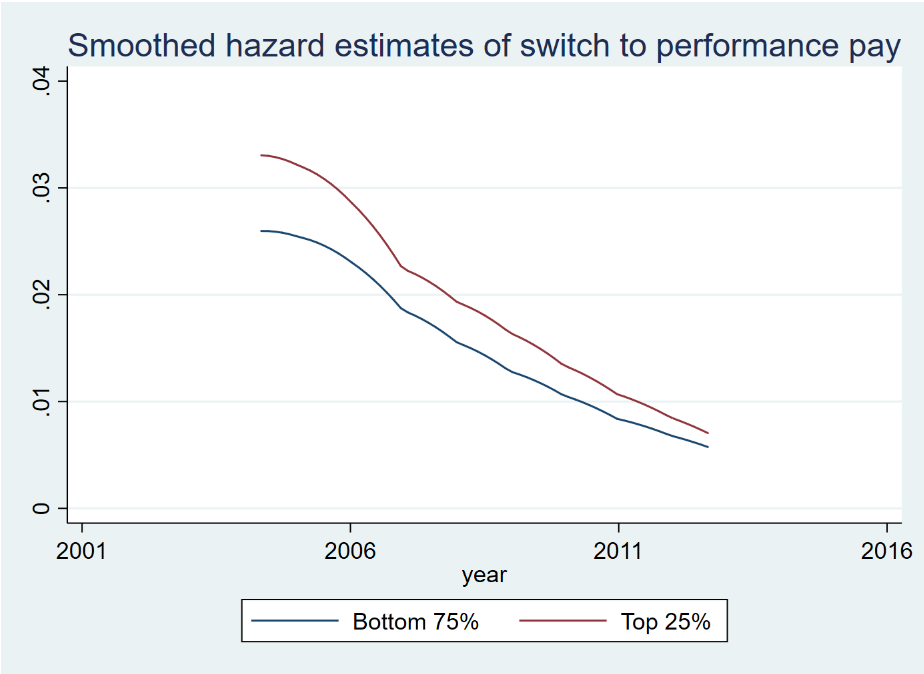
The figures depict the coefficient estimates and corresponding 95% confidence intervals of the interactions of a treatment dummy and time-to-tenure fixed effects in regressions with the following dependent variables: number of publications (panel a), impact factor-rated number of publications (panel b), total number of citations (panel c), average impact factor rating (panel d), and average number of citations. I control for 15 time-to-tenure fixed effects and a full set of interactions with the treatment variable as well as individual and time fixed effects. The dependent variables are calculated for academic i in field f and year t, backdated by average publication lag in field f as reported n Björk and Solomon (2013). The unit of observation is academic i. The sample is restricted to academics who started their first tenured affiliation at a German public university in 2001, 2002, 2003 or 2004 and did not hold a foreign affiliation immediately prior. *Treatment* is 1 if an academic makes tenure at a public university in 2003 or 2004 and 0 if they becomes tenured in 2001 or 2002 (the control group). Estimation as conditional quasi-maximum likelihood fixed effects Poisson model with robust standard errors.

Fig. 4: Smoothed Hazard Estimates of Switch to Performance Pay

(a) Quantiles Based on Impact Factor-Rated Number of Publications



(b) Quantiles Based on Impact Factor-Rated Number of Publications Weighted by Number of Authors



The above figures depict the Epanechnikov kernel-density estimates of the hazard function for switching to the performance pay scheme for academics in the top quartile (red line) and bottom three quartiles (blue line) of the average productivity distribution. Switches to performance pay are defined as any contract, position or affiliation change after 2005. Quantiles are determined based on three pre-implementation year averages (2002, 2003 and 2004) of, respectively, the number of impact factor-rated publications (Figure a) and the number of impact factor-rated publications weighted by number of authors on a paper (Figure b) of academic i in field f and year t, lagged by average publication lag in field f as reported in Björk and Solomon (2013), Quantiles are derived separately by academic field and tenure year. The sample is restricted to academics who held a tenured affiliation at a public univesity before 2005.

Tab. 1: Summary Statistics

Panel A: Research productivity variables	N	Mean	SD	Median	Min	Max
Number of publications	108363	2.907	7.134	0	0	195
IF-rated publications	108363	9.874	31.38	0	0	1082.363
Citations	108363	102.271	333.992	0	0	11086
Average IF-rating	48552	2.709	2.848	2.075	0	52.589
Average citations	48552	32.691	61.322	19.118	0	2759
Maximum citations	48552	99.164	216.016	42	0	8796
Minimum citations	48552	10.612	41.576	1	0	2759
Panel B: Research productivity variables, weighted by number of authors						
Number of publications	108363	0.707	1.516	0	0	45.967
IF-rated publications	108363	1.876	5.424	0	0	156.061
Citations	108363	19.278	62.726	0	0	3987.705
Average IF-rating	48552	2.544	2.829	1.897	0	52.589
Average citations	48552	29.27	57.714	16.521	0	2759
Maximum citations	48552	21.485	56.744	9	0	3126
Minimum citations	48552	2.903	11.464	0.143	0	786
Panel C: Novelty Metrics						
Comparison with publications in academic field						
Minimum backward cosine similarity	39602	0	0	0	0	0.001
Maximum backward cosine similarity	39602	0.003	0.003	0.002	0	0.215
Average backward cosine similarity	39602	0.277	0.345	0.202	0	19.267
Minimum forward cosine similarity	38347	0	0	0	0	0.001
Maximum forward cosine similarity	38347	0.003	0.003	0.002	0	0.205
Average forward cosine similarity	38347	0.286	0.329	0.215	0	10.244
Comparison with own publications						
Minimum backward cosine similarity	32938	0	0.001	0	0	0.026
Maximum backward cosine similarity	32938	0.002	0.003	0.001	0	0.048
Average backward cosine similarity	32938	0.002	0.003	0.001	0	0.121
Minimum forward cosine similarity	33039	0	0.001	0	0	0.027
Maximum forward cosine similarity	33039	0.002	0.003	0.001	0	0.047
Average forward cosine similarity	33039	0.002	0.004	0.001	0	0.151
Ages at milestones, restricted to tenured professors at public universities						
Age at habilitation	190594	37.21	4.69	37	23	64
Age at PhD	171089	29.63	3.65	29	20	61
Years between PhD and habilitation	127788	8.02	3.77	7	1	34
Age at first tenured affiliation	57886	43.89	6.69	43	23	64

Notes: The unit of observation is academic *i*. The sample is restricted to academics who started their first tenured affiliation at a German public university in 2002, 2003, 2004, 2005, 2006 or 2007 and includes data from 1993 until and including 2010.

Tab. 2: Baseline Effort Effect Estimation

Panel A: Unweighted Publication Variables

	# Publications	IF-rated publications	Citations	Average IF-rating	Average citations
Post'02 * Treatment	0.168***	0.133***	0.129**	-0.094***	-0.103*
	(0.038)	(0.051)	(0.063)	(0.035)	(0.062)
Tenure * Treatment	-0.011	0.026	0.017	0.025	-0.018
	(0.040)	(0.052)	(0.064)	(0.036)	(0.067)
Number of Observations	83937	74326	78308	47052	47789
Number of Individuals	4671	4136	4357	3917	4110
Log Likelihood	-136647.270	-338248.006	-4508200.290	-70677.097	-736179.301
Chi-squared	2863.231	2856.117	1101.473	611.818	476.042

Panel A.II: Maximum and Minimum Number of Citations

	Maximum citations	Minimum citations
Post'02 * Treatment	0.001	-0.176
	(0.066)	(0.136)
Tenure * Treatment	-0.063	-0.058
	(0.078)	(0.171)
Number of Observations	47789	45928
Number of Individuals	4110	3967
Log Likelihood	-2366039.795	-525485.210
Chi-squared	430.387	457.933

Panel B: Publication Variables Weighted by Number of Authors

	# Publications	IF-rated publications	Citations	Average IF-rating	Average citations
Post'02 * Treatment	0.177***	0.110**	0.111	-0.099***	-0.105
	(0.036)	(0.053)	(0.072)	(0.037)	(0.068)
Tenure * Treatment	0.015	0.117*	0.117	0.028	0.011
	(0.042)	(0.064)	(0.079)	(0.040)	(0.077)
Number of Observations	83937	74326	78308	47052	47789
Number of Individuals	4671	4136	4357	3917	4110
Log Likelihood	-64857.187	-109658.385	-1074350.682	-70155.980	-689216.004
Chi-squared	1941.485	1785.932	771.481	417.241	606.337

Panel B.II: Maximum and Minimum Number of Citations, Weighted by Number of Authors

	Maximum citations	Minimum citations
Post'02 * Treatment	0.012	-0.133
	(0.094)	(0.134)
Tenure * Treatment	0.079	-0.152
	(0.100)	(0.143)
Number of Observations	47789	45928
Number of Individuals	4110	3967
Log Likelihood	-634373.285	-150432.757
Chi-squared	394.960	881.988

Notes: * denotes significance at 10%, ** at 5% and *** at 1%. The unit of observation is academic i. The sample is restricted to academics who started their first tenured affiliation at a German public university in 2002 to 2007 (excluding those with a foreign affiliation directly prior to this) and includes data from 1993 until and including 2012. The dependent variables in Panels A.I and B.I are, respectively, the number of publications, the impact factor-rated number of publications, the total number of citations to all publications published in a given year, the average impact factor rating of publications published in a given year, and the average number of citations to publications published in a given year. The dependent variables in Panels A.II and B.II are, respectively, the maximum and minimum number of citations to publications published in a given year. All dependent variables are defined for academic i in field f and year t, lagged by average publication lag in field f as reported in Björk and Solomon (2013). $Post'02$ is 0 before 2002 and 1 thereafter, $Tenure$ is 0 before 2005 and 1 thereafter, and $Treatment$ is 1 if an academic makes tenure at a public university in 2005, 2006 or 2007 and 0 if they make tenure at a public university in 2002, 2003 or 2004 (the control group). All specifications control for year and individual fixed effects and fifteen time-to-tenure fixed effects (from seven years before the tenure year to seven years after). Estimation as conditional quasi-maximum likelihood estimation of Poisson fixed effects models with robust standard errors.

Tab. 3: Robustness Checks - Alternative Models

Panel A: Linear FE Model	# Publications	IF-rated publications	Citations	Average IF-rating	Average citations
Post'02 * Treatment	0.300**	0.875	7.634	-0.216**	-3.561*
	(0.120)	(0.594)	(6.770)	(0.093)	(2.160)
Tenure * Treatment	-0.197	-0.489	-3.141	0.069	-0.421
	(0.166)	(0.815)	(8.319)	(0.102)	(2.365)
Number of Observations	108363	108363	108363	48552	48552
Number of Individuals	6039	6039	6039	4671	4671
Log Likelihood	-294977.403	-471443.345	-740680.174	-102809.237	-262588.501
Panel B: With absolute(time-to-tenure) instead of time-to-tenure dummies					
Post'02 * Treatment	0.166***	0.166***	0.180***	-0.037*	-0.023
	(0.031)	(0.041)	(0.047)	(0.022)	(0.043)
Tenure * Treatment	-0.042	-0.038	-0.041	-0.025	-0.045
	(0.034)	(0.044)	(0.051)	(0.025)	(0.049)
Number of Observations	83937	74326	78308	47052	47789
Number of Individuals	4671	4136	4357	3917	4110
Log Likelihood	-136676.95	-338396.973	-4510367.867	-70699.505	-736831.695
Chi-squared	2791.631	2771.287	1091.63	578.407	446.108
Panel C: Without Switchers					
Post'02 * Treatment	0.207***	0.178***	0.158**	-0.093**	-0.103
	(0.040)	(0.055)	(0.067)	(0.040)	(0.066)
Tenure * Treatment	-0.009	0.004	0.001	0.021	-0.012
	(0.044)	(0.059)	(0.074)	(0.042)	(0.075)
Number of Observations	73272	64920	68380	41135	41764
Number of Individuals	4078	3613	3805	3421	3579
Log Likelihood	-118583.219	-296315.765	-3973456.609	-62391.047	-656767.678
Chi-squared	2508.109	2425.435	933.761	519.380	402.585
Panel D: Synthetic Cohorts (assignment to treat-control based on synthetic tenure year; actual tenure year controlled for)					
Post'02 * Treatment	0.099**	0.146**	0.027	-0.022	-0.132**
	(0.041)	(0.059)	(0.069)	(0.031)	(0.055)
Tenure * Treatment	-0.029	-0.046	-0.028	0.042	0.074
	(0.044)	(0.064)	(0.073)	(0.033)	(0.059)
Number of Observations	36548	31116	33204	20975	21403
Number of Individuals	2033	1731	1847	1636	1752
Log Likelihood	-60018.452	-148221.305	-1863790.551	-31268.893	-287069.199
Chi-squared	1291.025	1200.136	524.503	328.138	269.700

Notes: * denotes significance at 10%, ** at 5% and *** at 1%. The unit of observation is academic i. The sample is restricted to academics who started their first tenured affiliation at a German public university in 2002 to 2007 (excluding those with a foreign affiliation directly prior to this) and includes data from 1993 until and including 2012. The dependent variables are, respectively, the number of publications, the impact factor-rated number of publications, the total number of citations to all publications published in a given year, the average impact factor rating of publications published in a given year, and the average number of citations to publications published in a given year. All dependent variables are defined for academic i in field f and year t, lagged by average publication lag in field f as reported in Björk and Solomon (2013). $Post'02$ is 0 before 2002 and 1 thereafter, $Tenure$ is 0 before 2005 and 1 thereafter, and $Treatment$ is 1 if an academic makes tenure at a public university in 2005, 2006 or 2007 and 0 if they make tenure at a public university in 2002, 2003 or 2004 (the control group). All specifications control for year and individual fixed effects and fifteen time-to-tenure fixed effects (from seven years before the tenure year to seven years after). Estimation as conditional quasi-maximum likelihood estimation of Poisson fixed effects models with robust standard errors.

Tab. 4: Heterogeneous Effort Effect Estimation

Panel A: Top Quartile Academics	# Publications	IF-rated publications	Citations	Average IF-rating	Average citations
Post'02 * Treatment	0.144***	0.071	0.104	-0.185***	-0.191*
	(0.050)	(0.064)	(0.079)	(0.058)	(0.108)
Tenure * Treatment	0.001	0.016	0.002	0.102	-0.026
	(0.050)	(0.062)	(0.078)	(0.065)	(0.113)
Number of Observations	21616	21378	23065	11343	11400
Number of Individuals	1203	1189	1283	824	833
Log Likelihood	-54174.311	-153690.389	-2308500.808	-21893.611	-276939.181
Chi-squared	1431.931	1534.550	670.836	281.771	634.649
Panel B: Bottom 3 Quartile Academics					
Post'02 * Treatment	0.184***	0.242***	0.165*	-0.037	-0.042
	(0.056)	(0.086)	(0.097)	(0.042)	(0.068)
Tenure * Treatment	0.002	0.064	0.067	-0.018	-0.046
	(0.063)	(0.087)	(0.102)	(0.042)	(0.082)
Number of Observations	59337	50162	52439	31505	31948
Number of Individuals	3302	2792	2918	2389	2495
Log Likelihood	-76718.648	-167967.463	-1967318.384	-43239.362	-376507.832
Chi-squared	2142.238	1892.701	1013.084	1327.835	372.162
Panel C: Top Quartile vs. Bottom 3 quartiles					
Post'02 * Q4	-0.303***	-0.246***	-0.344***	-0.278***	-0.507***
	(0.041)	(0.053)	(0.054)	(0.025)	(0.053)
Post'02 * Treatment	0.167***	0.214***	0.144*	-0.097**	-0.117*
	(0.048)	(0.068)	(0.079)	(0.039)	(0.068)
Post'02 * Treatment * Q4	-0.003	-0.125*	-0.025	0.014	0.026
	(0.056)	(0.075)	(0.084)	(0.042)	(0.083)
Tenure * Q4	0.125***	0.073**	0.077	0.042	0.092
	(0.025)	(0.036)	(0.058)	(0.030)	(0.061)
Tenure * Treatment	0.120***	0.181***	0.148*	0.065*	0.049
	(0.045)	(0.060)	(0.076)	(0.038)	(0.072)
Tenure * Treatment * Q4	-0.202***	-0.227***	-0.189***	-0.101***	-0.217***
	(0.034)	(0.047)	(0.061)	(0.029)	(0.060)
Number of Observations	80953	71540	75504	42848	43348
Number of Individuals	4505	3981	4201	3213	3328
Log Likelihood	-131098.018	-322615.021	-4300277.3	-65371.248	-665536.009
Chi-squared	3232.728	3085.574	1348.783	946.904	615.399

Notes: * denotes significance at 10%, ** at 5% and *** at 1%. The unit of observation is academic i. The sample is restricted to academics who started their first tenured affiliation at a German public university in 2002 to 2007 (excluding those with a foreign affiliation directly prior to this) and includes data from 1993 until and including 2012. The dependent variables in Panels A.I and B.I are, respectively, the number of publications, the impact factor-rated number of publications, the total number of citations to all publications published in a given year, the average impact factor rating of publications published in a given year, and the average number of citations to publications published in a given year. The dependent variables in Panels A.II and B.II are, respectively, the maximum and minimum number of citations to publications published in a given year. All dependent variables are defined for academic i in field f and year t, lagged by average publication lag in field f as reported in Björk and Solomon (2013). $Post'02$ is 0 before 2002 and 1 thereafter, $Tenure$ is 0 before 2005 and 1 thereafter, and $Treatment$ is 1 if an academic makes tenure at a public university in 2005, 2006 or 2007 and 0 if they make tenure at a public university in 2002, 2003 or 2004 (the control group). Productivity quantiles are determined on the basis of the averages of the various dependent variables over the three pre-announcement years 1999, 2000 and 2001, separately by academic field and tenure cohort. All specifications control for year and individual fixed effects and fifteen time-to-tenure fixed effects (from seven years before the tenure year to seven years after). Estimation as conditional quasi-maximum likelihood estimation of Poisson fixed effects models with robust standard errors.

Tab. 5: Heterogeneous Treatment Effects By Academic Field

Panel A: Science	# Publications	IF-rated publications	Citations	Average IF-rating	Average citations
Post'02 * Q4	-0.327***	-0.372***	-0.323***	-0.327***	-0.515***
	(0.056)	(0.072)	(0.076)	(0.036)	(0.078)
Post'02 * Treatment	0.155**	0.127	0.231*	-0.109*	-0.100
	(0.067)	(0.092)	(0.119)	(0.062)	(0.101)
Post'02 * Treatment * Q4	-0.055	-0.098	-0.161	0.011	0.003
	(0.079)	(0.104)	(0.125)	(0.068)	(0.130)
Tenure * Q4	0.076**	0.116**	0.016	0.060	0.049
	(0.037)	(0.053)	(0.074)	(0.048)	(0.074)
Tenure * Treatment	0.153**	0.236***	0.116	0.107*	-0.052
	(0.066)	(0.081)	(0.112)	(0.059)	(0.104)
Tenure * Treatment * Q4	-0.217***	-0.282***	-0.178*	-0.161***	-0.144
	(0.051)	(0.065)	(0.098)	(0.042)	(0.094)
Number of Observations	28042	27520	27646	19039	19040
Number of Individuals	1560	1531	1538	1340	1339
Log Likelihood	-53919.215	-144395.302	-2178935.955	-30894.089	-344911.194
Chi-squared	1795.177	2112.709	689.675	683.510	307.123

Panel B: Applied Science					
Post'02 * Q4	-0.269***	-0.156**	-0.340***	-0.238***	-0.459***
	(0.061)	(0.076)	(0.078)	(0.036)	(0.072)
Post'02 * Treatment	0.168**	0.283***	0.065	-0.065	-0.133
	(0.070)	(0.096)	(0.110)	(0.048)	(0.092)
Post'02 * Treatment * Q4	0.040	-0.152	0.105	0.020	0.043
	(0.083)	(0.105)	(0.118)	(0.054)	(0.109)
Tenure * Q4	0.162***	0.049	0.146	0.027	0.134
	(0.034)	(0.048)	(0.093)	(0.037)	(0.111)
Tenure * Treatment	0.091	0.159*	0.186*	0.026	0.176*
	(0.065)	(0.087)	(0.107)	(0.046)	(0.105)
Tenure * Treatment * Q4	-0.180***	-0.201***	-0.192**	-0.056	-0.260***
	(0.049)	(0.065)	(0.079)	(0.043)	(0.081)
Number of Observations	29773	29348	29169	19841	19834
Number of Individuals	1658	1634	1624	1402	1401
Log Likelihood	-61997.055	-164075.453	-1853816.140	-30810.064	-268245.487
Chi-squared	1564.143	1441.839	789.367	344.549	342.013

Tab. 5: Heterogeneous Treatment Effects By Academic Field

Panel C: Social Science	# Publications	IF-rated publications	Citations	Average IF-rating	Average citations
Post'02 * Q4	-0.375***	-0.866***	-1.194***	-0.358***	-0.732***
	(0.124)	(0.193)	(0.228)	(0.101)	(0.174)
Post'02 * Treatment	0.341**	0.025	-0.017	-0.271	-0.169
	(0.164)	(0.300)	(0.399)	(0.228)	(0.283)
Post'02 * Treatment * Q4	-0.255	-0.246	-0.182	-0.154	-0.085
	(0.182)	(0.301)	(0.387)	(0.157)	(0.281)
Tenure * Q4	0.131	0.262	-0.059	-0.021	0.013
	(0.107)	(0.175)	(0.269)	(0.113)	(0.204)
Tenure * Treatment	0.270*	0.499**	0.372	0.137	0.252
	(0.142)	(0.248)	(0.344)	(0.267)	(0.247)
Tenure * Treatment * Q4	-0.367***	-0.472***	-0.472*	-0.092	-0.445***
	(0.108)	(0.168)	(0.249)	(0.107)	(0.172)
Number of Observations	13108	12244	11887	3462	3454
Number of Individuals	729	681	661	393	391
Log Likelihood	-10162.547	-11407.902	-215604.240	-3284.157	-44420.943
Chi-squared	773.488	722.098	317.928	293.667	266.027
Panel D: Humanities					
Post'02 * Q4	-0.778***	-1.270	-1.529***	-1.489***	-1.378***
	(0.217)	(0.828)	(0.510)	(0.508)	(0.388)
Post'02 * Treatment	0.476	0.029	1.231**	1.063	1.480***
	(0.293)	(0.970)	(0.593)	(1.251)	(0.548)
Post'02 * Treatment * Q4	-0.065	-0.996	0.145	-0.160	0.612
	(0.293)	(0.981)	(0.736)	(1.216)	(0.570)
Tenure * Q4	0.015	-0.641	-0.019	1.325*	0.572
	(0.195)	(0.632)	(0.512)	(0.713)	(0.383)
Tenure * Treatment	-0.292	0.235	-1.974**	-1.565	-2.206***
	(0.286)	(0.908)	(0.917)	(1.237)	(0.683)
Tenure * Treatment * Q4	-0.189	0.175	-0.414	0.836	-0.348
	(0.194)	(0.609)	(0.718)	(1.117)	(0.613)
Number of Observations	10030	2428	6802	506	1020
Number of Individuals	558	135	378	78	197
Log Likelihood	-4448.843	-963.401	-22416.469	-220.324	-4201.452
Chi-squared	183.736	1081.615	680.362	519.248	655.384

Notes: * denotes significance at 10%, ** at 5% and *** at 1%. The unit of observation is academic i. The sample is restricted to academics who started their first tenured affiliation at a German public university in 2002 to 2007 (excluding those with a foreign affiliation directly prior to this) and includes data from 1993 until and including 2012. The dependent variables are, respectively, the number of publications, the impact factor-rated number of publications, the total number of citations to all publications published in a given year, the average impact factor rating of publications published in a given year, and the average number of citations to publications published in a given year. All dependent variables are defined for academic i in field f and year t, lagged by average publication lag in field f as reported in Björk and Solomon (2013). $Post'02$ is 0 before 2002 and 1 thereafter, $Tenure$ is 0 before 2005 and 1 thereafter, and $Treatment$ is 1 if an academic makes tenure at a public university in 2005, 2006 or 2007 and 0 if they make tenure at a public university in 2002, 2003 or 2004 (the control group). Productivity quantiles are determined on the basis of the averages of the various dependent variables over the three pre-announcement years 1999, 2000 and 2001, separately by academic field and tenure cohort. All specifications control for year and individual fixed effects and fifteen time-to-tenure fixed effects (from seven years before the tenure year to seven years after). Estimation as conditional quasi-maximum likelihood estimation of Poisson fixed effects models with robust standard errors.

Tab. 6: Novelty Estimation - By Productivity Quantile

Panel A. Within field comparison novelty metrics

	Backward similarity			Forward Similarity		
A.I: Top Quartile Academics	Minimum	Maximum	Average	Minimum	Maximum	Average
Post'02 * Treatment	-0.177	0.038	0.042	0.029	0.170**	0.073
	(0.132)	(0.068)	(0.056)	(0.139)	(0.070)	(0.063)
Tenure * Treatment	0.340**	-0.051	0.022	0.114	-0.032	-0.019
	(0.172)	(0.070)	(0.062)	(0.135)	(0.069)	(0.055)
Number of Observations	14609	14609	14609	14621	14621	14621
Number of Individuals	1150	1150	1150	1149	1149	1149
Log Likelihood	-0.089	-126.431	-5755.428	-0.091	-125.072	-6001.988
Chi-squared	716.390	115.499	1440.996	1061.534	96.749	1261.526
A.II: Bottom 3 Quartile Academics						
Post'02 * Treatment	0.048	0.003	0.062	-0.129	-0.001	-0.002
	(0.130)	(0.067)	(0.082)	(0.154)	(0.064)	(0.073)
Tenure * Treatment	-0.187	0.001	-0.066	-0.317**	-0.061	-0.036
	(0.151)	(0.060)	(0.105)	(0.159)	(0.094)	(0.062)
Number of Observations	23357	23357	23357	22119	22119	22119
Number of Individuals	2489	2489	2489	2404	2404	2404
Log Likelihood	-0.182	-126.677	-9179.561	-0.178	-116.478	-8769.347
Chi-squared	769.601	98.374	966.021	892.126	76.700	665.049

Panel B. Within author comparison novelty metrics

B.I: Top Quartile Academics						
Post'02 * Treatment	-0.402	0.047	0.042	0.740*	0.147	0.098
	(0.439)	(0.092)	(0.114)	(0.422)	(0.100)	(0.133)
Tenure * Treatment	-0.795	-0.059	0.015	-0.803**	-0.091	-0.204
	(0.644)	(0.095)	(0.142)	(0.325)	(0.101)	(0.180)
Number of Observations	13381	13381	13381	13624	13624	13624
Number of Individuals	1093	1093	1093	1088	1088	1088
Log Likelihood	-1.520	-84.804	-71.424	-1.372	-85.220	-80.760
Chi-squared	293.211	75.491	140.578	356.511	95.434	77.766
B.II: Bottom 3 Quartile Academics						
Post'02 * Treatment	0.012	0.081	-0.053	-0.583*	0.061	-0.062
	(0.298)	(0.115)	(0.116)	(0.312)	(0.106)	(0.108)
Tenure * Treatment	-0.207	-0.008	0.012	0.338	-0.001	-0.068
	(0.285)	(0.100)	(0.103)	(0.277)	(0.093)	(0.095)
Number of Observations	18438	18438	18438	18300	18300	18300
Number of Individuals	2008	2008	2008	1994	1994	1994
Log Likelihood	-4.630	-67.825	-54.593	-4.406	-65.444	-59.756
Chi-squared	250.022	53.811	73.038	262.344	53.216	54.368

Notes: * denotes significance at 10%, ** at 5% and *** at 1%. The unit of observation is academic i. The sample is restricted to academics who started their first tenured affiliation at a German public university in 2002 to 2007 (excluding those with a foreign affiliation directly prior to this) and includes data from 1993 until and including 2012. The dependent variables in Panels A.I and A.II are the minimum, maximum and average backward and forward similarity of the publications of author i in year t compared to publications in the same field, while in Panels B.I and B.II the comparison is between papers by the same author. All dependent variables are defined for academic i in field f and year t, lagged by average publication lag in field f as reported in Björk and Solomon (2013). *Post'02* is 0 before 2002 and 1 thereafter, *Tenure* is 0 before 2005 and 1 thereafter, and *Treatment* is 1 if an academic makes tenure at a public university in 2005, 2006 or 2007 and 0 if they make tenure at a public university in 2002, 2003 or 2004 (the control group). Productivity quantiles are determined on the basis of the average of the impact factor-rated number of publications calculated over the three pre-announcement years 1999, 2000 and 2001, separately by academic field and tenure cohort. All specifications control for year and individual fixed effects and fifteen time-to-tenure fixed effects (from seven years before the tenure year to seven years after). Estimation as conditional quasi-maximum likelihood estimation of Poisson fixed effects models with robust standard errors.

Tab. 7: Selection Analysis

	Treatment vs. Control		Placebo	
Panel A: With Continuous Productivity Variable	1a	1b	2a	2b
Treatment	-0.564***	0.658	-0.185***	-0.933
	(0.047)	(0.401)	(0.067)	(0.594)
Age	-0.142***	-0.126***	-0.155***	-0.164***
	(0.005)	(0.007)	(0.007)	(0.010)
Avg Productivity	0.002**	-0.001	0.002**	0.003**
	(0.001)	(0.002)	(0.001)	(0.001)
Treatment * Age		-0.028***		0.016
		(0.009)		(0.012)
Treatment * Avg Productivity		0.004**		-0.001
		(0.002)		(0.002)
Constant	3.061***	2.381***	3.233***	3.601***
	(0.242)	(0.329)	(0.335)	(0.427)
Number of Observations	80131	80131	51431	51431
Number of Subjects	14972	14972	6960	6960
Number of Switches	2435	2435	1099	1099
Log Likelihood	-7553.739	-7545.484	-3234.631	-3233.591
Chi-squared	1421.044	1365.450	576.432	601.114
p				
Panel B: Above vs Below Median Productivity				
Treatment	-0.553***	0.545	-0.185***	-0.929
	(0.047)	(0.403)	(0.067)	(0.597)
Age	-0.142***	-0.127***	-0.154***	-0.162***
	(0.005)	(0.007)	(0.007)	(0.010)
Above Median	0.164***	0.068	0.243***	0.240***
	(0.049)	(0.063)	(0.072)	(0.092)
Treatment * Age		-0.026***		0.016
		(0.009)		(0.012)
Treatment * Above Median		0.188**		0.007
		(0.084)		(0.124)
Constant	3.039***	2.420***	3.177***	3.536***
	(0.242)	(0.330)	(0.334)	(0.429)
Number of Observations	80131	80131	51431	51431
Number of Subjects	14972	14972	6960	6960
Number of Switches	2435	2435	1099	1099
Log Likelihood	-7549.737	-7542.467	-3230.879	-3230.053
Chi-squared	1434.546	1387.434	596.157	614.786
p				

Notes: * denotes significance at 10%, ** at 5% and *** at 1%. The unit of observation is academic *i*. The table reports estimation results of Weibull propotional hazard models of selection into performance pay. Any first affiliation, position or contract change after implementation of the pay reform (as of 2005) is considered a switch to performance pay (the "failure" event) and the time until such a switch is used as duration variable. Academics are "at risk" of switching after their most recent affiliation/position/contract change if they are tenured, at a public university and not retired. The treatment variable is 1 for academics who have made tenure before 2005 and 0 for those who make tenure afterwards. "Avg Productivity" is calculated as three year pre-implementation averages (2002-2004) of the unweighted impact factor-rated number of publications. "Above Median" is a dummy variable that is 1 for academics whose pre-reform average productivity is above the median average productivity of academics in the same field and tenure cohort. The synthetic age is calculated using the average age at habilitation, promotion or tenure. All models control for age at tenure and are estimated for years t>2004 and for academics who make tenure as of 1999. Standard errors are robust and clustered by individual academic.

Appendix A

Proof of Proposition 1:

For second period effort, we have that $e_2^{p,*}(y_1) > e_2^{a,*}$ because $E_{\theta_i,\theta_j}[b * g(\triangle\theta)|y_1] = C'\left(e_2^{p,*}(y_1)\right) > 0 = C'\left(e_2^{a,*}(y_1)\right)$ for any $g(.)$ with the entire real line as support.

For first period effort, $e_1^{p,*}(y_1) > e_2^{a,*}$ because $C'\left(e_{i,1}^{p,*}\right) = \delta\frac{\partial}{\partial e_{i,1}}E[\theta_i|y_i] > C'\left(e_{i,1}^{a,*}\right) = 0$. To prove the latter inequality, note that the derivative of the conditional expectation of ability with respect to first period effort is equal to

$$\frac{\partial}{\partial e_{i,1}}E[\theta_i|y_i] = \int_\theta \theta\frac{\frac{\partial g(y|\theta)}{\partial e_{i,1}}f(\theta)\int_\theta g(y|\theta)f(\theta)d\theta - g(y|\theta)f(\theta)\int_\theta\frac{\partial g(y|\theta)}{\partial e_{i,1}}f(\theta)d\theta}{\left[\int_\theta\frac{\partial g(y|\theta)}{\partial e_{i,1}}f(\theta)d\theta\right]^2} \tag{25}$$

This derivative is positive if the numerator is positive. Rewrite this numerator as

$$\int_{\hat\theta}\int_\theta\hat\theta\left[\frac{\partial g\left(y|\hat\theta\right)}{\partial e_{i,1}}g(y|\theta) - g\left(y|\hat\theta\right)\frac{\partial g(y|\theta)}{\partial e_{i,1}}\right]f(\theta)df(\hat\theta)d\hat\theta \tag{26}$$

The derivative with respect to $e_{i,1}$ of the distribution of $y_{i,1}$ conditional on θ for normally distributed mean-zero error terms is equal to

$$\frac{\partial g\left(y|\hat\theta\right)}{\partial e_{i,1}} = \frac{-(y_{i,1}-\theta-\hat{e_1})}{\sigma^2\sqrt{2\pi\sigma^2}}\left(e^{\frac{-\left(y_{i,1}-\theta-\hat{e_1}\right)^2}{2\sigma^2}}\right)$$

Substituting this into (26) and simplifying yields

$$\int_{\hat\theta}\int_\theta\hat\theta\left[\frac{1}{\sigma^2\sqrt{2\pi\sigma^2}}\left(e^{\frac{-\left(y_{i,1}-\hat\theta-\hat{e_1}\right)^2-\left(y_{i,1}-\theta-\hat{e_1}\right)^2}{2\sigma^2}}\right)\left[-\left(y_{i,1}-\hat\theta-\hat{e_1}\right)-(y_{i,1}-\theta-\hat{e_1})\right]\right]f(\theta)d\theta f(\hat\theta)d\hat\theta \tag{27}$$

The bracketed term $-\left(y_{i,1}-\hat\theta-\hat{e_1}\right)-(y_{i,1}-\theta-\hat{e_1})$ simplifies to $\left(\hat\theta-\theta\right)$, which is larger than zero for $\hat\theta > \theta$. Denote

$$K\left(\hat\theta,\theta\right) := \frac{1}{\sigma^2\sqrt{2\pi\sigma^2}}\left(e^{\frac{-\left(y_{i,1}-\hat\theta-\hat{e_1}\right)^2-\left(y_{i,1}-\theta-\hat{e_1}\right)^2}{2\sigma^2}}\right)$$

and rewrite (27) as

$$\int_{\hat\theta=\bar\theta}^{\theta_{max}}\int_{\theta=\theta_{min}}^{\bar\theta}\hat\theta K\left(\hat\theta,\theta\right)\left(\hat\theta-\theta\right)f(\theta)d\theta f(\hat\theta)d\hat\theta + \int_{\hat\theta=\theta_{min}}^{\bar\theta}\int_{\theta=\bar\theta}^{\theta_{max}}\hat\theta K\left(\hat\theta,\theta\right)\left(\hat\theta-\theta\right)f(\theta)d\theta f(\hat\theta)d\hat\theta \tag{28}$$

$$\int_{\hat\theta=\theta_{min}}^{\bar\theta}\int_{\theta=\theta_{min}}^{\bar\theta}\hat\theta K\left(\hat\theta,\theta\right)\left(\hat\theta-\theta\right)f(\theta)d\theta f(\hat\theta)d\hat\theta + \int_{\hat\theta=\bar\theta}^{\theta_{max}}\int_{\theta=\bar\theta}^{\theta_{max}}\hat\theta K\left(\hat\theta,\theta\right)\left(\hat\theta-\theta\right)f(\theta)d\theta f(\hat\theta)d\hat\theta$$

where $\bar\theta$ denotes the ability type.

Note that $K\left(\hat\theta,\theta\right)$=$K\left(\theta,\hat\theta\right)$, so that $K\left(\hat\theta,\theta\right)\left(\hat\theta-\theta\right)$=$-K\left(\theta,\hat\theta\right)\left(\theta-\hat\theta\right)$, so that the first two terms reduce to

$$\int_{\hat\theta=\bar\theta}^{\theta_{max}}\int_{\theta=\theta_{min}}^{\bar\theta}\left(\hat\theta-\theta\right)^2 K\left(\hat\theta,\theta\right)f(\theta)d\theta f(\hat\theta)d\hat\theta > 0 \tag{29}$$

Iteration of this same approach of dividing the integration ranges in two for the remaining terms until the integration intervals approach zero shows that (27) is larger than zero. QED

Proof of Proposition 2

I first show that optimal second period effort under performance pay decreases with the difference in ability of competitors. Implicitly differentiating the first-order condition for optimal second period effort (7) with respect to the difference in ability types $\triangle\theta$ yields

$$\frac{\partial e_2^*}{\partial\triangle\theta} = \frac{-E_{\theta_i,\theta_j}\left[b*g'\left(\triangle\theta\right)|y_1\right]}{E_{\theta_i,\theta_j}\left[b*g'\left(\triangle\theta\right)|y_1\right]-C''\left(e_2^*\right)} \tag{30}$$

The denominator is the second-order derivative for optimal second period effort, which is assumed to be negative. Because $g\left(.\right)$ is the probability density function of the difference of two iid random variables that are each distributed according to a unimodal, mean-zero distribution, $g'\left(x\right)<0$ if $x>0$ while $g'\left(x\right)>0$ if $x<0$. Hence $\frac{\partial e_2^*}{\partial\triangle\theta}$<0 if $\triangle\theta=\theta_i-\theta_j>0$, which is the case for the competitor with the higher ability. By symmetry, $\frac{\partial e_{i,2}^*}{\partial\triangle\theta}=\frac{\partial e_2^*(\triangle\theta)}{\partial\triangle\theta}=-\frac{\partial e_2^*(-\triangle\theta)}{\partial\triangle\theta}=\frac{\partial e_{j,2}^*}{\partial\triangle\theta}$, so that $\frac{\partial e_2^*}{\partial\triangle\theta}$>0 for the competitor with the lower ability as well.

Next, I show that the unique pairwise stable matching of competitors is such that the lowest ability academic sorts into a tournament where they compete against the next lowest ability academic and so on. First, observe that second period utility increases in ability difference:

$$\frac{\partial E\left[U_{i,2}\right]}{\partial\triangle\theta}=E_{\theta_i,\theta_j}\left[b*g\left(\triangle\theta+\triangle e_2^*\right)|y_1\right]+\left(E_{\theta_i,\theta_j}\left[b*g\left(\triangle\theta+\triangle e_2^*\right)|y_1\right]-C'\left(e_2^*\right)\right)\frac{\partial e_2^*}{\partial\triangle\theta}=E_{\theta_i,\theta_j}\left[b*g\left(\triangle\theta\right)|y_1\right]>0 \tag{31}$$

where the second equality follows from the envelope theorem, and the inequality at the end from the assumption that the probability distribution function of the noise terms has full support. Because $g\left(x\right)$ is symmetric around 0, we then have that $dU_{i,2}\approx E_{\theta_i,\theta_j}\left[b*g\left(\triangle\theta\right)|y_1\right]*d\triangle\theta>0$ for $d\triangle\theta>0$ while $dU_{i,2}\approx E_{\theta_i,\theta_j}\left[b*g\left(\triangle\theta\right)|y_1\right]*d\triangle\theta<0$ for $d\triangle\theta<0$. Every academic therefore wants to compete with an academic of as low ability as possible, so that $\triangle\theta$ is as large positive as possible. For the academic with the lowest ability however, this means that they should want to compete against the academic of the next lowest ability, and so on. That is, the unique pairwise stable matching is positively assortative. The proof is by contradiction.

Suppose that there are four academics, with ability types $\theta_1<\theta_2<\theta_3<\theta_4$. With slight abuse of notation, I will refer to academics by their ability type in what follows. Suppose that the following matching $\hat{m}$ is a pairwise stable matching: θ_1 is matched to θ_3 and θ_2 to θ_4. For this to be a pairwise stable matching, there should be no two academics who would rather be matched with each other than with their match partners in $\hat{m}$. But θ_1 and θ_2 can increase their utility by matching with each other, since $d\triangle\theta=\left(\theta_1-\theta_2\right)-\left(\theta_1-\theta_3\right)>0$ and $d\triangle\theta=\left(\theta_2-\theta_1\right)-\left(\theta_2-\theta_4\right)>0$. Hence $\hat{m}$ cannot be pairwise stable. By iteration, the same can be shown for any matching other than the positively assortative matching m^*in which θ_1 is matched to θ_2 and θ_3 to θ_4.

For a finite number of academics, whose abilities are drawn from a right-skewed distribution $f(\theta)$ of ability types, the expected difference in ability between two academics at the lower end of the ability distribution is smaller than the difference in ability between two academics at the higher end of the ability distribution because the median ability is smaller than the mean ability. Combined with the fact that optimal second period effort decreases in the ability difference between two competitors, it must be that the optimal effort of tenured academics is smaller for higher ability academics. QED

Proof of Proposition 3

The first part of the proposition follows immediately from the first-order derivative for optimal second period risk (9). This is negative for $\triangle\theta > 0$ and positive $\triangle\theta < 0$ because $g(.)$ is symmetric around 0. Therefore, optimal second period risk is σ_{min} for the higher ability academic of a pair of competitors, while it is σ_{max} for the lower ability academic.

The second part of the proposition requires deriving the sign of the derivative of the expected ability conditional on first period output with respect to first period risk. This derivative is given by

$$\frac{\partial}{\partial\sigma_{i,1}}E\left[\theta_i|y_i\right]=\int_\theta \theta\frac{\frac{\partial g(y|\theta)}{\partial\sigma_{i,1}}f(\theta)\int_\theta g\left(y|\theta\right)f(\theta)d\theta-g\left(y|\theta\right)f(\theta)\int_\theta\frac{\partial g(y|\theta)}{\partial\sigma_{i,1}}f(\theta)d\theta}{\left[\int_\theta\frac{\partial g(y|\theta)}{\partial\sigma_{i,1}}f(\theta)d\theta\right]^2} \tag{32}$$

This derivative is positive if the numerator is positive. Rewrite this numerator as

$$\int_{\hat{\theta}}\int_\theta \hat{\theta}\left[\frac{\partial g\left(y|\hat{\theta}\right)}{\partial\sigma_{i,1}}g\left(y|\theta\right)-g\left(y|\hat{\theta}\right)\frac{\partial g\left(y|\theta\right)}{\partial\sigma_{i,1}}\right]f(\theta)d\theta f(\hat{\theta})d\hat{\theta} \tag{33}$$

The derivative with respect to $\sigma_{i,1}$ of the distribution of $y_{i,1}$ conditional on θ for normally distributed mean-zero error terms is equal to

$$\frac{1}{\sigma\sqrt{2\pi\sigma^2}}\left(e^{\frac{-\left(y_{i,1}-\theta-\hat{e_1}\right)^2}{2\sigma^2}}\right)\left[\frac{\left(y_{i,1}-\theta-\hat{e_1}\right)^2-1}{\sigma^2}\right]$$

Substituting this into (33) and simplifying yields

$$\int_{\hat{\theta}}\int_\theta \hat{\theta}\left[\frac{1}{\sigma^2 2\pi\sigma^3}\left(e^{\frac{-\left(y_{i,1}-\hat{\theta}-\hat{e_1}\right)^2-\left(y_{i,1}-\theta-\hat{e_1}\right)^2}{2\sigma^2}}\right)\left[\left(y_{i,1}-\hat{\theta}-\hat{e_1}\right)^2-\left(y_{i,1}-\theta-\hat{e_1}\right)^2\right]\right]f(\theta)d\theta f(\hat{\theta})d\hat{\theta} \tag{34}$$

The bracketed term $\left(y_{i,1}-\hat{\theta}-\hat{e_1}\right)^2-\left(y_{i,1}-\theta-\hat{e_1}\right)^2$is smaller than zero for $\hat{\theta}>\theta$. Denote

$$M\left(\hat{\theta},\theta\right):=\frac{1}{\sigma^2 2\pi\sigma^3}\left(e^{\frac{-\left(y_{i,1}-\hat{\theta}-\hat{e_1}\right)^2-\left(y_{i,1}-\theta-\hat{e_1}\right)^2}{2\sigma^2}}\right)$$

and $\triangle D\left(\hat{\theta},\theta\right):=\left(y_{i,1}-\hat{\theta}-\hat{e_1}\right)^2-\left(y_{i,1}-\theta-\hat{e_1}\right)^2$
and rewrite (34) as

$$\int_{\hat{\theta}=\bar{\theta}}^{\theta_{max}}\int_{\theta=\theta_{min}}^{\bar{\theta}}\hat{\theta}M\left(\hat{\theta},\theta\right)\triangle D\left(\hat{\theta},\theta\right)f(\theta)d\theta f(\hat{\theta})d\hat{\theta}+\int_{\hat{\theta}=\theta_{min}}^{\bar{\theta}}\int_{\theta=\bar{\theta}}^{\theta_{max}}\hat{\theta}M\left(\hat{\theta},\theta\right)\triangle D\left(\hat{\theta},\theta\right)f(\theta)d\theta f(\hat{\theta})d\hat{\theta} \tag{35}$$

$$\int_{\hat{\theta}=\theta_{min}}^{\bar{\theta}}\int_{\theta=\theta_{min}}^{\bar{\theta}}\hat{\theta}M\left(\hat{\theta},\theta\right)\triangle D\left(\hat{\theta},\theta\right)f(\theta)d\theta f(\hat{\theta})d\hat{\theta}+\int_{\hat{\theta}=\bar{\theta}}^{\theta_{max}}\int_{\theta=\bar{\theta}}^{\theta_{max}}\hat{\theta}M\left(\hat{\theta},\theta\right)\triangle D\left(\hat{\theta},\theta\right)f(\theta)d\theta f(\hat{\theta})d\hat{\theta}$$

Note that $M\left(\hat{\theta},\theta\right)=M\left(\theta,\hat{\theta}\right)$, so that $M\left(\hat{\theta},\theta\right)\triangle D\left(\hat{\theta},\theta\right)=-M\left(\theta,\hat{\theta}\right)\triangle D\left(\theta,\hat{\theta}\right)$, so that the first two terms reduce to

$$\int_{\hat{\theta}=\bar{\theta}}^{\theta_{max}}\int_{\theta=\theta_{min}}^{\bar{\theta}}\left(\hat{\theta}-\theta\right)M\left(\hat{\theta},\theta\right)\triangle D\left(\hat{\theta},\theta\right)f(\theta)d\theta f(\hat{\theta})d\hat{\theta}<0 \tag{36}$$

Iteration of this same approach of dividing the integration ranges in two for the remaining terms until the integration intervals approach zero shows that (34) is smaller than zero. Optimal first period risk is therefore σ_{min}. QED

Proof of Proposition 4

In period 2, an academic's expected equilibrium utility when working under the age-related pay scheme is given by:

$$E\left[U_{i,2}^{a}\left(\sigma_{i,2}^{a*},e_{i,2}^{a*}\right)|y\right]=w_{2}^{a}-C\left(e_{i,2}^{a*}\right)=w_{2}^{a} \tag{37}$$

while their expected life-time utility when working under the performance pay scheme is given by:

$$E\left[U_{i,2}^{p}\left(\sigma_{i,2}^{p*},e_{i,2}^{p*}\right)|y_{1}\right]=w_{2}^{p}+bG\left(\triangle\theta\right)-C\left(e_{i,2}^{p*}\right)=E\left[\theta_{i}|y_{i}\right]+e_{i,2}^{p*}-C\left(e_{i,2}^{p*}\right) \tag{38}$$

The second equality follows from perfect competition in the labor market. As noted before, competition in the labor market means that total pay in the second period equals total expected output, conditional on first period.

From (16), we know that $e_{i,2}^{p*}>0$. By optimality of $e_{i,2}^{p*}$, it must be that $e_{i,2}^{p*}-C\left(e_{i,2}^{p*}\right)$>0. Suppose not, then academic i could increase their utility by reducing effort, and hence cost of effort, and increase utilility. This would contradict optimality of of $e_{i,2}^{p*}$.

It follows that any academic with expected ability $E\left[\theta_{i}|y_{i}\right]>w_{2}^{a}-\left(e_{i,2}^{p*}-C\left(e_{i,2}^{p*}\right)\right)$ would increase their utility by switching to performance pay. A derivation akin to the one given in the proof of proposition 1 shows that the derivative of the conditional expectation of ability increases with actual ability. More able academics are more likely to select into performance pay. QED

Tab. 8: Publication Variables Restricted to Articles Only

Panel A: Unweighted Publication Variables

	# Publications	IF-rated publications	Citations	Average IF-rating	Average citations
Post'02 * Treatment	0.165***	0.129**	0.130**	-0.092**	-0.110*
	(0.040)	(0.055)	(0.061)	(0.039)	(0.056)
Tenure * Treatment	-0.046	-0.033	-0.026	0.025	-0.033
	(0.041)	(0.053)	(0.063)	(0.040)	(0.060)
Number of Observations	81057	72316	77308	44499	45106
Number of Individuals	4510	4024	4301	3790	3974
Log Likelihood	-120114.906	-278455.246	-4048814.424	-66542.784	-679164.726
Chi-squared	2556.845	3229.310	1158.145	897.142	452.101

Panel A.II: Maximum and Minimum Number of Citations

	Maximum citations	Minimum citations
Post'02 * Treatment	-0.005	-0.240**
	(0.062)	(0.113)
Tenure * Treatment	-0.112	0.035
	(0.075)	(0.130)
Number of Observations	45106	44851
Number of Individuals	3974	3946
Log Likelihood	-1923612.159	-532740.222
Chi-squared	344.780	468.234

Panel B: Publication Variables, Weighted by Number of Authors

	# Publications	IF-rated publications	Citations	Average IF-rating	Average citations
Post'02 * Treatment	0.166***	0.103**	0.077	-0.101**	-0.132**
	(0.038)	(0.048)	(0.060)	(0.041)	(0.059)
Tenure * Treatment	-0.016	0.044	0.075	0.037	-0.001
	(0.043)	(0.051)	(0.062)	(0.042)	(0.062)
Number of Observations	81057	72316	77308	44499	45106
Number of Individuals	4510	4024	4301	3790	3974
Log Likelihood	-54807.809	-81724.728	-856200.546	-64194.652	-620292.140
Chi-squared	1571.130	2060.060	812.743	625.075	684.985

Panel B.II: Maximum and Minimum Number of Citations, Weighted by Number of Authors

	Maximum citations	Minimum citations
Post'02 * Treatment	-0.052	-0.175
	(0.066)	(0.112)
Tenure * Treatment	0.021	-0.111
	(0.072)	(0.110)
Number of Observations	45106	44851
Number of Individuals	3974	3946
Log Likelihood	-422129.834	-147387.134
Chi-squared	699.102	1067.075

Notes: * denotes significance at 10%, ** at 5% and *** at 1%. The unit of observation is academic i. The sample is restricted to academics who started their first tenured affiliation at a German public university in 2002 to 2007 (excluding those with a foreign affiliation directly prior to this) and includes data from 1993 until and including 2012. The dependent variables in Panels A.I and B.I are, respectively, the number of publications, the impact factor-rated number of publications, the total number of citations to all publications published in a given year, the average impact factor rating of publications published in a given year, and the average number of citations to publications published in a given year. The dependent variables in Panels A.II and B.II are, respectively, the maximum and minimum number of citations to publications published in a given year. All dependent variables are defined for academic i in field f and year t, lagged by average publication lag in field f as reported in Björk and Solomon (2013). $Post'02$ is 0 before 2002 and 1 thereafter, $Tenure$ is 0 before 2005 and 1 thereafter, and $Treatment$ is 1 if an academic makes tenure at a public university in 2005, 2006 or 2007 and 0 if they make tenure at a public university in 2002, 2003 or 2004 (the control group). Productivity quantiles are determined on the basis of the averages of the various dependent variables over the three pre-announcement years 1999, 2000 and 2001, separately by academic field and tenure cohort. All specifications control for year and individual fixed effects and fifteen time-to-tenure fixed effects (from seven years before the tenure year to seven years after). Estimation as conditional quasi-maximum likelihood estimation of Poisson fixed effects models with robust standard errors.

Tab. 9: Further Robustness Checks to Alternative Specifications

Panel A: 4-Year Treatment and Control Groups	# Publications	IF-rated publications	Citations	Average IF-rating	Average citations
Post'02 * Treatment	0.138***	0.093**	0.066	-0.064**	-0.094*
	(0.034)	(0.046)	(0.055)	(0.031)	(0.053)
Tenure * Treatment	-0.005	0.027	0.081	0.018	0.032
	(0.035)	(0.046)	(0.056)	(0.031)	(0.059)
Number of Observations	104987	93038	97636	57688	58553
Number of Individuals	5842	5177	5432	4888	5115
Log Likelihood	-167728.232	-416551.362	-5601408.411	-86337.296	-907781.141
Chi-squared	3541.787	3438.503	1344.787	756.156	574.359
Panel B: 2-Year Treatment and Control Groups					
Post'02 * Treatment	0.118*	0.028	0.015	-0.107*	-0.059
	(0.068)	(0.091)	(0.105)	(0.059)	(0.125)
Tenure * Treatment	-0.076	-0.018	0.012	0.058	-0.05
	(0.061)	(0.082)	(0.098)	(0.056)	(0.102)
Number of Observations	46810	41180	43562	25910	26341
Number of Individuals	2604	2291	2423	2175	2287
Log Likelihood	-76523.381	-187507.992	-2535716.145	-38137.036	-401924.46
Chi-squared	1678.502	1748.001	645.885	421.276	339.662
Panel C: Controlling for 8 Years Before to 8 Years after Tenure					
Post'02 * Treatment	0.119***	0.107*	0.104	-0.086**	-0.101
	(0.044)	(0.060)	(0.072)	(0.039)	(0.072)
Tenure * Treatment	-0.007	0.028	0.018	0.025	-0.018
	(0.040)	(0.051)	(0.064)	(0.036)	(0.067)
Number of Observations	83937	74326	78308	47052	47789
Number of Individuals	4671	4136	4357	3917	4110
Log Likelihood	-136629.216	-338216.424	-4507838.674	-70675.656	-736169.531
Chi-squared	2856.093	2861.711	1100.645	619.954	476.975

Tab. 9: Further Robustness Checks to Alternative Specifications

Panel D: Controlling for 6 Years Before to 6 Years after Tenure					
	# Publications	IF-rated publications	Citations	Average IF-rating	Average citations
Post'02 * Treatment	0.168***	0.131***	0.161***	-0.071**	-0.044
	(0.038)	(0.050)	(0.061)	(0.034)	(0.062)
Tenure * Treatment	-0.01	0.03	0.018	0.026	-0.018
	(0.041)	(0.053)	(0.064)	(0.036)	(0.067)
Number of Observations	83937	74326	78308	47052	47789
Number of Individuals	4671	4136	4357	3917	4110
Log Likelihood	-136647.71	-338272.363	-4509293.616	-70684.686	-736633.658
Chi-squared	2833.915	2827.513	1099.907	601.016	469.047
Panel E: With Post'05*interaction					
Post'02 * Treatment	0.158***	0.132***	0.119*	-0.092***	-0.076
	(0.035)	(0.049)	(0.063)	(0.032)	(0.065)
Post'05 * Treatment	0.006	0.028	0.037	0.021	-0.070
	(0.037)	(0.051)	(0.064)	(0.032)	(0.065)
Number of Observations	83937	74326	78308	47052	47789
Number of Individuals	4671	4136	4357	3917	4110
Log Likelihood	-136647.402	-338246.655	-4508120.087	-70677.188	-736134.759
Chi-squared	2860.852	2846.786	1100.804	608.041	478.565
Panel F: Inverse Hyperbolic Sine Transform Specification					
Post'02 * Treatment	0.118***	0.075***	0.092***	-0.040***	-0.012
	(0.024)	(0.025)	(0.025)	(0.015)	(0.015)
Tenure * Treatment	0.004	0.02	0.023	0.004	0.011
	(0.023)	(0.024)	(0.024)	(0.016)	(0.016)
Number of Observations	83937	74326	78308	47052	47789
Number of Individuals	4671	4136	4357	3917	4110
Log Likelihood	-75124.742	-80201.805	-125433.185	-46584.775	-71901.016
Chi-squared	2927.707	2987.478	2115.081	1454.155	540.056

Notes: * denotes significance at 10%, ** at 5% and *** at 1%. The unit of observation is academic i. The sample is restricted to academics who started their first tenured affiliation at a German public university in 2002 to 2007 (excluding those with a foreign affiliation directly prior to this) and includes data from 1993 until and including 2012. The dependent variables are, respectively, the number of publications, the impact factor-rated number of publications, the total number of citations to all publications published in a given year, the average impact factor rating of publications published in a given year, and the average number of citations to publications published in a given year. All dependent variables are defined for academic i in field f and year t, lagged by average publication lag in field f as reported in Björk and Solomon (2013). $Post'02$ is 0 before 2002 and 1 thereafter, *Tenure* is 0 before 2005 and 1 thereafter, and $Treatment$ is 1 if an academic makes tenure at a public university in 2005, 2006 or 2007 and 0 if they make tenure at a public university in 2002, 2003 or 2004 (the control group). Productivity quantiles are determined on the basis of the averages of the various dependent variables over the three pre-announcement years 1999, 2000 and 2001, separately by academic field and tenure cohort. All specifications control for year and individual fixed effects and fifteen time-to-tenure fixed effects (from seven years before the tenure year to seven years after). Estimation as conditional quasi-maximum likelihood estimation of Poisson fixed effects models with robust standard errors.

Tab. 10: Heterogeneous Treatment Effects - Robustness Checks

Panel A: Control without switchers	# Publications	IF-rated publications	Citations	Average IF-rating	Average citations
	Coef./SE	Coef./SE	Coef./SE	Coef./SE	Coef./SE
Post'02 * Q4	-0.299***	-0.242***	-0.317***	-0.262***	-0.479***
	(0.046)	(0.064)	(0.064)	(0.030)	(0.065)
Post'02 * Treatment	0.214***	0.267***	0.193**	-0.086**	-0.106
	(0.051)	(0.075)	(0.086)	(0.044)	(0.074)
Post'02 * Treatment * Q4	-0.007	-0.129	-0.052	-0.003	-0.002
	(0.061)	(0.083)	(0.091)	(0.045)	(0.092)
Tenure * Q4	0.117***	0.056	0.051	0.022	0.093
	(0.027)	(0.039)	(0.065)	(0.033)	(0.068)
Tenure * Treatment	0.125***	0.163**	0.133	0.055	0.053
	(0.048)	(0.066)	(0.083)	(0.044)	(0.080)
Tenure * Treatment * Q4	-0.201***	-0.224***	-0.183***	-0.097***	-0.217***
	(0.035)	(0.047)	(0.061)	(0.030)	(0.060)
Number of Observations	70414	62260	65702	37361	37802
Number of Individuals	3919	3465	3656	2794	2892
Log Likelihood	-113329.049	-281376.858	-3775756.779	-57492.391	-591618.356
Chi-squared	2818.059	2613.658	1119.943	816.683	529.397

Tab. 10: Heterogeneous Treatment Effects - Robustness Checks

Panel B: Heterogeneous responses by quartile	# Publications	IF-rated publications	Citations	Average IF-rating	Average citations
Post'02 * Q2	-0.575***	-1.061***	-0.984***	-0.231***	-0.361***
	(0.108)	(0.161)	(0.156)	(0.046)	(0.070)
Post'02 * Q3	-0.782***	-1.200***	-1.188***	-0.303***	-0.493***
	(0.104)	(0.150)	(0.140)	(0.045)	(0.067)
Post'02 * Q4	-0.946***	-1.335***	-1.410***	-0.503***	-0.867***
	(0.099)	(0.145)	(0.137)	(0.043)	(0.075)
Post'02 * Treatment	-0.120	-0.207	-0.310	-0.134*	0.040
	(0.133)	(0.229)	(0.199)	(0.071)	(0.126)
Post'02 * Treatment * Q2	0.284*	0.386	0.626***	0.044	-0.160
	(0.149)	(0.251)	(0.237)	(0.074)	(0.138)
Post'02 * Treatment * Q3	0.326**	0.461*	0.400*	0.054	-0.196
	(0.142)	(0.238)	(0.207)	(0.071)	(0.129)
Post'02 * Treatment * Q4	0.281**	0.296	0.425**	0.051	-0.129
	(0.137)	(0.233)	(0.203)	(0.071)	(0.137)
Tenure * Q2	0.162**	0.206*	0.112	0.167***	-0.109
	(0.068)	(0.113)	(0.183)	(0.046)	(0.117)
Tenure * Q3	0.246***	0.273**	0.126	0.141***	-0.130
	(0.064)	(0.108)	(0.169)	(0.041)	(0.113)
Tenure * Q4	0.311***	0.297***	0.184	0.162***	-0.001
	(0.062)	(0.105)	(0.169)	(0.043)	(0.116)
Tenure * Treatment	0.475***	0.545***	0.503***	0.206***	0.043
	(0.092)	(0.159)	(0.162)	(0.063)	(0.108)
Tenure * Treatment * Q2	-0.383***	-0.320**	-0.379**	-0.148**	-0.036
	(0.092)	(0.163)	(0.171)	(0.059)	(0.103)
Tenure * Treatment * Q3	-0.437***	-0.455***	-0.421***	-0.192***	0.036
	(0.089)	(0.155)	(0.161)	(0.057)	(0.096)
Tenure * Treatment * Q4	-0.561***	-0.591***	-0.550***	-0.239***	-0.209**
	(0.086)	(0.152)	(0.153)	(0.057)	(0.098)
Number of Observations	80953	71540	75504	42848	43348
Number of Individuals	4505	3981	4201	3213	3328
Log Likelihood	-130449.894	-320080.403	-4271644.740	-65244.459	-660980.749
Chi-squared	3381.317	3321.157	1563.662	1128.193	715.831

Notes: * denotes significance at 10%, ** at 5% and *** at 1%. The unit of observation is academic i. The sample is restricted to academics who started their first tenured affiliation at a German public university in 2002 to 2007 (excluding those with a foreign affiliation directly prior to this) and includes data from 1993 until and including 2012. The dependent variables are, respectively, the number of publications, the impact factor-rated number of publications, the total number of citations to all publications published in a given year, the average impact factor rating of publications published in a given year, and the average number of citations to publications published in a given year. All dependent variables are defined for academic i in field f and year t, lagged by average publication lag in field f as reported in Björk and Solomon (2013). $Post'02$ is 0 before 2002 and 1 thereafter, *Tenure* is 0 before 2005 and 1 thereafter, and *Treatment* is 1 if an academic makes tenure at a public university in 2005, 2006 or 2007 and 0 if they make tenure at a public university in 2002, 2003 or 2004 (the control group). Productivity quantiles are determined on the basis of the averages of the various dependent variables over the three pre-announcement years 1999, 2000 and 2001, separately by academic field and tenure cohort. All specifications control for year and individual fixed effects and fifteen time-to-tenure fixed effects (from seven years before the tenure year to seven years after). Estimation as conditional quasi-maximum likelihood estimation of Poisson fixed effects models with robust standard errors.

Tab. 11: Treatment Effects by Academic Field

Panel A: Science	# Publications	IF-rated publications	Citations	Average IF-rating	Average citations
Post'02 * Treatment	0.102*	0.049	0.108	-0.101*	-0.069
	(0.058)	(0.073)	(0.087)	(0.056)	(0.092)
Tenure * Treatment	0.028	0.074	0.029	0.045	-0.083
	(0.061)	(0.072)	(0.095)	(0.057)	(0.098)
Number of Observations	31167	30627	30767	20998	21012
Number of Individuals	1562	1535	1542	1492	1496
Log Likelihood	-60148.605	-167256.273	-2451073.01	-34541.278	-378025.729
Chi-squared	261968.506	254035.508	330751.11	482.564	350.497
Panel B: Applied Science					
Post'02 * Treatment	0.200***	0.183**	0.127	-0.056	-0.083
	(0.053)	(0.072)	(0.091)	(0.042)	(0.086)
Tenure * Treatment	-0.015	0.055	0.057	0.002	0.033
	(0.059)	(0.079)	(0.093)	(0.044)	(0.099)
Number of Observations	33198	32671	32530	22086	22084
Number of Individuals	1666	1639	1632	1589	1587
Log Likelihood	-68048.835	-184412.454	-2002109.887	-34405.783	-297410.926
Chi-squared	309901.975	225418.21	297284.653	299.824	359.567
Panel C: Social Science					
Post'02 * Treatment	0.175	-0.23	-0.238	-0.345*	-0.311
	(0.121)	(0.223)	(0.248)	(0.208)	(0.224)
Tenure * Treatment	0.078	0.145	-0.018	0.136	0.053
	(0.135)	(0.221)	(0.261)	(0.250)	(0.218)
Number of Observations	14712	13752	13318	4467	4469
Number of Individuals	737	689	667	589	589
Log Likelihood	-11160.792	-12843.436	-232556.65	-4085.868	-55668.751
Chi-squared	111242.847	76001.014	96073.857	263.46	208.252
Panel D: Humanities					
Post'02 * Treatment	0.554**	-0.848	0.976	-0.332	1.034*
	(0.258)	(0.595)	(0.600)	(0.601)	(0.575)
Tenure * Treatment	-0.262	0.694	-1.992*	0.994	-0.911
	(0.281)	(1.085)	(1.136)	(0.780)	(0.844)
Number of Observations	11332	3015	7713	766	1482
Number of Individuals	568	151	386	132	320
Log Likelihood	-4854.905	-1156.475	-25256.859	-360.749	-6237.449
Chi-squared	67166.629	17509.513	25678.314	240.636	253.874

Notes: * denotes significance at 10%, ** at 5% and *** at 1%. The unit of observation is academic i. The sample is restricted to academics who started their first tenured affiliation at a German public university in 2002 to 2007 (excluding those with a foreign affiliation directly prior to this) and includes data from 1993 until and including 2012. The dependent variables in Panels A.I and B.I are, respectively, the number of publications, the impact factor-rated number of publications, the total number of citations to all publications published in a given year, the average impact factor rating of publications published in a given year, and the average number of citations to publications published in a given year. The dependent variables in Panels A.II and B.II are, respectively, the maximum and minimum number of citations to publications published in a given year. All dependent variables are defined for academic i in field f and year t, lagged by average publication lag in field f as reported in Björk and Solomon (2013). $Post'02$ is 0 before 2002 and 1 thereafter, $Tenure$ is 0 before 2005 and 1 thereafter, and $Treatment$ is 1 if an academic makes tenure at a public university in 2005, 2006 or 2007 and 0 if they make tenure at a public university in 2002, 2003 or 2004 (the control group). Productivity quantiles are determined on the basis of the averages of the various dependent variables over the three pre-announcement years 1999, 2000 and 2001, separately by academic field and tenure cohort. All specifications control for year and individual fixed effects and fifteen time-to-tenure fixed effects (from seven years before the tenure year to seven years after). Estimation as conditional quasi-maximum likelihood estimation of Poisson fixed effects models with robust standard errors.

Tab. 12: Novelty Effects

Panel A: Baseline Novelty Effects

	Backward similarity			Forward similarity		
Panel A.1: Comparison with papers in same field	Minimum	Maximum	Average	Minimum	Maximum	Average
Post'02 * Treatment	-0.004	0.023	0.056	-0.051	0.087*	0.022
	(0.098)	(0.047)	(0.054)	(0.110)	(0.047)	(0.049)
Tenure * Treatment	-0.054	-0.026	-0.038	-0.207*	-0.044	-0.034
	(0.120)	(0.045)	(0.068)	(0.120)	(0.060)	(0.043)
Number of Observations	39160	39160	39160	37883	37883	37883
Number of Individuals	3774	3774	3774	3686	3686	3686
Log Likelihood	-0.283	-259.413	-15372.354	-0.28	-247.593	-15216.338
Chi-squared	1244.328	152.801	2025.993	1646.646	110.726	1686.01

Panel A.2: Comparison within author						
Post'02 * Treatment	-0.058	0.074	0.033	-0.292	0.112	-0.008
	(0.244)	(0.071)	(0.084)	(0.267)	(0.072)	(0.099)
Tenure * Treatment	-0.378	-0.038	-0.032	0.068	-0.057	-0.152
	(0.263)	(0.069)	(0.097)	(0.234)	(0.069)	(0.110)
Number of Observations	32672	32672	32672	32761	32761	32761
Number of Individuals	3204	3204	3204	3186	3186	3186
Log Likelihood	-6.402	-156.136	-129.77	-6.03	-153.997	-144.893
Chi-squared	324.209	87.361	148.722	386.493	95.226	52.904

Tab. 12: Novelty Effects

Panel B: Heterogeneous Novelty Effects - Top Quartile vs. Bottom Three Quartiles

	Backward similarity			Forward similarity		
Panel B.1: Comparison with papers in same field	Minimum	Maximum	Average	Minimum	Maximum	Average
Post'02 * Q4	0.081	0.005	-0.087***	-0.026	0.009	-0.016
	(0.072)	(0.034)	(0.030)	(0.082)	(0.043)	(0.029)
Post'02 * Treatment	0.001	0.003	0.069	-0.076	0.066	0.03
	(0.111)	(0.054)	(0.064)	(0.125)	(0.055)	(0.056)
Post'02 * Treatment * Q4	-0.087	0.037	-0.02	0.003	0.049	-0.001
	(0.104)	(0.053)	(0.051)	(0.115)	(0.058)	(0.048)
Tenure * Q4	-0.185**	0.114***	0.005	-0.253**	0.086*	0.02
	(0.093)	(0.044)	(0.058)	(0.104)	(0.050)	(0.036)
Tenure * Treatment	-0.079	0.056	-0.065	-0.242*	-0.001	-0.043
	(0.127)	(0.052)	(0.078)	(0.129)	(0.072)	(0.048)
Tenure * Treatment * Q4	0.159*	-0.170***	0.071	0.170*	-0.108	0.035
	(0.084)	(0.055)	(0.046)	(0.096)	(0.069)	(0.037)
Number of Observations	37966	37966	37966	36740	36740	36740
Number of Individuals	3639	3639	3639	3553	3553	3553
Log Likelihood	-0.272	-253.152	-14939.986	-0.27	-241.642	-14775.204
Chi-squared	1228.209	170.175	2050.172	1635.785	110.117	1676.254

Panel B.2: Comparison within author						
Post'02 * Q4	0.299	0.064	0.146*	-0.033	0.04	0.152*
	(0.224)	(0.055)	(0.079)	(0.183)	(0.050)	(0.086)
Post'02 * Treatment	-0.032	0.049	-0.044	-0.388	0.123	0.087
	(0.256)	(0.087)	(0.093)	(0.273)	(0.082)	(0.109)
Post'02 * Treatment * Q4	-0.323	0.032	0.089	0.376	-0.015	-0.098
	(0.348)	(0.085)	(0.120)	(0.257)	(0.077)	(0.123)
Tenure * Q4	0.157	0.121**	0.079	0.083	0.110*	-0.035
	(0.299)	(0.055)	(0.061)	(0.323)	(0.057)	(0.062)
Tenure * Treatment	-0.41	0.077	0.131	0.026	0.002	-0.157
	(0.267)	(0.076)	(0.107)	(0.279)	(0.076)	(0.106)
Tenure * Treatment * Q4	0.158	-0.216***	-0.211*	0.225	-0.115*	0.025
	(0.251)	(0.063)	(0.121)	(0.532)	(0.067)	(0.110)
Number of Observations	31819	31819	31819	31924	31924	31924
Number of Individuals	3101	3101	3101	3082	3082	3082
Log Likelihood	-6.18	-152.691	-126.059	-5.818	-150.753	-140.596
Chi-squared	325.292	97.068	161.992	398.339	96.057	70.231

Notes: * denotes significance at 10%, ** at 5% and *** at 1%. The unit of observation is academic i. The sample is restricted to academics who started their first tenured affiliation at a German public university in 2002 to 2007 (excluding those with a foreign affiliation directly prior to this) and includes data from 1993 until and including 2012. The dependent variables in Panels A.I and A.II are the minimum, maximum and average backward and forward similarity of the publications of author i in year t compared to publications in the same field, while in Panels B.I and B.II the comparison is between papers by the same author. All dependent variables are defined for academic i in field f and year t, lagged by average publication lag in field f as reported in Björk and Solomon (2013). $Post'02$ is 0 before 2002 and 1 thereafter, $Tenure$ is 0 before 2005 and 1 thereafter, and $Treatment$ is 1 if an academic makes tenure at a public university in 2005, 2006 or 2007 and 0 if they make tenure at a public university in 2002, 2003 or 2004 (the control group). Productivity quantiles are determined on the basis of the average of the impact factor-rated number of publications calculated over the three pre-announcement years 1999, 2000 and 2001, separately by academic field and tenure cohort. All specifications control for year and individual fixed effects and fifteen time-to-tenure fixed effects (from seven years before the tenure year to seven years after). Estimation as conditional quasi-maximum likelihood estimation of Poisson fixed effects models with robust standard errors.

Tab. 13: Selection Analysis - Robustness

	Treatment vs Control		Placebo	
Panel A: Cox Proportional Hazard Model	1a	1b	2a	2b
Treatment	-0.185***	1.301***	-0.037	-0.581
	(0.049)	(0.410)	(0.069)	(0.616)
Age	-0.104***	-0.087***	-0.126***	-0.132***
	(0.005)	(0.007)	(0.008)	(0.010)
Avg Productivity	0.002***	-0.001	0.002***	0.003**
	(0.001)	(0.002)	(0.001)	(0.001)
Treatment * Age		-0.034***		0.012
		(0.009)		(0.013)
Treatment * Avg Productivity		0.004**		-0.001
		(0.002)		(0.002)
Number of Observations	80131	80131	51431	51431
Number of Subjects	14972	14972	6960	6960
Number of Switches	2435	2435	1099	1099
Log Likelihood	-21444.778	-21435.296	-8946.454	-8945.883
Chi-squared	677.668	679.827	394.490	402.815
Panel B: With Avg Productivity variable based on impact factor-rated number of publications, weighted by number of authors				
Treatment	-0.564***	0.641	-0.186***	-0.893
	(0.047)	(0.401)	(0.067)	(0.596)
Age	-0.142***	-0.126***	-0.155***	-0.163***
	(0.005)	(0.007)	(0.007)	(0.010)
Avg Productivity	0.009**	-0.008	0.012**	0.018**
	(0.004)	(0.009)	(0.005)	(0.007)
Treatment * Age		-0.027***		0.015
		(0.009)		(0.012)
Treatment * Avg Productivity		0.023**		-0.011
		(0.009)		(0.010)
Constant	3.057***	2.382***	3.219***	3.569***
	(0.242)	(0.330)	(0.335)	(0.428)
Number of Observations	80131	80131	51431	51431
Number of Subjects	14972	14972	6960	6960
Number of Switches	2435	2435	1099	1099
Log Likelihood	-7553.857	-7545.288	-3234.650	-3233.113
Chi-squared	1419.511	1363.612	574.554	599.063

Tab. 13: Selection Analysis - Robustness

	Treatment vs Control		Placebo	
Panel C: With Field Strata	1a	1b	2a	2b
Treatment	-0.569***	0.758*	-0.195***	-0.837
	(0.047)	(0.402)	(0.067)	(0.593)
Age	-0.142***	-0.125***	-0.156***	-0.163***
	(0.005)	(0.007)	(0.007)	(0.010)
Avg Productivity	0.002**	-0.001	0.002**	0.003**
	(0.001)	(0.002)	(0.001)	(0.001)
Age * Treatment		-0.030***		0.014
		(0.009)		(0.012)
Avg Productivity * Treatment		0.003*		-0.001
		(0.002)		(0.002)
Constant	2.870***	2.115***	3.184***	3.515***
Ln_p constant	(0.320)	(0.396)	(0.522)	(0.587)
	0.371***	0.382***	0.379***	0.383***
	(0.080)	(0.082)	(0.127)	(0.126)
Number of Observations	80131	80131	51431	51431
Number of Subjects	14972	14972	6960	6960
Number of Switches	2435	2435	1099	1099
Log Likelihood	-7545.685	-7537.590	-3228.694	-3227.706
Chi-squared	1278.907	1223.781	494.686	516.367

Notes: * denotes significance at 10%, ** at 5% and *** at 1%. The unit of observation is academic *i*. The table reports estimation results of Cox (Panel A) and Weibull propotional hazard models of selection into performance pay. Any first affiliation, position or contract change after implementation of the pay reform (as of 2005) is considered a switch to performance pay (the "failure" event) and the time until such a switch is used as duration variable. Academics are "at risk" of switching after their most recent affiliation/position/contract change if they are tenured, at a public university and not retired. The treatment variable is 1 for academics who have made tenure before 2005 and 0 for those who make tenure afterwards. "Avg Productivity" is calculated as three year pre-implementation averages (2002-2004) of the impact factor-rated number of publications, weighted by number of authors in Panel B only.. The age variable is equal to an author's self-reported age if known, and equal to a synthetic age otherwise. The synthetic age is calculated using the average age at habilitation, promotion or tenure. All models control for age at tenure and are estimated for years t>2004 and for academics who make tenure as of 1999. Standard errors are robust and clustered by individual academic.

Tab. 14: Switching Analysis

	Weibull Model		Cox PH Model	
Panel A: Productivity variables based on impact factor-rated number of publications	1a	1b	2a	2b
Age	-0.160***	-0.151***	-0.123***	-0.115***
	(0.006)	(0.008)	(0.006)	(0.008)
Avg Productivity	0.002**	0.000	0.002***	0.001
	(0.001)	(0.001)	(0.001)	(0.001)
Post		-0.022		0.628
		(0.411)		(0.440)
Post * Age		-0.012		-0.015
		(0.009)		(0.010)
Post * Avg Productivity		0.003*		0.003*
		(0.002)		(0.001)
Constant	3.597***	3.262***		
	(0.286)	(0.378)		
Number of Observations	65639	65639	65639	65639
Number of Subjects	7248	7248	7248	7248
Number of Switches	1599	1599	1599	1599
Log Likelihood	-5122.780	-5089.336	-13575.028	-13572.396
Chi-squared	872.500	952.756	638.909	655.997
Panel B: Productivity variables based on impact factor-rated number of publications weighted by number of authors				
Age	-0.160***	-0.151***	-0.122***	-0.115***
	(0.006)	(0.008)	(0.006)	(0.008)
Avg Productivity	0.011**	-0.000	0.012***	0.003
	(0.005)	(0.008)	(0.004)	(0.007)
Post		-0.041		0.605
		(0.411)		(0.440)
Post * Age		-0.012		-0.014
		(0.009)		(0.010)
Post * Avg Productivity		0.017**		0.016**
		(0.009)		(0.008)
Constant	3.590***	3.267***		
	(0.286)	(0.378)		
Number of Observations	65639	65639	65639	65639
Number of Subjects	7248	7248	7248	7248
Number of Switches	1599	1599	1599	1599
Log Likelihood	-5123.314	-5089.621	-13575.347	-13572.614
Chi-squared	874.000	951.401	639.645	655.795

Notes: * denotes significance at 10%, ** at 5% and *** at 1%. The unit of observation is academic *i*. The table reports estimation results of Weibull propotional hazard models of selection into performance pay. Any first affiliation, position or contract change after implementation of the pay reform (as of 2005) is considered a switch to performance pay (the "failure" event) and the time until such a switch is used as duration variable. Academics are "at risk" of switching after their most recent affiliation/position/contract change if they are tenured, at a public university and not retired. The sample used here is restricted to to academics who have made tenure before 2005. "Avg Productivity" is calculated as three year pre-implementation averages (2002-2004) of the unweighted impact factor-rated number of publications. "Post" is 1 as of 2005. The age variable is equal to an author's self-reported age if known, and equal to a synthetic age otherwise. The synthetic age is calculated using the average age at habilitation, promotion or tenure. All models control for age at tenure and are estimated for years t>2004 and for academics who make tenure as of 1999. Standard errors are robust and clustered by individual academic.