CESifo WORKING PAPERS

10153
2022

December 2022

Effort and Selection Effects of Performance Pay in Knowledge Creation

Erina Ytsma

CESifo

Impressum:

CESifo Working Papers
ISSN 2364-1428 (electronic version)
Publisher and distributor: Munich Society for the Promotion of Economic Research - CESifo GmbH
The international platform of Ludwigs-Maximilians University's Center for Economic Studies and the ifo Institute
Poschingerstr. 5, 81679 Munich, Germany
Telephone +49 (0)89 2180-2740, Telefax +49 (0)89 2180-17845, email office@cesifo.de
Editor: Clemens Fuest
https://www.cesifo.org/en/wp
An electronic version of the paper may be downloaded
· from the SSRN website: www.SSRN.com
· from the RePEc website: www.RePEc.org
· from the CESifo website: https://www.cesifo.org/en/wp

CESifo *Working Paper No. 10153*

Effort and Selection Effects of Performance Pay in Knowledge Creation

Abstract

The effects of performance pay in routine, easy to measure tasks are well-documented, but they are much less understood in knowledge creation. This paper studies the effects of explicit and implicit, career concerns incentives common in knowledge work in a multitasking model, and estimates their causal effort and selection effects in knowledge creation by exploiting the introduction of performance pay in German academia as a natural experiment. Using data encompassing the universe of German academics, I find that performance incentives attract more productive academics, and research quantity increases by 14 to 18%, but without increasing output of the highest quality. The latter is explained by response heterogeneity. The quantity effort response is strongest for low productivity academics, who do not produce high quality work. High ability academics also produce more publications, but not more of the highest quality. Medium ability academics do not increase quantity but produce fewer high-quality papers.

JEL-Codes: J330, M520, O310.

Keywords: performance pay, knowledge creation, career concerns, effort and selection effects, multitasking.

Erina Ytsma
Tepper School of Business
Carnegie Mellon University
Pittsburgh / PA / USA
eytsma@andrew.cmu.edu

1 INTRODUCTION

Knowledge work is an important pillar of present-day economies. It has become rapidly more prevalent over the last four decades and has exhibited consistent growth in occupational employment share[1] (Autor, 2019). Furthermore, knowledge creation has long been considered an important driver of economic growth (Romer, 1986; Lucas, 1988). Yet much is still unclear about how to motivate knowledge workers, including how they respond to performance pay. This paper sheds light on the effect of performance pay on knowledge creation by causally identifying the effort and selection effects of performance incentives in academia.

It is by now well-understood that performance pay increases productivity in routine tasks and settings in which output is readily measurable (e.g. car window replacement, fruit picking, students' test scores), through increases in effort or by attracting the most productive individuals (Lazear, 2000; Shearer, 2004; Bandiera, Barankay and Rasul, 2005; Dohmen and Falk, 2011; Leuven et al., 2011). However, it is not clear that performance pay would have the same effects in the context of knowledge work. For one, knowledge work comprises multiple, complex tasks, the output of which is often not measurable or only a noisy signal of effort. Multitasking problems are therefore likely to arise (Holmström and Milgrom, 1991; Hellmann and Thiele, 2011). Secondly, knowledge production is a function of effort as well as uncertain agent ability. In a competitive labor market with public output signals, workers then face implicit, career concerns incentives on top of any explicit incentives (Gibbons and Murphy, 1992; Holmström, 1999; Lerner and Wulf, 2007; Bonatti and Hörner, 2017). These incentives may well have a different effort and selection effect. Finally, knowledge workers may be particularly highly intrinsically motivated. Higher-powered extrinsic incentives could crowd-out this intrinsic motivation, thus potentially reducing knowledge output (Benabou and Tirole (2003; 2006), Besley and Ghatak (2005; 2018)).

In this paper, I study the effect of performance incentives on the quantity and quality of knowledge output and the productivity of knowledge workers attracted by high-powered incentives. I present a multitasking model with explicit and implicit, career concerns incentives commonly found in knowledge work industries and use this to derive predictions for both average incentive effects and heterogeneous responses across ability types. I test the model's predictions by exploiting the introduction of performance pay in German academia as a natural experiment, using a newly constructed data set of the universe of German academics. The specifics of the rollout of the performance pay scheme give rise to a differential incidence of performance incentives across tenure and age cohorts. This allows me to causally and separately identify the effort and selection effects in a difference-in-differences framework.

The theoretical model presented in this paper has three key features characteristic of knowledge work. First, effort and output have two dimensions, quantity and quality, and output quality is less precisely measured than quantity. The combination of complex tasks with differentially noisy output measures gives rise to multitasking problems (Holmström and Milgrom, 1991). Second, output is a function of both effort and uncertain ability. The market gradually learns about agent ability from the noisy output measures. In a competitive labor market, this learning process gives rise to implicit, market-based incentives; career con-

[1] Knowledge work is defined here as nonroutine cognitive jobs that comprise a host of intellectual tasks.

cerns (Fama, 1980; Holmström, 1999). Thirdly, agents are risk averse and principals can offer contingent contracts, which introduce explicit performance incentives (Gibbons and Murphy, 1992)[2].

The model predicts that, relative to a flat wage, output quantity increases in response to performance pay, but there is potential crowding out of quality effort - more so if risk aversion is high and the quality measure noisy. These responses are not uniform across ability types, but monotonically increasing, decreasing, U- or J-shaped in ability, depending on noise in the output measures and risk aversion. Furthermore, while selection is positive when risk aversion is low, it can be positive, negative or neither when risk aversion is high. Finally, I show that, relative to performance pay systems with explicit incentives only, the presence of career concerns incentives decreases the likelihood of positive quality effort and selection effects, when output signals are differentially noisy. Whether performance pay increases output quantity *and* quality, and attracts more productive workers in knowledge creation therefore requires empirical analysis.

For the empirical analysis, I constructed a data set comprising the affiliations, publications and related information of the universe of academics in Germany by consolidating information from various, unstructured data sources. The performance pay reform that I exploit introduced a performance pay scheme that comprises both implicit, career concerns incentives and explicit performance incentives. These incentives take effect at different points in (relative) time for different tenure cohorts, which allows me to identify their effort effect separately.

To estimate effort effects, I use the fact that any professorial appointment after implementation of the reform falls under the new performance pay scheme, while any existing contract continues to fall under the old age-related pay scheme. Thus, academics who start their first tenured affiliation after the reform are paid according to the performance pay scheme, but academics who are appointed to their first tenured affiliation before the reform start out in the age-related pay scheme. If the timing of the start of the first tenured affiliation is exogenous, any differential productivity change from before to after the reform between academics who start their first tenured position just before and just after the reform can be interpreted as the causal effort effect of performance pay.

I find that performance pay increases research quantity and quality-adjusted quantity by at least 14 to 18% on average. The average quality of publications however decreases by 9 to 10%. These effort effects are highly persistent, and arise in response to the career concerns component of performance pay, with no additional response to explicit performance incentives. The response in output quantity is equivalent to treated academics publishing almost one extra paper every three years, while the decline in average quality is equivalent to a decrease of almost 0.22 in the impact factor of the journal in which the average publication appears (this is roughly equal to the difference between the Economic Journal and RAND Journal in 2005 (Clarivate Analytics, 2000-2012*b*)) or 3.6 fewer citations to publications. Analyzing the distribution of citations and impact factors, I find that treated academics produce more of low to medium quality work, but not more of the highest quality research.

I find no evidence of pre-existing trends, and show that responses align closely with the time when

[2] Market-based wages determined in contract negotiations (career concerns) and on-the-job performance bonuses are common performance incentives in academia and knowledge creation jobs more generally, as well as managerial jobs and professional jobs such as in law and finance (Gibbons and Murphy, 1992; Chevalier and Ellison, 1999; Hong, Kubik and Solomon, 2000; Lerner and Wulf, 2007; Aghion, Van Reenen and Zingales, 2013; Ferrer, 2016; Bonatti and Hörner, 2017).

performance incentives take effect. I also find no evidence of differential selection into the treatment or control cohort. Results are robust to estimation with synthetic cohorts, where assignment to the treatment and control cohort is determined by the average age at which academics start their first tenured affiliation instead of the actual timing of the first tenured affiliation. I also find no evidence that more productive academics are more likely to start their first tenured affiliation after implementation of the reform, nor any significant productivity differentials in a placebo difference-in-differences estimation with two cohorts that tenure before the reform. These tests lend support to the causal interpretation of the effort effect estimates.

Professors appointed in the age-related pay scheme can switch to performance pay by changing affiliation or position, or renegotiating their contract after the reform is implemented. Academics in the control group can thus become treated. This implies that the above estimates provide a lower bound of the effort effects. Indeed, estimating the effort effects using a control group without switchers yields larger estimates of the pure quantity- and quality-adjusted research output increase of 17-23%. On the other hand, estimates of the effects on average quality remain virtually unchanged.

There is considerable heterogeneity in the effort response, with both the quantity and quality response U-shaped in ability. This heterogeneity explains the decrease in average quality and lack of more work of the highest quality. The quantity effort response is strongest for low productivity academics, who do not produce high quality work. High ability academics also produce more publications, but not more of the highest quality. Medium ability academics, finally, do not increase output quantity and produce fewer high-quality papers.

To provide further evidence on the quality and impact of the work produced in response to performance incentives, I perform textual analysis to construct metrics for the similarity of papers to past and future publications to gauge their novelty and impact, respectively, similar to Kelly et al. (2021). The analyses confirm the quality response results: the additional work produced in response to performance pay is, on average, not the most novel or impactful. The most productive academics produce more low to medium novel work that generates mid- to high-level follow-on, low productivity academics produce additional papers that are just above the median in terms of both similarity to past and future work, while medium productive academics produce more papers that are very novel, but also garner very little follow-on work.

Finally, I estimate the selection effect of performance pay by testing for differential changes in switching rates by age and cohort for academics across productivity levels in another difference-in-differences framework. I use the fact that only academics who already hold a tenured affiliation before the reform can select into the performance pay scheme, by changing position or renegotiating contract. Accordingly, the treated-control designation for the selection analysis is the opposite of what it was for the effort effect analysis. Age is the second dimension of variation I exploit. The basic wage schemes of the age-related and performance pay schemes compare differently at different ages. The schemes intersect only once, and wages increase with age only in the age-related scheme. The performance pay scheme is therefore relatively less attractive for older academics, and selection incentives decrease by age. If selection into performance pay is positive, older academics thus need to be relatively more productive to want to switch. Indeed, I find that a higher productivity increases the switching rates of treated academics less when they are older.

Performance pay thus attracts more productive academics, and the selection effect is positive.

Taken together, this paper shows that performance pay incentives commonly found in knowledge creation jobs – particularly ones involving career concerns incentives – attract more productive workers and significantly increase knowledge output quantity. They do not yield more of the highest quality work, however. Importantly, this means that both the nature of the task and the kind of incentives matter for the effect performance pay has on output. If tasks are complex, with multi-dimensional effort and output and noisy output measures, the effort response might be more positive or positive only in the dimension(s) with less noisy output measures. Moreover, relative to explicit incentives only, the introduction of career concerns incentives in such a setting decreases the likelihood of positive quality effort and selection effects.

This paper contributes to the literature on performance incentives. While this literature provides ample evidence of the effects of performance pay - especially effort effects - in more routine tasks with precise output measures (e.g. Lazear, 2000; Shearer, 2004; Bandiera, Barankay and Rasul, 2005), there is relatively scant evidence on the effects of performance incentives in contexts with highly complex tasks and imprecise output measures. Furthermore, the (empirical) literature on incentives has mostly studied the effects of explicit performance incentives, such as piece rates, bonus pay, tournaments and monitoring schemes (see e.g. Oyer and Schaefer, 2011; Lazear and Oyer, 2012). The contribution of the current paper therefore lies in studying effort and selection effects of explicit and implicit career concerns incentives in knowledge creation (a highly complex task with noisy output measures). In doing so, the paper contributes to the literature on career concerns as well. Gibbons and Murphy (1992) and Camargo, Lange and Pastorino (2021) provide evidence of wage profiles consistent with career concerns and other forms of performance pay, while Miklós-Thal and Ullrich (2016) and Xu, Nian and Cabral (2020) document effort patterns consistent with career concerns. To the best of my knowledge, however, this paper is the first to present causal evidence of the effort and selection effects of career concerns incentives. Furthermore, the paper studies multitasking problems that career concerns incentives give rise to when tasks are complex and output measures noisy.[3] As such, it contributes to the multitasking literature as well. Alexander (2020) and Gupta (2021) document multitasking behaviors in response to explicit incentives in health care, and Glewwe, Ilias and Kremer (2010) and Andrabi and Brown (2022) in education. The current paper provides evidence of multitasking behaviors in knowledge creation in response to career concerns, and shows that these behaviors can be highly non-linear, varying with agent ability.

The paper also contributes to the literature on incentives for innovation and knowledge creation, as well as the literature on university governance (Aghion et al., 2010; Haeck and Verboven, 2012; McCormack, Propper and Smith, 2014; MacLeod and Urquiola, 2021) and the process and organization of knowledge creation (e.g. Jones, 2009; Waldinger, 2012; Azoulay, Fons-Rosen and Graff Zivin, 2019; Agarwal and Gaule, 2020). Many papers in the first literature look at incentives for commercializable knowledge and the commercialization of knowledge (patenting) (Lach and Schankerman, 2004, 2008; Azoulay, Ding and Stuart, 2009; Hall and Harhoff, 2012; Hvide and Jones, 2018). In doing so, the academic fields that can be studied is generally restricted to the (applied) sciences. Moreover, it is frequently difficult to differentiate

[3]Dewatripont, Jewitt and Tirole (1999) also study a career concerns model with multitasking, but their analysis centers around total effort across equally noisy output signals, while this model focuses on effort allocation across tasks (quantity and quality) that differ in the noise with which their output is measured.

between changes in the production of knowledge that can be commercialized and the rates at which such knowledge is commercialized (patented). Of the papers that study incentives for academic (basic) research more generally, many focus on the effects of funding or awards (Azoulay, Graff Zivin and Manso, 2011; Chan et al., 2014; Borjas and Doran, 2015). In these settings, it is difficult to distinguish between the effect of funding itself and the incentive effect of funding or awards on output. Furthermore, it is difficult to distinguish between the effort effect of performance incentives and selection into the funding or award schemes. The contribution of the current paper thus lies in providing causal and separate evidence of the effort and selection effects of common performance incentives in knowledge creation. Moreover, the paper studies these effects across all academic fields and productivity levels, and documents and analyzes the substantial heterogeneity along both dimensions.

Finally, by studying the effect of performance incentives on the quality, impact and novelty of knowledge output, the paper contributes to the literature on incentives for novelty and creativity. A number of recent papers study the effects of competition in online platforms on the creativity or novelty of outputs such as logo designs (Gross, 2020), novels (Wu and Zhu, 2022) and software development (Boudreau, Lacetera and Lakhani, 2011; Boudreau, Lakhani and Menietti, 2016; Graff Zivin and Lyons, 2021), with mixed results. Gibbs, Neckermann and Siemroth (2017) find that rewarding employees for ideas for product and process improvements raises the quality of ideas submitted in a field experiment setting, while Erat and Gneezy (2016) find that piece-rate pay does not alter creativity, but competitive incentives reduce creativity in a lab experiment. Ederer and Manso (2013) show that the structure of rewards, allowing for early failure and rewarding long-term success, is important for innovation in another lab setting. In this literature, the incentives studied are generally explicit, short-term incentives, and the tasks relatively narrowly defined and short. The current paper thus contributes by studying the effects of long-term explicit and implicit incentives on broad scope innovation effected over a long time.

The paper is structured as follows: section 2 provides information on the institutional background and section 3 presents the theoretical model. The empirical analysis makes up section 4, while 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 (flat) wage (BMBF, 2002). Before the reform, professors were paid according to an age-related pay scheme ("C-pay") in which pay increased at a pre-determined rate with age (Oeffentlicher Dienst, 2004; Detmer and Preissler, 2006).

2.1 Performance Pay (W-Pay)

Only tenured professors can earn substantial bonuses in the performance pay scheme[4]. Accordingly, I restrict attention to tenured professors in the empirical analyses of the paper. There are two tenured

[4] Untenured Professors can earn only small, non-pensionable supplements, and only in special circumstances (Detmer and Preissler, 2005).

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").

In the performance pay scheme, professors can earn performance bonuses in two ways[5]: as attraction or retention bonus, or for on-the-job performance (BMBF, 2002). These bonuses can potentially more than double a professor's monthly wage, and thus constitute high-powered incentives. Although federal and state laws lay down ground rules, universities have discretion in whom to award performance bonuses, and how much (Handel, 2005; Detmer and Preissler, 2006).

Attraction or retention bonuses are wage premia that are determined in contract (re)negotiations between a candidate and university. They are awarded on the basis of a professor's qualifications and past achievements and performance, taking into account applicant pool quality and labor market tightness (Detmer and Preissler, 2005). Performance is assessed using measures such as the number and impact or ranking of publications, funding grants and prizes, teaching evaluations and placement record of students. To be able to negotiate an attraction or retention bonus, professors often need to show proof of another offer (Detmer and Preissler, 2005).

An academic's past performance is thus used to update beliefs about their ability, and the (bonus) pay offered is driven to reflect beliefs about the academic's expected productivity under the influence of competition in the labor market. As such, the attraction and retention bonuses constitute implicit, market-based incentives in the form of career concerns[6]. These incentives do not derive from an explicit performance contract between an academic and any particular university, but rather from the academic's expectation to be able to influence future attraction or retention bonuses - either at the current university or another - by exerting more effort now. Career concerns incentives, and indeed attraction or retention bonuses, are very common in academia and knowledge creation jobs more generally, as well as in managerial and professional jobs such as in law and finance (Gibbons and Murphy, 1992; Chevalier and Ellison, 1999; MacLeod and Parent, 2000; Lerner and Wulf, 2007; Aghion, Van Reenen and Zingales, 2013; Ferrer, 2016; Bonatti and Hörner, 2017).

The on-the-job performance bonuses are awarded for performance over multiple years (often at least 3) in the current professorial position. They can be awarded for performance in research, art, teaching, mentoring and supervision (BMBF, 2002; Handel, 2005). Much like for attraction and retention bonuses, universities assess research performance for on-the-job performance bonuses using the number and impact or rank of publications, external research grants, patents and research prizes, while exceptional teaching evaluations, the development of didactic methods and teaching grants and prizes can serve as evidence of special teaching achievements (Detmer and Preissler, 2005; Universitaet Regensburg, 2016; Gien, 2017). University statutes stipulate award procedures for on-the-job bonuses, including explicit performance criteria and corresponding bonus amounts.

Universities can award both attraction/retention and on-the-job bonuses on a permanent basis, for a fixed term (initially) or even as a one-off payment (Detmer and Preissler, 2004, 2005). If bonuses are

[5] A third type of bonus, for taking on management roles or tasks, is not linked to performance, and therefore not discussed here.

[6] The reform also enabled professors to extract pay supplements from third-party awarded funds (BMBF, 2002). To the extent that grant application committees use an academic's past performance to update their beliefs about the academic's ability and chances of success, these grant pay supplements give rise to career concerns incentives as well.

awarded for a fixed term with renewal option, universities frequently enter into an indivualized target agreement with the respective professor (Detmer and Preissler, 2006). The target agreement specifies 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 bonus continues to be paid, either for another 3- to 5-year period, or permanently. Target agreements may also allow for partial fulfillment: when a professor secures external funds below a certain threshold for instance, the bonus that they (continue to) receive is lower (Detmer and Preissler, 2006). On-the-job bonuses and target agreements constitute explicit performance incentives, of the sort commonly seen in knowledge creation, managerial and professional jobs more generally (Gibbons and Murphy, 1992; Hong, Kubik and Solomon, 2000; Lerner and Wulf, 2007).

Since the incentives of the on-the-job performance bonuses and target agreements derive from an explicit performance contract between a professor and their university for professors appointed in the performance pay scheme, their incentive effect commences upon entry into the performance pay scheme. In contrast, the career concerns incentives of the attraction and retention bonuses take effect from the moment the reform is announced, for academics who expect to fall under the performance pay scheme in the future. For those academics, improved performance - even before implementation of the reform - increases the probability that they can command higher attraction or retention bonuses once they enter into the performance pay scheme, after reform implementation. I use this differential timing to separately identify the effort effect of the career concerns and explicit performance incentives in the empirical analyses.

The award of performance bonuses is not a rare occurrence in the performance pay scheme. In 2006, for instance, 77% of professors in the performance pay scheme received a bonus (BMI, 2007). The most common and largest bonuses are the attraction and retention bonuses. They were awarded almost six times as often as on-the-job performance bonuses in 2005 and more than three times as often in 2006 (BMI, 2007). Moreover, about 75% of the total amount of bonus pay in the performance pay scheme up until 2008 was awarded as attraction or retention bonus (Biester, 2010). This too aligns with performance pay in knowledge creation jobs more generally, with wage premia negotiated at hiring often far exceeding (short-term) bonuses linked to performance on the job (Lerner and Wulf, 2007).

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

In the age-related pay scheme, wages increase every two years, from the age of 21 to the age of 49 (Oeffentlicher Dienst, 2004; Detmer and Preissler, 2006). In contrast, the basic wage in the performance pay scheme does not vary with age, and the level is such that the basic wage schedules of the two pay schemes have a single crossing point.[7] Depending on the specific pay level of a tenured university professor (C3 or C4 in the age-related system; W2 or W3 in the performance pay system), the age-related wage starts to exceed the basic wage in the performance pay scheme at age 33 or 43 (Cf. Figure 6a) (Oeffentlicher Dienst, 2004; Handel, 2005).

The age-related pay system featured performance incentives as well, but these incentives were much weaker and impacted only the most productive professors. In the age-related pay scheme, only professors

[7]Some age-related pay increases were introduced in the W-pay scheme as of 2013, after the German federal constitutional court ruled that basic wages in the W-scheme were too low on 2 Feb. 2012 (Jochheim, 2015). These changes fall outside the sample period used for the analyses in this paper.

in the highest pay level (C4) could earn bonuses, and only when they received further C4 offers. Consequently, only a small fraction of professors qualified for and received bonuses under the age-related pay system(Detmer and Preissler, 2006; Jochheim, 2015). Moreover, these bonuses were standardized and substantially lower than those in the performance pay system, generally ranging from around 490 to 750 euro per month (Preissler, 2006; BMI, 2007).

By comparison, professors in the performance pay scheme can earn bonuses totaling up to 5,241 euro per month, or more if needed to attract or retain a professor from abroad or who already earns this amount of bonuses[8] (BMBF, 2002; BMI, 2007). Combined with the fact that the basic wage is lower in the performance pay scheme than at most ages in the age-related scheme, this gives rise to a greater spread in professorial pay in the performance pay system. Furthermore, any tenured professor can earn bonuses in the performance pay system, and not just when receiving additional offers. In 2006, for instance, 77% of professors in the performance pay scheme received bonus pay (BMI, 2007), while Handel (2005) calculates that only 16.5% of professors received bonuses in the age-related pay system. As a result, only 3.55% of the total professorial pay volume was spent on bonuses in the age-related system, while an estimated 26% of the professorial pay volume was available for performance bonuses under the performance pay scheme(Expertenkommission, 2000; Handel, 2005). Indeed, since the academic pay reform was mandated to be cost-neutral (the average professorial pay at the federal and state level was to remain at the respective pre-reform levels (BMBF, 2002)) and because of the single-crossing property of the basic wage schedules, a larger portion of pay depends on performance in the performance pay scheme.

All in all, the performance pay system offers substantially higher-powered performance incentives than the age-related pay system. Moreover, performance incentives are stronger across the ability distribution. For lower productivity academics, because they can now earn bonuses. For higher productivity academics, because they can now earn bonuses on-the-job (without outside offers), and bonuses can be larger[9].

2.3 Implementation

The federal law introducing the performance pay scheme was passed 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 by 1 January 2005. Only three states (Bremen, Niedersachsen and Rheinland-Pfalz) did so before the end of 2004 (Detmer and Preissler, 2005).

Any professorial appointment ("Ernennung") after implementation of the reform falls under the new performance pay scheme, while existing professorial contracts continue to fall under the age-related pay scheme (Detmer and Preissler, 2004). Hence I observe professors in both the age-related and performance pay scheme as of 1 January 2005. I exploit this to estimate the effort effect of the performance incentives in the empirical analysis. Academics who are appointed to a new affiliation or position, or who renegotiate their contract after implementation of the pay reform switch from age-related to performance pay (Detmer and Preissler, 2004). Once a professor switches to performance pay, they can never go back to age-related

[8] Consequently, in the performance pay scheme professors can earn more than 9866 Euro per month (the B10 pay level, another pay scale for civil servants) (BMBF, 2002). In contrast, the maximum salary for C4 professors was capped at the B10 pay level, and generally fell between the B5 and B7 pay level (6821 to 7582 Euro per month) (Flämig et al., 2013).

[9] The most productive academics, especially those with the potential to move internationally and attract large grants, were already highly incentivized. Their performance incentives likely increased the least.

pay. I exploit such switches to estimate the selection effect of the performance incentives. Online appendix A2 provides supplementary institutional details.

3 THEORETICAL FRAMEWORK

The model outlined here and presented in detail in Online appendix A1.2 has three key features characteristic of knowledge work. First, effort and output have two dimensions, quantity and quality, and output quality is less precisely measured than quantity. The combination of complex tasks with differentially noisy output measures gives rise to multitasking problems (Holmström and Milgrom, 1991). Second, output is a function of both effort and ability, and there is uncertainty about ability. Since output is publicly observable, the market gradually learns about agent ability from the noisy output measures. Combined with market forces - such as in a competitive labor market - this learning process gives rise to implicit, market-based incentives: career concerns (Fama, 1980; Holmström, 1999). Third, principals can offer contingent contracts and agents are risk averse. The contracts introduce explicit performance incentives. These do not fully eliminate career concerns, because of agents' risk aversion (Gibbons and Murphy, 1992).[10]

The combination of career concerns and explicit performance incentives is representative of incentive structures in academia and knowledge creation jobs more generally, as well as managerial jobs and professional jobs such as in law and finance (Gibbons and Murphy, 1992; Chevalier and Ellison, 1999; Hong, Kubik and Solomon, 2000; Lerner and Wulf, 2007; Ferrer, 2016; Bonatti and Hörner, 2017). It is also the structure of the performance pay scheme introduced in German academia, with attraction and retention bonuses that give rise to career concerns, plus explicit on-the-job bonuses. The model thus captures key characteristics of the production function and performance incentives in knowledge work and, as such, allows for the evaluation of the effort and selection effects of performance pay in knowledge creation.

Consider a perfectly competitive labor market in which risk neutral principals interact with infinitely lived, risk averse agents. Each period ($t = 0, 1, 2, ...$) agents choose effort $\overrightarrow{e_t} \geq \vec{0}$ and produce output $\overrightarrow{y_t}$. Effort cannot be observed by principals, but output is observable to all market participants. Output has two dimensions, quantity $y_{p,t}$ and quality $y_{q,t}$, each of which is a noisy signal of agent ability θ and effort put towards the respective output dimension: $y_{p,t} = \theta + e_{p,t} + \varepsilon_t$ and $y_{q,t} = \theta + e_{q,t} + \nu_t$. Here ε_t, ν_t are iid sequences of normally distributed noise terms: $\varepsilon_t \sim N\left(0, \sigma_\varepsilon^2\right), \nu_t \sim N\left(0, \sigma_\nu^2\right), \sigma_\varepsilon^2 > 0, \sigma_\nu^2 > 0$. I assume that the quality measure is more noisy, $\sigma_\varepsilon^2 < \sigma_\nu^2$.

Ability is not known either by agents or principals, but there is common knowledge about the prior ability distribution. In particular, an agent's ability θ_i is an iid draw from a normal distribution with mean $m_{i,0} \in [\underline{m}, \overline{m}], \underline{m} \geq 0$ and variance $\sigma_0^2 > 0$. I allow for abilities to be drawn from distributions with different means, to reflect the possibility that agents can distinguish themselves before entering the labor market, for instance during their studies.[11]

Agents have CARA preferences with risk aversion parameter r. Their cost of effort is multivariate

[10] Dewatripont, Jewitt and Tirole (1999) also study a career concerns model with multitasking, but their analysis focuses on total effort across equally noisy output signals. This model focuses on effort allocation across tasks (quantity and quality) that differ in the noise with which their output is measured.

[11] This is in line with Harris and Holmstrom (1982), who assume that the mean of the ability distribution is a function of education.

quadratic, specifically

$$g\left(\overrightarrow{e_t}\right)=\left[\overrightarrow{e_t}-\vec{\bar{e}}\right]^T\frac{1}{2}G\left[\overrightarrow{e_t}-\vec{\bar{e}}\right]=\left[\overrightarrow{e_t}-\vec{\bar{e}}\right]^T\frac{1}{2}\begin{bmatrix} c & d \\ d & c \end{bmatrix}\left[\overrightarrow{e_t}-\vec{\bar{e}}\right]. \tag{1}$$

Inclusion of $\vec{\bar{e}}$ follows Holmström and Milgrom (1991) and ensures positive effort levels even in the absence of performance pay. These minimum effort levels capture other mechanisms and institutions that drive effort, such as intrinsic motivation or minimum output requirements (e.g. tenure requirements). For simplicity, and without loss of generality, set $c=1$. Assume further that $0<d<1$, so that G is positive definite and the relevant optimization problems are concave. This means that effort towards quantity $e_{p,t}$ and effort towards quality $e_{q,t}$ are substitutes. I allow for the degree of substitution to vary by ability class, $d=d\left(m_0\right)$. In particular, I follow Rubin, Samek and Sheremeta (2018) in assuming that quantity and quality effort are weaker substitutes for agents in higher ability classes, $\frac{\partial d}{\partial m_0}<0$.

I assume only short-term contracts are feasible and restrict attention to linear contracts of the form $w_t\left(\vec{y}_t\right)=c_t+\overrightarrow{b_t}^T\vec{y}_t$, as in Gibbons and Murphy (1992).[12] In each period, the timing is as follows: principals offer a contract (w_t) to agents; agents pick the contract that yields the highest expected utility; agents choose and exert effort; output materializes, principals receive the output produced by agents they employ and agents are paid according to their contract terms.

Because of perfect competition, principals pay agents their expected productivity and hence earn zero profits. To determine expected productivity, the market uses the noisy output measures to update beliefs about agent ability. This gives rise to incentives for the agent to exert effort: career concerns. Also because of perfect competition, principals set the explicit bonuses $\overrightarrow{b_t}^T$ so as to maximize an agent's certainty equivalent. Proposition 1 then characterizes the unique equilibrium:

Proposition 1 - Equilibrium Performance Pay Contracts and Effort: *In a perfectly competitive labor market where agents face both career concerns and linear performance pay contracts, the unique equilibrium effort level for output quantity $e^*_{p,t}$ is given by the first argument in* (2) *when $\vec{e}_t\gg 0$ and the second argument when $e^*_{q,t}=0$, while equilibrium quallity effort $e^*_{q,t}$ is given by* (3). *The corresponding unique equilibrium performance pay rates $b^*_{p,t}$ and quality $b^*_{q,t}$ are given by the first argument of* (4) *and* (5) *when $\vec{e}_t\gg 0$ and the second argument when $e^*_{q,t}=0$.*

$$e^*_{p,t}=\left[\bar{e}_{p,t}+d\bar{e}_{q,t}+\frac{\sigma_t^2+\sigma_\nu^2\left(1+CV_tCC_t\sigma_\nu^2\right)}{\sigma_t^2+\sigma_\nu^2+CV_t},\bar{e}_{p,t}+\frac{1+r\left(\sigma_\nu^2-d\sigma_\varepsilon^2\right)\left(1+CC_t\left(1+d\right)CV_t\right)}{\left(1+d\right)D_t}\right] \tag{2}$$

$$e^*_{q,t}=max\left[0,\bar{e}_{q,t}+\frac{1+r\left(\sigma_\varepsilon^2-d\sigma_\nu^2\right)\left(1+CC_t\left(1+d\right)CV_t\right)}{\left(1+d\right)D_t}\right] \tag{3}$$

[12] This not only ensures tractability, but Holmström and Milgrom (1991) show that optimal contracts are also linear in a setting with comparable assumptions about agent preferences and output noise.

$$b_{p,t}^{*}=\left[\frac{1+(1-d)\,r\sigma_{\nu}^{2}-CC_{t}\left(\sigma_{\nu}^{2}\left(1+r\left(\sigma_{\nu}^{2}-d\sigma_{\varepsilon}^{2}\right)\right)-r\left(1+d\right)\sigma_{t}^{2}\left(\sigma_{\varepsilon}^{2}-\sigma_{\nu}^{2}\right)\right)}{1+r\left(2\left(1+d\right)\sigma_{t}^{2}+\sigma_{\varepsilon}^{2}+\sigma_{\nu}^{2}\right)+\left(1-d^{2}\right)r^{2}\left(\sigma_{\varepsilon}^{2}\sigma_{\nu}^{2}+\sigma_{t}^{2}\left(\sigma_{\varepsilon}^{2}+\sigma_{\nu}^{2}\right)\right)},\frac{-\left(\sigma_{t}^{2}+\sigma_{\nu}^{2}\right)\left(CC_{t}\sigma_{\nu}^{2}-1\right)}{\sigma_{t}^{2}+\sigma_{\nu}^{2}+CV_{t}}\right] \tag{4}$$

$$b_{q,t}^{*}=\left[\frac{1+(1-d)\,r\sigma_{\varepsilon}^{2}-CC_{t}\left(\sigma_{\varepsilon}^{2}\left(1+r\left(\sigma_{\varepsilon}^{2}-d\sigma_{\nu}^{2}\right)\right)-r\left(1+d\right)\sigma_{t}^{2}\left(\sigma_{\nu}^{2}-\sigma_{\varepsilon}^{2}\right)\right)}{1+r\left(2\left(1+d\right)\sigma_{t}^{2}+\sigma_{\varepsilon}^{2}+\sigma_{\nu}^{2}\right)+\left(1-d^{2}\right)r^{2}\left(\sigma_{\varepsilon}^{2}\sigma_{\nu}^{2}+\sigma_{t}^{2}\left(\sigma_{\varepsilon}^{2}+\sigma_{\nu}^{2}\right)\right)},\frac{\sigma_{t}^{2}\left(CC_{t}\sigma_{\nu}^{2}-1\right)}{\sigma_{t}^{2}+\sigma_{\nu}^{2}+CV_{t}}\right] \tag{5}$$

Here $CV_t = r\left(\sigma_{\varepsilon}^2\sigma_{\nu}^2 + \sigma_t^2\left(\sigma_{\varepsilon}^2+\sigma_{\nu}^2\right)\right)$, $D_t = 1 + r\left(2\left(1+d\right)\sigma_t^2 + \sigma_{\varepsilon}^2 + \sigma_{\nu}^2\right) + r\left(1-d^2\right)CV_t$, $CC_t = \sum_{\tau=1}^{\infty}\delta^{\tau}\frac{\sigma_0^2}{\sigma_{\varepsilon}^2\sigma_{\nu}^2+(t+\tau)\sigma_0^2(\sigma_{\varepsilon}^2+\sigma_{\nu}^2)}$ and $\sigma_t^2 := var\left(\theta_t\right) = \frac{\sigma_0^2\sigma_{\varepsilon}^2\sigma_{\nu}^2}{\sigma_{\varepsilon}^2\sigma_{\nu}^2+t\sigma_0^2(\sigma_{\varepsilon}^2+\sigma_{\nu}^2)}$.

The proof relies on the realization that ability enters output additively and does not affect the marginal cost of effort, so the optimal bonus $\overrightarrow{b_t}$ does not depend on m_t (the mean of the posterior distribution of θ) (Macleod, 2022). Future income risk is therefore unaffected by effort, and career concerns incentives are unaffected by the introduction of explicit performance incentives. Equilibrium effort is then characterized by first order conditions that equate the marginal cost of effort to the sum of the respective career concerns incentive and optimal bonus. Similarly, $\overrightarrow{b_t}$ affects only the effort and income risk in period t and is therefore chosen to maximize agent utility that period. This amounts to optimizing the certainty equivalent with respect to $\overrightarrow{b_t}$. Online appendix A1.2 provides a complete, formal proof.

The career concerns incentive for quantity effort is $\sigma_{\nu}^2 CC_t$. This increases with uncertainty about ability σ_0^2 and output quality σ_{ν}^2, but decreases with uncertainty about output quantity σ_{ε}^2. The relationship with σ_0^2 and σ_{ε}^2 is standard (cf. Gibbons and Murphy (1992); Holmström (1999)). The relationship with σ_{ν}^2 follows because the market updates beliefs about agent ability using both quantity and quality output. Since the quality signal is less informative when σ_{ν}^2 is large, agents have a stronger incentive to exert quantity effort in a bid to sway beliefs when σ_{ν}^2 increases. An equivalent result holds for quality effort.

Absent career concerns, the optimal bonuses for both quantity and quality effort decrease with noise in either output measure as well as risk aversion.[13] This is a standard result; the greater uncertainty and risk aversion, the higher the risk premium for the agent (Holmström and Milgrom, 1991). With career concerns, the explicit incentives also decrease with career concerns incentives, thus balancing out - though not eliminating - these incentives.[14]

Finally, because quantity and quality effort are substitutes, optimal quantity effort is a response to the quantity effort career concern incentive and the quantity bonus, as well as the quality effort career concern incentive and the quality bonus - the latter two mitigated by the rate of substitution d.[15] Because d is assumed to be lower for higher ability academics, quantity effort varies with ability. Equivalent results apply for quality effort. Taken together, quantity and quality effort are complex functions of risk aversion, output noise and the substitution rate. Because output quality is noisier, risk aversion has a bigger impact on quality effort - even more so in lower ability classes due to higher substitution rates. This gives rise to interesting nonlinearities in the effort and selection response.

In order to analyze these responses, I compare behavior in the above equilibrium to that under flat wage

[13] These optimal bonuses are $b_p^* = \frac{1+(1-d)r\sigma_{\nu}^2}{1+r(\sigma_{\varepsilon}^2+\sigma_{\nu}^2)+(1-d^2)r^2\sigma_{\varepsilon}^2\sigma_{\nu}^2}$ and $b_q^* = \frac{1+(1-d)r\sigma_{\varepsilon}^2}{1+r(\sigma_{\varepsilon}^2+\sigma_{\nu}^2)+(1-d^2)r^2\sigma_{\varepsilon}^2\sigma_{\nu}^2}$ (for $e_{q,t}^* > 0$). See (74)-(75) in A1.4.2.

[14] This follows from comparing equations (74)-(75) in online appendix A1.4.2 to (4)-(5) above.

[15] This follows from the first-order conditions (18)-(19) in online appendix (A1.2): $e_{p,t}^* = \bar{e}_p + \frac{1}{1-d^2}\left(b_{p,t} + \sigma_{\nu}^2 CC_t - d\left(b_{q,t} + \sigma_{\varepsilon}^2 CC_t\right)\right)$ and $e_{q,t}^* = \bar{e}_p + \frac{1}{1-d^2}\left(b_{q,t} + \sigma_{\varepsilon}^2 CC_t - d\left(b_{p,t} + \sigma_{\nu}^2 CC_t\right)\right)$.

pay. If agents are paid flat wages, as in the age-related pay system in German academia, all agents provide minimal effort in equilibrium; $e^*_{p,t} = \bar{e}_p$ and $e^*_{q,t} = \bar{e}_q$ (See online appendix A1.1). The comparison yields four predictions: for the effort response, heterogeneity in the quantity and quality response, and the selection effect. These are summarized below and formally stated in Propositions 2, 3a, 3b and 4 in online appendix A1.3.1-A1.3.3 (including proofs). The proofs, which are cumbersome but otherwise straightforward, amount to showing 1) when (2) and (3) are larger than $\bar{e}_p$,$\bar{e}_q$ or 0 ; 2) when the derivatives of (2) and (3) with respect to m_0 are positive or negative, and 3) when the certainty equivalent under performance pay is greater than under a flat wage, and how this varies with m_0.

First, compared to a flat wage, quantity effort is unambiguously larger under performance pay. Quality effort also increases in all ability classes, but only if risk aversion is low. If risk aversion is high, quality effort decreases in ability classes below a threshold m^{pp} (Proposition 2). This threshold increases with uncertainty about output quality, so that quality may decrease in all ability classes if the quality measure is very noisy (Corollary 1). Quality effort can even be reduced to zero across ability classes if risk aversion is sufficiently high and minimum effort levels low (Lemma 1). The theory thus predicts a positive quantity effort effect, but potential crowding out of quality effort - more so if risk aversion is high and the quality measure noisy.

Second, quantity effort increases monotonically with ability if risk aversion is low, but is U-shaped when risk aversion is high, and monotonically decreasing with ability when quality effort is zero for all ability classes (Proposition 3a). Relative to a flat wage, output quantity thus increases most in high ability and least in low ability classes in response to performance pay when risk aversion is low. When risk aversion is high, the quantity increase is smallest in intermediate ability classes. It is smallest in high ability classes when there is complete crowding out of quality effort across ability classes.

Third, if all ability classes exert positive quality effort (if risk aversion is low or minimum quality effort high), output quality increases monotonically with ability. Relative to a flat wage, the highest ability class then increases quality effort most or decreases it least in response to performance pay. If there is crowding out of quality under performance pay in low ability classes only, the change in output quality relative to flat wage pay is J-shaped or U-shaped in ability. The U-shape occurs if, for instance, tenure requirements or intrinsic motivation are higher for higher ability classes, so that complete crowding out amounts to a smaller drop in quality effort for the lowest ability classes (Corollary 2). The change in quality effort is either constant or decreasing in ability class if there is complete crowding out of quality effort across the ability distribution (Proposition 3b).

Fourth, selection into performance pay is positive when risk aversion is low, but ambiguous (positive, negative or neither) when risk aversion is high (Proposition 4). These selection effect predictions carry over to selection into academia.

Taken together, the theory yields an unambiguous prediction for the average quantity effort response only. The average quality effort response, heterogeneity in both quantity and quality responses, and the selection effect depend on parameters like risk aversion, output quality noise and minimum effort requirements because of the multitasking nature of the knowledge production function. They are therefore ultimately empirical questions.

Online appendix A1.4 provides two supplementary sets of comparative statics. Using steps similar to the ones outlined above, it characterizes the equilibrium in two alternative performance pay systems, with career concerns only or explicit incentives only, and analyzes their effort and selection effects relative to flat wage pay. It compares these effects to those of a system with both career concerns and explicit incentives. Relative to a system with career concerns and explicit incentives, the selection effect of explicit incentives only is more likely to be unambiguously positive, the quality effort effect more likely to be positive, and quantity effort less likely to be monotonically decreasing with ability (Corollary 4). A system with career concerns only, on the other hand, is less likely to give rise to unambiguously positive selection or a positive quality effort effect than a system with career concerns and explicit incentives, and quantity effort is less likely to increase monotonically with ability (Corollary 3). Thus, the presence of career concerns incentives in performance pay systems when output signals are differentially noisy decreases the likelihood of positive quality effort and selection effects, relative to systems with explicit incentives only. Whether performance pay increases output quantity *and* quality, and attracts more productive workers in knowledge creation therefore requires empirical analysis. I turn to this next.

4 EMPIRICAL ANALYSIS

4.1 Data Description

To estimate the effort and selection effects of performance pay in knowledge creation, I constructed a data set that encompasses the affiliations of the universe of academics in German academia for the years 1999-2013, as well as their publication records in 1993-2012. The data set also contains personal information, the year in which an academic completed their PhD, obtained their postdoctoral qualification, and started working in academia. The data set encompasses 50174 academics who held a tenured position at a German public university at some point between 1999 and 2013.

The performance pay reform applied to all public institutions of higher education. I restrict attention to public *universities*, since the research output of other institutions is incomparable to that of universities, and I focus on research productivity as outcome variable (BMBF, 2002). There were 89 public universities in the years spanned by the panel[16] (HRK, 2014). I restrict attention to academics who hold a tenured affiliation at some point in the sample period, because performance bonuses can be earned in tenured positions only.

To construct the data set, I draw from three main, mostly unstructured, input data sets; Kuerschners Deutscher Gelehrten Kalender (hereafter: DGK) and Forschung & Lehre Magazine (hereafter: FuL) for affiliations, and ISI Web of Science for publications data. DGK is a comprehensive encyclopedia of tenured and untenured academics at German universities[17] (De Gruyter, 2006, 2008, 2013). I use DGK as a register of the universe of academics who are eligible for or who hold a tenured professorial position at a German university, and extract personal as well as professional information from it (academic affiliations over time, start and end year of academic career in Germany, degrees earned etc.).

[16] This constitutes the vast majority of universities in Germany, since there were only a few - small - private universities during this period.

[17] DGK includes everyone who is eligible for a tenured professorial position and actively teaching and researching at universities that can award doctoral degrees in Germany, Austria and Switzerland. People are eligible for a tenured position once they have passed a post-doctoral and teaching qualification ("Habilitation" and "Lehrbefugnis"), or equivalent.

I cross-check and supplement the professional information in DGK with information from FuL (DHV, 1999-2013). FuL is Germany's largest academic professional magazine. Every month, it publishes an overview of scholars who obtained their post-doctoral or teaching qualification ("Habilitation" and "Lehrbefugnis"), professorial appointments and offers that were extended, accepted or rejected. I extract publication records and citation counts for the years 1993-2012 from the ISI Web of Science database to construct measures of research productivity and use journal impact factors from the ISI journal citation report (JCR) of the year of publication to construct impact factor-rated publication measures (Clarivate Analytics, 2000-2012*b*, 1993-2012*a*).[18] Online appendix A3.1, A3.2 and A3.3 provide supplementary information about DGK, FuL and ISI, and how I use these data sources.

I match academics across the three input data sets on the basis of last name, initials and field, and discard any resulting duplicate matches. I further improve the matching by exploiting additional information such as the start or end date of academic careers to rule out implausible matches. Doing so yields an 83% match rate of academics whom FuL reports as having a tenured affiliation at 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. Where possible, I resolve such inconsistencies manually. Online appendix A3.4 and A3.5 provide supplementary information about the matching procedures used.

To construct an individual-level affiliations panel from the information in DGK and FuL, I use an iterative procedure to fill the affiliation information from DGK over time. I cross-check this information with offer and appointment information from FuL, after backdating by the average printing lag of FuL announcements.[19] Online appendix A3.6 provides supplementary information about the construction of the affiliations panel. Together with the publication records, this panel forms the basis for the data used for the effort effect analysis.

I code FuL information about extended and rejected professorial offers as contract renegotiations at the current university. After implementation of the reform, professors tenured in the age-related pay scheme switch to performance pay through succesful contract renegotiations and new appointments (Detmer and Preissler, 2004). I therefore use the renegotiations data, together with data on affiliation or position changes from the affiliations panel described above, to determine which professors switch from age-related to performance pay for the selection effect analysis.[20] In what follows, supplementary tables

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

[19]Specifically, I backdate information about accepted offers and appointments by four months, and take these dates to be the time of appointment to a new position. I effectively code acceptance as appointment to err on the side of conservative effort and selection effect estimates. Since there can be 3-4 months between offer acceptance and appointment (Wissenschaftsrat, 2005), academics are slightly more likely marked as having a (first) tenured appointment before the implementation of the reform based on FuL information. Some academics may thus be assigned to the control cohort for the effort effect estimation, when in fact they are treated. If anything, this would lead to an underestimation of the effort effect. For the selection effect estimation, some academics may be recorded as treated (starting out in age-related pay and hence able to switch to performance pay) when in fact they are control. This too would lead to a conservative estimate of the selection effect. Reassuringly, consistency checks reveal that the information in DGK regarding the timing (year) of appointments or affiliation changes differs from that in FuL for only 5% of individuals who change a (tenured) affiliation at least once. This is mostly down to printing lags and, where possible, is corrected manually.

[20]I assume that, whenever an academic receives an offer, they either accept and change affiliation or position (which would show in the affiliations panel), or reject and renegotiate their current contract. 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 lower productivity type. Including them in the pool of switchers would then yield a conservative estimate of the selection effect. Coded this way, renegotiations make up around 21% of switches in the data. Reassuringly, (BMI, 2007) reports that circa 18% of entries into performance pay are due to renegotiations ("Bleibeverhandlungen"). Since Preissler (2006) reports that only a very small number of professors chose to opt into the W-pay scheme without another/outside offer, the renegotiations and affiliation/position changes in the data should capture the vast majority

and figures in the online appendix are referenced with a number preceded by "A".

4.2 Effort Effect

In order to identify the effort effect of the career concerns and explicit performance incentives introduced with the pay reform in Germany, I use the fact that any professorial appointment as of 1 January 2005 falls under the performance pay scheme, whereas any appointment before this date falls under the old, age-related pay scheme.[21] Accordingly, academics who start their first tenured affiliation after 2004 switch to the performance pay scheme upon their appointment, while academics who start their first tenured affiliation before 2005 continue to fall under the age-related pay scheme. 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 that of academics who start their first tenured affiliation as of 1 January 2005 (the treatment group).

In the baseline, I use a three-year window to define the treatment and control group, in order to abstract from seniority effects. Results are, however, robust to longer or shorter cohort windows (Table A.1 panels F and G). I exclude academics who hold a foreign affiliation before their first tenured affiliation in Germany to avoid confounding effort and selection effects. The treated cohort comprises 2,844 academics, the control cohort 3,197.

4.2.1 Descriptive Statistics

I construct several measures of research productivity that are based on the publications of academics. I backdate publications of academic i in field f by the average publication lag in field f (rounded up to the nearest year) as reported in Björk and Solomon (2013). After correcting for average publication lags, I have productivity measures for (at least) 18 years, from 1993 through 2010. The productivity measures take all available publication types into account, including journal articles, books, book chapters and conference proceedings.[22]

Table 1 Panel A shows that academics in the treatment and control cohort had almost 3 publications per year on average, though the median academic does not have a publication in any given year. Weighting publications by the impact factor of the outlet in which it appears brings this sum to almost 10.[23] To put this in perspective, the top five general interest journals in economics had an impact factor rating between 1.806 (AER) and 4.775 (QJE) in 2005, while Science had an impact factor rating of 30.927 and Nature of 29.273 (Clarivate Analytics, 2000-2012*b*). The average total number of citations to publications from a given year is 102. Citations data was extracted from ISI Web of Science in January 2019, so there are at least 6 years between the publication date and the time at which citations were counted for each publication.

of contract switches.

[21] Using 2005 as uniform before-after cut-off yields a conservative estimate of the effort effect, since three states introduced performance pay a little earlier. Some of the control group is thus already treated before 2005.

[22] Citations are available only for journal articles, and do not include citations to books, chapters, patents, etc.

[23] I am using the two-year rather than the five-year impact factor, since the latter is available only as of 2007.

4.2.2 Baseline Difference-in-Differences

I estimate the effort effect in the following difference-in-differences model:

$$E\left[Y_{i,f,t-x_f}|X_{i,f,t}\right] = exp[\alpha_i + \beta_1 Post'02 * Treatment_i + \beta_2 Tenure_{i,j} * Treatment_i + \sum_{j=-7}^{7} ttt_{i,j} + \gamma_t] \tag{6}$$

The dependent variable, $Y_{i,f,t-x_f}$ is a productivity measure 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 α_i and γ_t, are individual and calendar year fixed effects, respectively. The individual fixed effects subsume academic field fixed effects here, because an academic's field is kept constant throughout. $Treatment$ is 1 for academics who start their first tenured affiliation at a public university in 2005, 2006 or 2007, and 0 for those who start their first tenured affiliation at a public university in 2002, 2003 or 2004 (the control cohort). $Post'02$ is 1 as of 2002 and 0 beforehand, while $Tenure$ is 1 as of the year in which an academic starts their first tenured affiliation and 0 beforehand.

This difference-in-differences specification distinguishes two before and after periods: before and after announcement of the reform, and before and after commencement of the first tenured affiliation. The moment the reform is announced, the lure of future attraction and retention bonuses takes effect, and with it career concerns incentives (see section 2.1). These career concerns apply to untenured professors as well, since the absence of tenure-track positions in Germany means academics need to move to a new university and negotiate a new contract to obtain tenure. Academics who anticipate starting their first tenured affilliation in the performance pay system thus face strong career concerns, as their pre-tenure performance can influence their tenure contract negotiations and pay in the performance pay system. I can therefore identify the effort effect of the career concerns component of performance pay from the differential change in productivity of about-to-be-tenured academics in the treatment and control cohort from before to after announcement of the reform ($Post'02 * Treatment_i$).

The incentive effect of the explicit on-the-job performance bonuses takes effect only after academics enter into the performance pay scheme. This coincides with the start of the first tenured affiliation for academics in the treated cohort. These explicit incentives come in addition to the career concerns incentives, which remain in effect, and unchanged.[24] I can therefore identify the effort effect of the explicit incentives component of performance pay from the differential change in productivity in the treatment and control cohort from before to after the start of the first tenured affilition ($Tenure * Treatment_i$).[25]

The announcement of the reform occurs at the same calendar time for all tenure cohorts, but at a different time relative to tenure. The start of the first tenured affiliation, on the other hand, occurs at the same relative time, but at a different calendar time for all cohorts. This allows me to identify the effort

[24] As noted in the theoretical framework, adding explicit incentives to career concerns incentives does not affect the career concerns incentives Macleod, 2022. Upon entry into the performance pay scheme, the only change in incentives is therefore the addition of the explicit performance incentives to career concerns. Hence $Post'02 * Treatment_i$ and $Tenure * Treatment_i$ separately identify the effort effect of career concerns and explicit incentives, respectively, in an incentive system that combines both.

[25] Universities generally announce either the total amount or number of on-the-job bonuses in a year at the beginning of the year (Lünstroth, 2011). These incentives thus vary by university and year. On top of that is the variation, at the individual academic level, in target agreements. Identifying the effort effect of the explicit performance incentives by exploiting cross-university variation is therefore infeasible, even aside from the potential bias due to sorting. Hence I estimate the effort effect of average explicit performance incentives here.

effect of the career concerns and explicit incentives components of performance pay separately. The sum of $Post'02 * Treatment_i$ and $Tenure * Treatment_i$ provide a difference-in-differences estimate of the total effort effect of career concerns plus explicit performance incentives in knowledge creation. Because of the way $Post'02$ and $Tenure$ are defined, the interactions capture **persistent** differential changes in research productivity.

The $ttt_{i,j}$ variables are relative time fixed effects in the form of time-to-tenure dummies. They control flexibly for productivity patterns in the run up to and after tenure that are common to treatment and control cohorts.[26] Because the specification includes individual fixed effects, including all calendar time and relative time fixed effects would yield a specification that is underidentified. I therefore include at most 15 year-to-tenure (relative time) fixed effects for each cohort, from seven years before to seven years after tenure.[27] If the timing of the start of the first tenured affiliation is exogenous, it follows from Athey and Imbens (2022) that estimation of (6) yields an unbiased estimate of the average effort effect of even the explicit performance incentives (with staggered adoption). I provide evidence of exogeneity of tenure timing below.

I estimate the model as a conditional quasi-maximum likelihood fixed-effect Poisson model[28], because the dependent variables are highly skewed with a large mass at zero and long right tail. The corresponding estimation results are shown in Table 1. Robust standard errors, clustered at the individual level, are reported throughout.

<Table 1 about here>

4.2.3 Baseline Results

Table 1 Panel B shows that research quantity and quality-adjusted quantity increase in response to performance incentives, but average quality decreases. The positive and significant (at 1%) $Post'02*Treatment$ interaction in column 1 implies that there is a persistent 18.3% increase in the number of publications in response to the career concerns incentives of performance pay.[29] There is no additional change in the number of publications in response to the explicit incentives of performance pay ($Tenure * Treatment_i$). To allow for a more easily interpretable result, I also estimate a linear fixed effects version of the baseline regression. The results from this estimation suggest that academics produce almost one extra publication every three years in response to career concerns incentives (column 1 in Panel A of Table A.1).

I find comparable results with quality-adjusted measures of productivity (cf. columns 2 and 3 in Table 1 Panel B). The impact-factor weighted number of publications increases by 14.2% in response to career

[26]Similar to Deshpande and Li (2019), the $ttt_{i,j}$ are included to avoid underweighting the long-run effect estimates of the performance incentives with staggered adoption, namely, the explicit incentives (De Chaisemartin and dHaultfoeuille, 2020; Borusyak, Jaravel and Spiess, 2021; Callaway and SantAnna, 2021; Sun and Abraham, 2021). Recall that the explicit incentives take effect at the start of the first tenured affiliation, which occurs at different points in calendar time for different tenure cohorts. In contrast, the career concerns incentives take effect at the same calendar time (2002) for all cohorts.

[27]Including 7 year-to-tenure dummies aligns with the institutional setting here. The median number of years between the end of the PhD and the completion of the habilitation is 7 and academics are, traditionally, required to have completed their habilitation before they become eligible for a tenured affiliation. Furthermore, dropping year-to-tenure fixed effects that are far from the time of treatment (here: tenure), allows for a more stable estimation of the effort effect (Borusyak, Jaravel and Spiess, 2021). Results are however robust to including other sets of year-to-tenure dummies (see e.g. Table A.1 Panel H), or including a non-linear function of relative time - the absolute value of time-to-tenure (Table A.1 Panel D).

[28]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, Monfort and Trognon (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.

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

concerns incentives, and the sum of citations to publications published in a given year increases by 13.8% (significant at 1% and 5%, respectively), with no additional changes in response to the explicit incentives of performance pay.

The average quality of publications, however, decreases in response to performance pay. The average impact factor rating of publications decreases by 9% (significant at 1%) in response to career concerns incentives, with no additional significant effect of explicit performance incentives (column 4 in Table 1)[30]. The coefficient estimate of the equivalent linear fixed effects estimation is -0.216 (column 4, Table A.1 panel A). This implies that an academic whose average publication under flat wage pay appeared in outlets like the Economic Journal - which had an impact factor of 1.440 in 2005 - publishes in outlets like the RAND journal of economics - which had an impact factor of 1.217 in 2005 - in response to the career concerns component of performance pay (Clarivate Analytics, 2000-2012*b*). The average number of citations decreases by a marginally significant 10% in response to the career concerns incentives (column 5 in Table 1 Panel B). This decrease is equivalent to publications receiving on average 3.6 fewer citations six or more years after publication (column 5 in panel A in Table A.1). Additional analyses show that there is no significant decrease in either the maximum or minimum number of citations to the publications of treated academics (Panel A, Table A.2).

To get a better idea of the quality of the work produced in response to performance pay, I analyze the distribution of citations and impact factors. I calculate the percentiles of citations and impact factor ratings separately by field and publication year and use the percentile cut-offs to generate quantile frequency variables for each author and publication year. To illustrate, suppose an author has three publications in a given year, one of which garners a number of citations that puts it in the top citation quartile of publications in the same field and publication year, while the other two papers fall in the bottom citation quartile. Then the top quartile frequency variable is equal to 1, the bottom quartile frequency variable 2, and other quartile frequency variables 0. I estimate the baseline model separately for all quantile frequency variables. The histograms in Figure 1a depict the resulting $Post'02*Treatment_i$ (grey bars) and $Tenure*Treatment_i$ (white bars) coefficient estimates and 95% confidence intervals. These figures clearly show that treated academics produce more of low to medium quality work, but not more of the highest quality research in response to (the career concerns incentives of) performance pay.

In short, I find evidence of a positive and significant average effort effect of performance incentives on the total raw and quality-adjusted quantity of *knowledge output*. This is a response to the *implicit, career concerns* component of performance pay, with no additional response to the explicit performance incentives component. The effect size ranges from 14 to 18%. This 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 (e.g. Lazear (2000); Shearer (2004)). However, the extra output produced is not of the highest quality, as only publications in low to medium citation and impact-factor quartiles increases. Indeed, there is a significant decrease in the average quality of knowledge output of around 9 to 10%.

The absence of a significant response to the explicit incentive component of performance pay can be ex-

[30]The difference in the number of observations across columns in this and further tables occurs for two reasons. First, in years in which an academic does not have any publications, the quantity and quality-adjusted quantity variables have zero entries, while observations for the average quality measures are missing. Second, in order to estimate the Poisson model, I have to drop authors for whom all observations are 0, or for whom I have only one observation.

plained by the fact that the explicit incentives are much lower-powered than the career concerns incentives in the German performance pay scheme. Gibbons and Murphy (1992) show that this is also what we would expect to see more generally for agents at the relative beginning of their career, if time horizons are finite.

<Figure 1 about here>

4.2.4 Validity of Identifying Assumptions

Interpretation of the above results as the causal effort effect of the higher-powered incentives of performance pay requires the parallel trends assumption to be met, and the timing of the start of the first tenured appointment to be exogenous. There are a number of potential threats to identification. The results could be driven by other events that occur around the time of the pay reform. There may also be selection into treatment: the timing of the first tenured appointment could be endogenous, there may be selection of higher productivity academics into the treated cohort, and academics in the control cohort could switch to the performance pay scheme and thus become treated after implementation of the reform. I adress each of these concerns below.

Around the time of the pay reform, a number of other events took place: the roll-out of a large nation-wide funding initiative for universities and research centers (the "Excellence Initiative") as of late 2006/early 2007 (DFG, 2016), the abolition of the professor's privilege in 2002[31] (Von Proff, Buenstorf and Hummel, 2012), and the introduction of the "Junior Professorship" in 2002 as an alternative path to professorships from the habilitation (Lutter and Schröder, 2016). The latter cannot be driving the results, because the first Junior Professors became eligible for a tenured position by 2008/9 and thus cannot be part of the treatment or control cohort. The abolition of the professor's privilege applied to all professors at the same time, and hence cannot explain the *differential* productivity changes between the treated and control cohort either. The Excellence Initiative funded a select set of universities, graduate schools and 'clusters of excellence' and thus does not affect the treatment and control cohorts differentially either. Besides, the initiative was announced in mid-2005. For both reasons, the initiative cannot give rise to the differential productivity changes between the treated and control cohort as of 2002. Moreover, results are robust to using 2-year treatment and control cohorts (Table A.1 Panel G). Given that the first funding decisions in the initiative were made late 2006, the restricted window excludes professors appointed after the first funding was distributed.

To provide further evidence that the results are driven by the introduction of the higher powered performance incentives introduced with the pay reform, I estimate the following conditional QML Poisson fixed effects model of dynamic productivity differences between the treated and control cohort:

$$E\left[Y_{i,f,t-x_f}|X_{i,f,t}\right] = exp[\alpha_i + \sum_{k=1}^{15}\beta_k ttt_{i,k-8} * Treatment_i + \sum_{j=-7}^{7} ttt_{i,j} + \gamma_t] \qquad (7)$$

Here, $ttt_{i,k-8} * Treatment_i$ denote interactions of the 15 time-to-tenure dummies with the treatment variable. All other variables are as before. This specification effectively aligns the relative time (time-to-

[31] Under the professor's privilege regime, professors owned the IPR of their inventions (Hvide and Jones, 2018). The abolition of this privilege should reduce incentives to produce commercializable (patentable) knowledge.

tenure) for different tenure cohorts, and allows me to estimate differences in output between the treated and control cohorts as they move towards and beyond the start of their first tenured affiliation. The coefficient estimates and 95% confidence intervals of the interaction terms are depicted in the left-hand side figures in Figure 2. The dashed line at $t-5$ indicates where in the tenure trajectory of the youngest academics of the treated cohort (those who start their first tenured affiliation in 2007) the announcement of the pay reform occurs and career concern incentives take effect. The dashed line at $tenure$ marks the time at which treated academics enter into the performance pay scheme and explicit incentives take effect.

All five figures display a similar pattern: the output metrics of the treated and control cohorts start to diverge after the announcement of the reform, when the treated cohort faces higher-powered, career concerns incentives, with no additional response to the explicit performance incentives. Since publications have been backdated by average publication lags, the differential increase in the number of publications directly after the announcement of the reform is consistent with an immediate effort response to the career concerns incentives. The response in quality-adjusted output and average quality occurs a bit later and more gradually, as expected if producing high quality research takes more time and is riskier. The figures also underline that the effort response to performance pay is not temporary, but persistent.

The clear alignment of the productivity response with the time at which performance incentives take effect supports the interpretation of the baseline results as the causal effort effect of performance pay, rather than another event. The absence of pre-existing trends lends support to the parallel trends assumption.

<Figure 2 about here>

Next, I address concerns about selection into treatment, starting with an assessment of the validity of the assumption that the timing of the start of the first tenured affiliation is exogenous. As a first test, I estimate the effort effects with synthetic treatment and control cohorts. To do so, I assign academics to the treatment and control cohort on the basis of the average age at which academics start their first tenured affiliation rather than the actual timing of the start of the first tenured affiliation. This assignment is orthogonal to any potential efforts to affect the timing of the start of the first tenured position. Reassuringly, results are robust to this alternative assignment to treatment and control (Table 2 Panel A). Online appendix section A4.1 provides further details.

<Table 2 about here>

As a second test, I estimate the dynamic productivity differences between a placebo treatment and control cohort. The placebo control cohort comprises academics who start their first tenured affiliation in 2001 or 2002, while those that start their first tenured affiliation directly before implementation of the performance pay scheme (2003-4) form the placebo treatment cohort. Both academics are appointed under the age-related pay scheme and thus face the same incentives. But if the latter cohort managed to "avoid" the performance pay scheme by publishing more and thereby moving up the start of the first tenured affiliation, there should be a (temporary) increase in their research output just before tenure (possibly coupled by a temporary drop after tenure). The absence of significant productivity differences in the right-hand side figures of Figure 2 show there is no evidence of this.

Note further that high productivity academics have no incentive to delay the start of their first tenured affiliation to after 2005, so as to fall under performance pay. Doing so would delay obtaining a permanent position with higher earnings, while they can always switch to performance pay from age-related pay. I discuss such switches further below.

A second selection concern is potential selection of higher productivity academics into the treated cohort. This could arise if tenure requirements were increased or universities delayed appointments so as to hire professors in the performance pay system. Apart from the synthetic cohorts test described above to address this concern, I also estimate the tenure probability using hazard rate analysis. As shown in Table 3, I find no evidence that more productive academics are more likely to obtain a tenured position after implementation of the reform: the $Post'05 * Productivity$ coefficients are all very close to zero and not significant[32]. Hence there is no evidence that the treated cohort comprises higher productivity academics. Online appendix section A4.3 provides further details.

<Table 3 about here>

A third selection concern pertains to academics in the control cohort selecting into performance pay. 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, renegotiating their contract, or by opting into the performance pay scheme while retaining the same position[33]. Academics in the control cohort could therefore end up being treated as well. The chances of this happening are, however, low. Academics appointed to a tenured position before 2005 have a 1.4% chance, on average, of switching to performance pay in any year after 2004 (cf. section 4.3.1). Moreover, any effort response of the control cohort would lead me to *underestimate* the effort effect of the treated cohort. If anything, the baseline results thus provide a conservative estimate of the effort effect. To test this, I re-estimate the baseline specification with a control group that excludes switchers, by labeling any academic who changes affiliation or position, or renegotiates their contract after implementation of the pay reform as a switcher (see section 4.1 for details). The effort effect estimates for output quantity and quality-adjusted quantity are indeed larger, ranging from a 17% increase in citations to a 23% increase in the number of publications (Table 2 Panel B), while the estimates of the effects on average quality are qualitatively the same (Table A.4 Panel B).

Finally, an instrumental variables estimation, instrumenting for selection into performance pay - and thus treatment status - with age-related variables and age cut-offs that align with the single crossing points of the basic wage schedules, yield qualitatively similar results (Online appendix section A4.2).

4.2.5 Robustness

There is no evidence that the documented changes are the result of strategic co-authorship behavior. The average number of co-authors on papers does not increase (Table A.2, Panel B) and the results of the baseline regression with dependent variables weighted by number of authors (so a paper with three authors

[32] The $Post'05$ coefficient is positive and significant as it picks up on any average change in tenure probability over time, since it is the only time-related variable included in the specification.

[33] Preissler (2006) reports that only a small number of professors chose to opt into the W-pay scheme in their current position.

counts for one-third) are very similar to the baseline results (Table 2, Panel C). Papers also do not become significantly shorter (Table A.2, Panel B).

Table A.1 shows that the effort effect results are also robust to restricting attention to articles and proceedings papers only[34] (Panel B); widening or narrowing the treatment and control cohort windows (Panels F and G); 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); transforming the dependent variables using the inverse hyperbolic sine transform and estimating as a panel fixed effects model (Panel C); including other sets of year-to-tenure dummies (Table A.1 Panel H), or including a non-linear function of relative time - the absolute value of time-to-tenure (Table A.1 Panel D).

4.2.6 Heterogeneous responses by academic field

The baseline results show how performance pay affects research productivity on average. Effort responses may differ across academic fields, however, for a number of reasons. For one, research teams in the natural and applied sciences tend to be much larger than those in the social sciences and humanities.[35] 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. On the other hand, the benefit of good management may have stronger effects in large teams. Highly incentivized team leaders may thus be able to bring about larger changes in output in large teams. Furthermore, fields may differ in the noise in output quality measures. To the extent that quality measures in the natural and applied sciences are more objective than in the social sciences and humanities, quality measures may be more noisy in the latter. The theory predicts that more ability classes decrease quality effort relative to flat wage pay when the quality measure is noisier, provided risk aversion is sufficiently high.

To study heterogeneous effort responses by academic field, I estimate the baseline regression (6) separately by broad academic field: natural and applied sciences, and social sciences and humanities. I classify mathematics, physics and informatics, biology, chemistry, earth sciences, pharmacology, engineering, medicine, dentistry, veterinary, agricultural science and nutrition science as *natural and applied sciences*; and theology, philosophy and history, philology and anthropology, law, economics and other social sciences as *social sciences and humanities*. Figure 3a depicts the estimation results of the baseline regression for these broad fields separately.

Quantity effort increases in both broad fields in response to the career concerns incentives of performance pay, by 17% in the natural and applied sciences and 31% in the social sciences and humanities. The difference between these responses is not statistically significant in a pooled regression with interaction terms with broad field indicators (Table A.3). Yet, while quality-adjusted measures of output increase significantly in the natural and applied sciences, there is no significant increase in these same measures in the social sciences and humanities (Figure 3a). Furthermore, the average impact factor-rating of publications decreases less in the natural and applied sciences in the pooled regression, though this difference is only marginally significant (Table A.3). Thus, while I find no evidence of a significant difference in the level of

[34] Specifically, I restrict attention to publications in the following ISI web of science categories only: “Article”, “Article: Book”, “Article: Book Chapter”, “Article: Proceedings Paper”, “Proceedings Paper”.

[35] The average team size is around 17 in the natural and applied sciences (median is 5), and only 3 in social sciences and humanities (median is 2) in the data.

the quantity effort response, there is suggestive evidence that the quality effort response is more negative in the social sciences and humanities. This would align with quality measures being more noisy in the latter.

<Figure 3 about here>

4.2.7 Heterogeneous responses by productivity quantile

Next, I analyze the quantity and quality effort response for different productivity quantiles. I determine productivity quantiles separately by academic field and treatment group on the basis of the averages of the impact factor-rated number of publications published in 1999, 2000 and 2001. I use pre-announcement averages to avoid simultaneity bias. Because the productivity distributions are highly right-skewed, with the median academic having no publications in an average year (cf. Table 1), I look at above-median academics separately by decile and below-median academics as one group. For the same reason, the 6th decile is significantly smaller than higher ones, so results for this decile should be taken to be indicative at best. The histograms in Figure 3b depict the $Post'02 * Treatment_i$ (grey bars) and $Tenure * Treatment_i$ (white bars) coefficient estimates and 95% confidence intervals of separate baseline regressions for each quantile.

Both low and high productivity academics increase pure quantity and quality-adjusted quantity in response to performance pay. Relative to the same quantile in the control group, treated below-median productive academics increase their number of publications by 24% and impact factor-rated number of publications by 31% in response to career concerns incentives. The top decile and 7th decile also increase the number of publications, by 22% and 24% respectively, as well as the sum of citations to publications, by 31% and 39%. There is no significant response in the 9th, 8th and 6th decile, though the lack of statistical significance in the 6th decile is likely due to its aforementioned small size.

The quantity effort response constitutes a positive and significant intensive margin response for top decile academics only (Fig A.1). Conditional on having at least one publication in a given year, their number of publications increases by 20% and citations to publications by 30%. The effort response of lower deciles is solely an extensive margin response. The probability that an academic has at least one publication in a given year increases for below-median academics as well as in all but the highest two above-median deciles.

To test whether the effort response differs across quantiles, I estimate the baseline regression model (6) augmented with interaction terms with indicator variables for the top five deciles. Table A.4 Panel A shows that the differences in effort responses are, in fact, significant. The coefficient of $post'02 * Treatment$ in column 1 implies that below-median productive academics in the treatment group produce 32% more publications in response to career concern incentives, relative to the same quantile in the control group. The triple interactions of $post'02$, $Treatment$ and indicator variables for the 10th and 9th decile imply that the effort response of the top deciles of the treated group are 14% and 28% less than the effort response of below-median treated academics (significant at the 10% and 1% level, respectively). Moreover, simple Wald tests for equality of the $post'02*Treatment*decile$ interactions show that the 9th decile interaction is significantly different from all other interactions, so the quantity effort response of this decile is less positive than that of both higher and lower productivity quantiles. A similar pattern emerges for the impact factor-rated number of publications. For the sum of citations, the effort response in both the 8th and 9th decile is significantly less positive than in all other quantiles. Results are robust to excluding academics who switch to performance pay after 2005 (Table A.4 Panel B), so there is no evidence that differential selection into performance pay is driving these heterogeneous treatment effects.[36]

Taken together, these results show that the pure quantity and quality-adjusted quantity effort response is U-shaped in productivity or ability class. This aligns with the theoretical prediction for the case when risk aversion is sufficiently high.

Turning to the quality effort response, I find that academics in the sub-top productivity deciles decrease quality effort, while quality remains unchanged in lower and higher productivity quantiles. Figure 3b shows that the average number of citations decreases significantly in the 8th decile in response to career concerns incentives, and the intensive margin response in the sum of citations is significantly negative for the 9th decile (Fig A.1d). Indeed, the response histograms in Figure 4 show that both these sub-top productivity deciles produce fewer papers in the top citation decile - a reduction of 25% and 32% respectively - in response to the career concerns incentives of performance pay. Hence, there is crowding out of quality effort in medium productive academics in response to performance pay.

There is no evidence that other productivity deciles change quality effort. There are increases in the lowest citation quartile bin for below-median productive academics and in the second and third quartile citation bins for the most productive academics, in line with their respective quantity effort response and commensurate with their ability class (Figure 4). But since there is no sign of substitution of higher citation decile bin papers for lower decile bins or vice versa, there is no evidence of changes in output quality for these productivity classes.

Taken together, the quality effort response also appears to be U-shaped in ability, with no significant increase at the top of the productivity distribution, a significant decline in quality in intermediate ability classes and no significant change in the lowest ability classes. This aligns with the theoretical prediction for the case when there is quality crowding out for some (lower) ability academics and tenure requirements increase with ability class. The absence of quality crowding in the lowest ability classes can be explained by quality effort at the bottom of the productivity distribution already being at the lowest possible level (average impact factor-rated number of publications is zero) before the introduction of performance pay, so that no further decrease is possible.

All in all, the heterogeneous responses explain the average effort effect, and in particular the decrease in average output quality. The largest quantity effort response comes from the least productive academics, who do not produce the highest quality work. The most productive academics also publish more in response to performance pay, but they do not increase quality effort and do not produce more of the highest quality publications. Medium productive academics, finally, do not increase quantity effort and, in fact, publish fewer of the highest quality papers.

<Figure 4 about here>

[36] Results are also robust to using deciles based on pre-announcement averages of the sum of citations (available from author).

4.2.8 Novelty and Impact

To provide further evidence on the quality and impact of the work produced in response to performance incentives, I perform textual analysis of paper abstracts. Specifically, I compare the similarity of abstracts of focal papers to abstracts of papers published before or after the focal paper, to gauge the novelty and impact of the focal paper. To construct similarity metrics, I represent each paper by a vector of terms used in the abstract, weighing those terms by their relative importance in capturing the paper's content (Gentzkow, Kelly and Taddy, 2019). To this end, I base similarity metrics on a leading metric from the textual analysis literature, the Term Frequency Inverse Document Frequency (TFIDF), as in Biasi and Ma (2022) and Kelly et al. (2021). This metric gives greater weight to terms that occur more frequently in a document, but less so if the term is very common in a set of comparison documents. In particular, the Term Frequency Inverse Document Frequency (TFIDF) index is defined as:

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

where $c_{w,d}$ denotes the count of a term w in document d (and equivalently for $c_{l,d}$) and $c_{d,\tau}$ denotes the count of documents that contain the term w and that were published in period τ. The first element of the expression is the frequency of term w in document d (the "Term Frequency" part). The second element is the log of the inverse of the frequency of documents in which term w appears (the "Inverse Document Frequency" part), which is smaller for more commonly used - and therefore less informative - terms.

As in Kelly et al. (2021), 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, to allow for some dependence between words and, thus, richer modeling (Gentzkow, Kelly and Taddy, 2019).[37] Before calculating TFIDFs, I drop common stop words from abstracts.

In order to measure the similarity of the bigrams 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 year t, the cosine similarity with comparison document $\tilde{d}$ published in year $\tilde{t}$ is:

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

For each paper pair, I calculate the inverse document frequency based on the set of all papers in the same field published in all years prior to the earlier of the two publication dates $(t,\tilde{t})$[38]. 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 similarity is to 1.

As in Kelly et al. (2021), I calculate two different cosine similarities: backward and forward similarity.

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

[38] Due to the way in which publication records were extracted from ISI Web of Science, namely, filtering by publications that have at least one author with a German (work) address, the set of comparison papers are from the same field, as well as the same country.

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. 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. A smaller backward similarity indicates that the focal publication contains more unique or rare bigrams, which may imply greater novelty. The forward similarity metric on the other hand captures the relatedness of follow-on research and, as such, constitutes an alternative measure of impact to citations. Indeed, Kelly et al. (2021) find that patents with a higher ratio of forward to backward similarity are more likely highly cited and have higher market value. This lends support to the notion that the TFIDF cosine similarity measures capture novelty and impact.[39]

Because the analysis in this paper is at the individual academic level rather than the publication or patent level, as is the case in (Kelly et al., 2021), and because publications are matched to academics on the basis of i.a. name and field, I deviate from the metrics in the latter paper in two ways. First, I restrict the set of comparison publications to the same field as the focal publication. Second, for each similarity measure, I calculate quantile frequency bins (in the same way as for citations above) to analyze the effect of performance pay on the distribution of the similarity of papers to past and future work.

The novelty metric analyses confirm the earlier quality response results: the additional work produced in response to performance pay is, on average, not the most novel or impactful. Panel A in Figure 1b shows that there is a significant increase in the frequency of top quartile backward cosine similarity papers as well as a marginally significant increase in the bottom decile and quartile bins. Thus, in response to performance pay, more papers that are very similar to previously published papers are produced, though there is also a slight increase in papers that are very dissimilar to prior work. At the same time, there is an increase in papers that give rise to considerable - but not the most - follow-on research and/or are part of a burgeoning literature (See third quartile bin of forward similarity metrics in Figure 1b Panel B).

Breaking down the novelty metric analysis by productivity quantiles provides further insight into the earlier heterogeneous quality response results. Top productivity academics produce more low to medium novel work that garners mid- to high-level follow-on. Low productivity academics produce additional papers that are just above the median in terms of both similarity to past and future work. Sub-top productivity academics finally produce more papers that are very novel, but also garner only very little follow-on work (Figures 5a and 5b).

<Figure 5 about here>

4.3 Selection Effect

I now turn attention to another channel through which performance pay can affect output: selection. If selection is positive, more productive academics are more likely to select into performance pay. As

[39] Relative to similarity metrics based on Latent Semantic Analysis (LSA) (Iaria, Schwarz and Waldinger, 2018), TFIDF is more likely to understate similarity, since it does not 'learn' about similar topics. It is therefore more likely to provide more conservative estimates of decreases in novelty and increases in impact, which is why it is used here.

shown in the theoretical section, selection into academia should follow the same pattern as selection into performance pay, since the latter drives the former (Online appendix section A1.3.3).

4.3.1 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 implementation of the reform in 2005. These academics are initially paid according to the age-related pay scheme and therefore have the choice of switching pay scheme. They do so by changing affiliation or position, or renegotiating their contract. Accordingly, I label any such event after implementation of the pay reform as a switch to performance pay (the "failure" event). As described in section 4.1, changes in affiliation or position are read from the affiliations panel, while FuL announcements about offer rejections or extensions only are recorded as contract renegotiations. The time since an academic's most recent contract renegotiation, affiliation or position change before the reform implementation is the "at risk" duration. I restrict attention to academics who start their first tenured affiliation after 1998[40], so that I observe their full tenured affiliation history, and hence the moment they become "at risk" of switching.[41] There are 11237 such academics, and I observe 1231 switches in a total of 85716 "at risk" periods (years). The average incidence rate of switches is 0.014, so academics in this group have a 1.4% chance, on average, of switching to performance pay in any year after 2004.

Figure 6b 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. As before, I base productivity quartiles on the average of the impact factor-rated number of publications in the three pre-implementation years (2002-04), calculating quartiles separately by academic field and tenure cohort. The hazard rate for switching to the performance pay scheme is greater for top quartile academics throughout. A log-rank test rejects equality of the survival functions at the 1% significance level. More productive academics are thus more likely to sort into performance pay.

<Figure 6 about here>

4.3.2 Difference-in-Differences Estimation

More productive academics may be more likely to change position or affiliation, or renegotiate their contract in general, since they are more likely to receive (outside) offers. In what follows, I will refer to contract renegotiations, affiliation and position changes collectively as "contract changes" for brevity. To distinguish general patterns in contract changes from the selection effect of performance pay, I estimate the selection effect in a difference-in-differences framework, exploiting variation along the tenure cohort and age dimension.

Any professor can (potentially) change contract, but only academics who hold a tenured affiliation before implementation of the pay reform switch from age-related to performance pay when such a change occurs

[40] As in the effort effect analysis, I also restrict the sample to academics that do not have a foreign affiliation before their first tenured affiliation, since I do not have full affiliation and publication records for academics from abroad.

[41] Including academics who enter into the data set in a tenured position would introduce left truncation into the survival analysis sample, since I would not know their risk origin date and hence risk duration. This could bias the analysis.

after 2004. Accordingly, I define academics who start their first tenured position before 2005 as treated cohort, while academics who make tenure after 2004 comprise the control cohort here. The event of interest is the first contract change after 2004 for the treated cohort, and the first contract change after appointment to the first tenured position for the control cohort. If selection into performance pay is positive, high productivity academics in the treated cohort should be more likely to change contract than high productivity academics in the control cohort. A simple difference in contract changing rates between the two cohorts could however be driven by factors other than the fact that only the treated cohort switches to performance pay when they change contract after 2004. I therefore also exploit variation in contract changing incentives along a second dimension: age.

Switching pay schemes is worthwhile only if the difference in expected utility net of basic wages between the performance pay and age-related pay scheme exceeds any difference in basic wages between the pay schemes. The difference in basic wages is relatively larger for older academics, because of the single-crossing property of the basic wage schemes for age-related and performance pay (Cf. Figure 6a). Older academics thus have weaker incentives to switch to performance pay, over and above any general differences across age groups in incentives to change contract (e.g. due to mobility differences). But then, if selection into performance pay is positive, older academics need to be relatively more productive to want to switch. Compared to academics in the control group, the risk of changing contract should therefore increase relatively less with productivity for older academics in the treated cohort if selection is positive.

I estimate the following Weibull proportional hazard model to test this:

$$\begin{aligned}\lambda_{i,t} = \rho * exp[\beta_0 + \beta_1 Treat_i + \beta_1 Age_{i,t} + \beta_2 AvgProd_i + \beta_3 Age_{i,t} * Treat_i + \beta_4 AvgProd_i * Treat_i \\ + \beta_1 AvgProd_i * Age_{i,t} + \beta_2 AvgProd_i * Age_{i,t} * Treat_i + X_i + u_{i,t}] * t^{\rho-1} \quad (8)\end{aligned}$$

Here $\lambda_{i,t}$ is the hazard (risk) of changing contract for academic i at time t and $Treat$ is a treatment group indicator.[42] $AvgProd$ is an academic's average productivity calculated as three-year pre-implementation averages (2002-2004) of the impact factor-rated number of publications. Age 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 after 1998 and do not hold a foreign position immediately prior. My preferred specification also controls for synthetic age at the start of the first tenured affiliation, to control for differences in career paths. Standard errors are robust and clustered by individual academic. Columns a and b in Panel A of Table 4 report estimation results of specifications without and with age-at-tenure as additional control, respectively.

The positive and significant coefficient estimates of $AvgProd * Treat$ in column 1 imply that a one standard deviation increase in the average productivity of treated academics increases the rate at which they change contract by 10.8%[43] to 20% more than academics in the control group. This holds while controlling for $Age * Treat$, which, as expected, is negative. The coefficient estimates of $AvgProd_i * Age_{i,t}$ and

[42] Risk duration is measured as time since last contract change (including start of first tenured affiliation) pre-2005 for the treated cohort, and start of first tenured affiliation for the control cohort.

[43] Calculated as EXP(0.004*25.53797)-1, where 25.53797 is the standard deviation of the average productivity variable here.

$AvgProd_i * Age_{i,t} * Treat_i$ in column 2 imply that a one standard deviation increase in average productivity reduces the negative effect of age on switching rates by 2.6% on average, but this moderating effect of productivity is 2.6% less for treated academics. That is, treated academics need to be of higher productivity in order to change contract when they are older. The selection effect of performance pay, net of general differential sorting patterns by age and productivity types, is thus positive and significant.

The finding that performance pay attracts more productive academics is robust to estimating the model as a Cox proportional hazard model or estimating the Weibull model with academic field strata (Columns 2-3 in Table A.5)[44]. Replacing the $AvgProd$ variable in the baseline Weibull model (8) by a dummy indicating whether an academic's pre-reform average productivity is above median shows that having above median productivity reduces the negative effect on switching rates of an extra year of age by 5.8% on average (Column 1 in Table A.5). However, this moderating effect of above median productivity is 3.8% less in treated academics, so treated academics need to be even more productive to change contract.

<Table 4 about here>

4.3.3 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 contract after 2004, so they face the same selection incentives. The estimation results of the baseline placebo estimation are reported in Panel B of Table 4 Reassuringly, neither the $AvgProd * Treat$ interaction nor the $AvgProd_i * Age_{i,t} * Treat_i$ triple interaction is ever 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.

As a final check, I also estimate any changes in contract changing rates from before to after the implementation of the reform for academics in the treated cohort (those whose first tenured affiliation started before 2005). While a one standard deviation increase in average productivity increases the likelihood of a contract change by around 5.5% on average, the increase in the likelihood of a change grows to 8 or 9% after the implementation of the reform (Table A.6). This consolidates the finding of a positive selection effect of performance pay.

5 CONCLUSION

This paper provides causal evidence that performance pay incentives commonly found in knowledge creation jobs – particularly ones involving career concerns incentives – attract more productive workers and significantly increase knowledge output quantity. However, they do not yield more output of the highest quality, and average quality decreases. This is because both the output quantity and quality response are U-shaped in ability. The least productive academics increase output quantity the most, but the quality of their output is low. The most productive academics also produce more output, but not more of the highest

[44] Results are robust to basing the average productivity variable on other productivity measures as well (results available on request).

quality (most novel and impactful). Medium productive academics do not increase output quantity, but they do produce less high quality work, suggesting there is crowding out of their output quality.

The paper thus documents multitasking behavior in response to performance incentives for knowledge creation. Such behavior may be particularly likely to arise in the context of knowledge creation for two reasons. For one, knowledge work comprises multiple complex tasks, with multi-dimensional effort and output and noisy output measures. Since Holmström and Milgrom (1991), multitasking problems have been predicted to arise in such settings. Apart from providing empirical evidence of this, the paper also shows theoretically that, relative to explicit incentives only, the introduction of career concerns incentives in these settings decreases the likelihood of positive selection effects as well as positive effort effects along less precisely measured dimensions. Given that career concerns incentives are common in knowledge work, as well as professional and managerial jobs, multitasking problems may thus be particularly likely in these contexts.

This suggests that it would be valuable to develop and use more informative measures of novelty, impact or other quality dimensions. This paper employs one potential measure; the TFIDF-based backward and forward cosine similarity of research output. With the advent of ever more powerful machine learning algorithms, the availability and precision of relevant performance metrics for knowledge creation could likely be improved. This seems a worthwhile avenue for research, not just in an academic context, but knowledge production in general.

When it comes to the effects of high-powered performance incentives in academia, it is an open question whether the benefits outweigh the costs. The paper shows that performance incentives significantly increase effort towards research quantity. With limited time and hence effort, the question is where this effort comes from, and whether the benefit of the redirection of effort exceeds its costs, especially since the additional output is not the most novel or impactful. It could be that some of the time dedicated to producing research quantity was previously enjoyed as leisure or spent providing consulting services. Or perhaps the additional quantity effort comes at the cost of teaching effort. In the latter case, whether teaching quality declines or improves will depend on whether research and teaching effort are substitites or complements. It is beyond the scope of the current paper to estimate these (additional) costs and benefits and this is left for future research.

Academic research is an important instance of knowledge work, and understanding the effect of performance pay on both effort and selection in this particular context is valuable in its own right. The types of performance incentives that are prevalent in academia, are common in many other knowledge work sectors as well, and the academic production function shares key features with that of other knowledge work. Academia is therefore also a useful setting in which to study the organization of knowledge work more generally. As such, the findings in this paper have implications for knowledge work in other contexts. Some characteristics of the academic environment may not be present to the same degree in other contexts, however. The knowledge created in academia is highly visible and available to a broad audience, and academics are (expected to be) highly mobile. Both of these characteristics are conducive to career concerns. In sectors in which these conditions are met to a lesser degree, the effects of performance incentives, specifically those including career concerns incentives, may thus not be as strong. It would therefore be valuable to study how the publicness of output and the mobility of agents affects the effects of performance incentives in knowledge creation. This is left for future research.

Acknowledgements: I would like to thank Laurence Ales, Pierre Azoulay, Oriana Bandiera, Tim Besley, Jordi Blanes-i-Vidal, Kenneth Corts, Pablo Casas-Arce, Baran Duzce, Florian Englmaier, Jeff Furman, Maitreesh Ghatak, Bob Gibbons, De Gruyter, Dietmar Harhoff, Rosario Macera, Bentley Macleod, Michal Matejka, Bob Miller, Steve Pischke, Andrea Prat, Carol Propper, John van Reenen, Mark Schankerman, Axel Schniederjuergen, Ananya Sen, Chris Stanton, Scott Stern, Neil Thompson, Fabian Waldinger, the ministries of education of the German states and seminar participants at Carnegie Mellon Tepper, Chicago Booth, Copenhagen Business School, ETH Zurich, Harvard Business School, MIT TIES, MIT IDE, MIT OE Lunch, MPI Munich, UANDES, Universidad Carlos III, Universite Laval and the University of Utah for helpful comments, information or data.

References

Academics.de. 2016. "Tenure Track: Professor Auf Lebenszeit."

Agarwal, Ruchir, and Patrick Gaule. 2020. "Invisible Geniuses: Could the Knowledge Frontier Advance Faster?" *American Economic Review: Insights*, 2(4): 409–24.

Aghion, Philippe, John Van Reenen, and Luigi Zingales. 2013. "Innovation and Institutional Ownership." *The American Economic Review*, 103(1): 277–304.

Aghion, Philippe, Mathias Dewatripont, Caroline Hoxby, Andreu Mas-Colell, and André Sapir. 2010. "The Governance and Performance of Universities: Evidence from Europe and the Us." *Economic Policy*, 25(61): 7–59.

Aghion, Philippe, Peter Howitt, and David Mayer-Foulkes. 2005. "The Effect of Financial Development on Convergence: Theory and Evidence." *The Quarterly Journal of Economics*, 120(1): 173–222.

Alexander, Diane. 2020. "How Do Doctors Respond to Incentives? Unintended Consequences of Paying Doctors to Reduce Costs." *Journal of Political Economy*, 128(11): 4046–4096.

Andrabi, Tahir, and Christina Brown. 2022. "Subjective Versus Objective Incentives and Employee Productivity." *Working Paper*.

Athey, Susan, and Guido W Imbens. 2022. "Design-based Analysis in Difference-in-differences Settings with Staggered Adoption." *Journal of Econometrics*, 226(1): 62–79.

Autor, David H. 2019. "Work of the Past, Work of the Future." 109: 1–32.

Azoulay, Pierre, Christian Fons-Rosen, and Joshua S Graff Zivin. 2019. "Does Science Advance One Funeral at a Time?" *American Economic Review*, 109(8): 2889–2920.

Azoulay, Pierre, Jeffrey L Furman, Joshua L Krieger, and Fiona Murray. 2015. "Retractions." *Review of Economics and Statistics*, 97(5): 1118–1136.

Azoulay, Pierre, Joshua S Graff Zivin, and Gustavo Manso. 2011. "Incentives and Creativity: Evidence from the Academic Life Sciences." *The Rand Journal of Economics*, 42(3): 527–554.

Azoulay, Pierre, Waverly Ding, and Toby Stuart. 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.

Bandiera, Oriana, Iwan Barankay, and Imran Rasul. 2005. "Social Preferences and the Response to Incentives: Evidence from Personnel Data." *The Quarterly Journal of Economics*, 917–962.

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

Bénabou, Roland, and Jean Tirole. 2006. "Incentives and Prosocial Behavior." *American Economic Review*, 96(5): 1652–1678.

Bergsdorf, Wolfgang. 2005. "Richtlinie Der Universitaet Erfurt Ueber Das Verfahren Und Die Vergabe Von Leistungsbezuegen."

Besley, Timothy, and Maitreesh Ghatak. 2005. "Competition and Incentives with Motivated Agents." *American Economic Review*, 95(3): 616–636.

Besley, Timothy, and Maitreesh Ghatak. 2018. "Prosocial Motivation and Incentives." *Annual Review of Economics*, 10: 411–438.

Biasi, Barbara, and Song Ma. 2022. "The Education-innovation Gap." working paper, Yale.

Biester, Christoph. 2010. "Der Universitaere Metabolismus, Die Buerokratisierung Der Leistungsorientierten Verguetung in Der W-besoldung." *Online powerpoint*, Accessed on 6 May 2015.

Björk, Bo-Christer, and David Solomon. 2013. "The Publishing Delay in Scholarly Peer-reviewed Journals." *Journal of Informetrics*, 7(4): 914–923.

BMBF. 2002. "Gesetz Zur Reform Der Professorenbesoldung." ProfBesReformG.

BMI. 2007. "Bericht Zum Besoldungsrechtlichen Vergaberahmen Bei Der Professorenbesoldung Nach p. 35 Abs. 5 Bundesbesoldungsgesetz." Bundesministerium des Innern. Stand 29 February 2008.

Bonatti, Alessandro, and Johannes Hörner. 2017. "Career Concerns with Exponential Learning." *Theoretical Economics*, 12(1): 425–475.

Borjas, George J, and Kirk B Doran. 2015. "Prizes and Productivity How Winning the Fields Medal Affects Scientific Output." *Journal of Human Resources*, 50(3): 728–758.

Borusyak, Kirill, Xavier Jaravel, and Jann Spiess. 2021. "Revisiting Event Study Designs: Robust and Efficient Estimation." *Arxiv Preprint Arxiv:2108.12419*.

Boudreau, Kevin J, Karim R Lakhani, and Michael Menietti. 2016. "Performance Responses to Competition across Skill Levels in Rank-order Tournaments: Field Evidence and Implications for Tournament Design." *The Rand Journal of Economics*, 47(1): 140–165.

Boudreau, Kevin J, Nicola Lacetera, and Karim R Lakhani. 2011. "Incentives and Problem Uncertainty in Innovation Contests: An Empirical Analysis." *Management Science*, 57(5): 843–863.

Bundesgesetzblatt. 1985. "Bundesbeamtengesetz."

Callaway, Brantly, and Pedro HC SantAnna. 2021. "Difference-in-differences with Multiple Time Periods." *Journal of Econometrics*, 225(2): 200–230.

Camargo, Braz, Fabian Lange, and Elena Pastorino. 2021. "On the Role of Learning, Human Capital, and Performance Incentives for Wages." Working Paper.

Chan, Ho Fai, Bruno S Frey, Jana Gallus, and Benno Torgler. 2014. "Academic Honors and Performance." *Labour Economics*, 31: 188–204.

Chevalier, Judith, and Glenn Ellison. 1999. "Career Concerns of Mutual Fund Managers." *The Quarterly Journal of Economics*, 114(2): 389–432.

Clarivate Analytics. 1993-2012*a*. "Isi Web of Science."

Clarivate Analytics. 2000-2012*b*. "Journal Citation Report."

De Chaisemartin, Clément, and Xavier dHaultfoeuille. 2020. "Two-way Fixed Effects Estimators with Heterogeneous Treatment Effects." *American Economic Review*, 110(9): 2964–96.

De Groot, Morris H. 1970. *Optimal Statistical Decisions.* McGraw-Hill.

De Gruyter. 2006. "Kuerschners Deutscher Gelehrten Kalender." *cd-rom*.

De Gruyter. 2008. "Kuerschners Deutscher Gelehrten Kalender." *cd-rom*.

De Gruyter. 2013. "Kuerschners Deutscher Gelehrten Kalender Online."

Deshpande, Manasi, and Yue Li. 2019. "Who Is Screened Out? Application Costs and the Targeting of Disability Programs." *American Economic Journal: Economic Policy*, 11(4): 213–48.

Detmer, Hubert, and Ulrike Preissler. 2004. "Abenteuer W, Strategien, Risiken Und Chancen." , (6): 308–311.

Detmer, Hubert, and Ulrike Preissler. 2005. "Die Neue Professorenbesoldung, Ein Ueberblick." , (5): 256–258.

Detmer, Hubert, and Ulrike Preissler. 2006. "Die W-besoldung Und Ihre Anwendung in Den Bundeslaendern." *Beitraege Zur Hochschulforschung*, 28(2): 50–65.

Dewatripont, Mathias, Ian Jewitt, and Jean Tirole. 1999. "The Economics of Career Concerns, Part Ii: Application to Missions and Accountability of Government Agencies." *The Review of Economic Studies*, 66(1): 199–217.

DFG. 2016. "Excellence Initiative 2005-2017."

DHV. 1999-2013. "Forschung Und Lehre."

DHV. 2002. "Habilitationen Und Berufungen." *Forschung Und Lehre*, 41.

DHV. 2014. "Forschung Und Lehre - Wir Ueber Uns."

Dohmen, Thomas, and Armin Falk. 2011. "Performance Pay and Multidimensional Sorting: Productivity, Preferences, and Gender." *The American Economic Review*, 556–590.

Ederer, Florian, and Gustavo Manso. 2013. "Is Pay for Performance Detrimental to Innovation?" *Management Science*, 59(7): 1496–1513.

Erat, Sanjiv, and Uri Gneezy. 2016. "Incentives for Creativity." *Experimental Economics*, 19(2): 269–280.

Expertenkommission. 2000. "Bericht Der Expertenkommission - Reform Des Hochschuldienstrechts."

Fama, Eugene F. 1980. "Agency Problems and the Theory of the Firm." *Journal of Political Economy*, 88(2): 288–307.

Ferrer, Rosa. 2016. "The Effect of Lawyers' Career Concerns on Litigation." Working Paper.

Fitzenberger, Bernd, and Ute Schulze. 2014. "Up or Out: Research Incentives and Career Prospects of Postdocs in Germany." *German Economic Review*, 15(2): 287–328.

Flämig, Christian, Volker Grellert, Otto Kimminich, E-J Meusel, H-H Rupp, Dieter Scheven, Hermann Josef Schuster, and Friedrich Stenbock-Fermor. 2013. *Handbuch des Wissenschaftsrechts.* Springer-Verlag.

Gentzkow, Matthew, Bryan Kelly, and Matt Taddy. 2019. "Text As Data." *Journal of Economic Literature*, 57(3): 535–74.

Gibbons, Robert, and Kevin J Murphy. 1992. "Optimal Incentive Contracts in the Presence of Career Concerns: Theory and Evidence." *Journal of Political Economy*, 100(3): 468–505.

Gibbs, Michael, Susanne Neckermann, and Christoph Siemroth. 2017. "A Field Experiment in Motivating Employee Ideas." *Review of Economics and Statistics*, 99(4): 577–590.

Gien, Gabriele. 2017. "Satzung Der Katholischen Universitaet Eichstaett-ingolstadt Zur Regelung Des Verfahrens Der Bewertung Der Besonderen Leistungen Zur Vergabe Der Besonderen Leistungsbezuege."

Glewwe, Paul, Nauman Ilias, and Michael Kremer. 2010. "Teacher Incentives." *American Economic Journal: Applied Economics*, 2(3): 205–27.

Gourieroux, Christian, Alain Monfort, and Alain Trognon. 1984. "Pseudo Maximum Likelihood Methods: Applications to Poisson Models." *Econometrica: Journal of the Econometric Society*, 701–720.

Graff Zivin, Joshua, and Elizabeth Lyons. 2021. "The Effects of Prize Structures on Innovative Performance." 111: 577–81.

Gross, Daniel P. 2020. "Creativity under Fire: The Effects of Competition on Creative Production." *Review of Economics and Statistics*, 102(3): 583–599.

Gupta, Atul. 2021. "Impacts of Performance Pay for Hospitals: The Readmissions Reduction Program." *American Economic Review*, 111(4): 1241–83.

Haeck, Catherine, and Frank Verboven. 2012. "The Internal Economics of a University: Evidence from Personnel Data." *Journal of Labor Economics*, 30(3): 591–626.

Hall, Bronwyn H, and Dietmar Harhoff. 2012. "Recent Research on the Economics of Patents." *Annual Review of Economics*, 4(1): 541–565.

Handel, Kai Christian. 2005. *Die Umsetzung Der Professorenbesoldungsreform in Den Bundesländern.* CHE.

Harbring, Christine, Bernd Irlenbusch, and Matthias Kräkel. 2004. "Ökonomische Analyse Der Professorenbesoldungsreform in Deutschland." 197–219.

Harris, Milton, and Bengt Holmstrom. 1982. "A Theory of Wage Dynamics." *The Review of Economic Studies*, 49(3): 315–333.

Hellmann, Thomas, and Veikko Thiele. 2011. "Incentives and Innovation: A Multitasking Approach." *American Economic Journal: Microeconomics*, 3(1): 78–128.

Holmström, Bengt. 1999. "Managerial Incentive Problems: A Dynamic Perspective." *The Review of Economic Studies*, 66(1): 169–182.

Holmström, Bengt, and Paul Milgrom. 1991. "Multitask Principal-agent Analyses: Incentive Contracts, Asset Ownership, and Job Design." *Journal of Law, Economics and Organization*, 7: 24.

Hong, Harrison, Jeffrey D Kubik, and Amit Solomon. 2000. "Security Analysts' Career Concerns and Herding of Earnings Forecasts." *The Rand Journal of Economics*, 121–144.

HRK. 2014. "Hochschulkompass."

Huber, Bernd. 2005. "Richtlinien Der Ludwig-maximilians-universitaet Muenchen Zur Regelung Der Grundsaetze Fuer Die Vergabe Von Leistungsbezuegen."

Hvide, Hans K, and Benjamin F Jones. 2018. "University Innovation and the Professor's Privilege." *American Economic Review*, 108(7): 1860–98.

Iaria, Alessandro, Carlo Schwarz, and Fabian Waldinger. 2018. "Frontier Knowledge and Scientific Production: Evidence from the Collapse of International Science." *The Quarterly Journal of Economics*, 133(2): 927–991.

Jochheim, Linda. 2015. "Leistungsförderung durch Leistungszulagen? Effekte der W-Besoldung untersucht am Beispiel deutscher Universitätsprofessoren."

Jones, Benjamin 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.

Kelly, Bryan, Dimitris Papanikolaou, Amit Seru, and Matt Taddy. 2021. "Measuring Technological Innovation Over the Long Run." *American Economic Review: Insights*, 3(3): 303–20.

Kräkel, Matthias. 2006. "Zur Reform Der Professorenbesoldung in Deutschland." *Perspektiven Der Wirtschaftspolitik*, 7(1): 105–126.

Lach, Saul, and Mark Schankerman. 2004. "Royalty Sharing and Technology Licensing in Universities." *Journal of the European Economic Association*, 2(2-3): 252–264.

Lach, Saul, and Mark Schankerman. 2008. "Incentives and Invention in Universities." *The Rand Journal of Economics*, 39(2): 403–433.

Lazear, Edward P. 2000. "Performance Pay and Productivity." *American Economic Review*, 90(5): 1346–1361.

Lazear, Edward P, and Paul Oyer. 2012. "Personnel Economics." *The Handbook of Organizational Economics*, 479.

Leitungsgremium, Universitaet Augsburg. 2005. "Grundsaetze Der Universitaet Augsburg Fuer Die Vergabe Von Leistungsbezuegen."

Lerner, Josh, and Julie Wulf. 2007. "Innovation and Incentives: Evidence from Corporate R&d." *The Review of Economics and Statistics*, 89(4): 634–644.

Leuven, Edwin, Hessel Oosterbeek, Joep Sonnemans, and Bas Van Der Klaauw. 2011. "Incentives Versus Sorting in Tournaments: Evidence from a Field Experiment." *Journal of Labor Economics*, 29(3): 637–658.

Lucas, R. E. 1988. "On the Mechanisms of Economic Development." *Journal of Monetary Economics*, 22(1): 3–32.

Lünstroth, Pia. 2011. "Leistungslohn Und Kooptation-eine ökonomische Analyse Der Reform Der Professorenbesoldung." PhD diss. Universitaet Trier.

Lutter, Mark, and Martin Schröder. 2016. "Who Becomes a Tenured Professor, and Why? Panel Data Evidence from German Sociology, 1980–2013." *Research Policy*, 45(5): 999–1013.

Macleod, Bentley. 2022. *Beyond Price Theory.* MIT Press.

MacLeod, W Bentley, and Daniel Parent. 2000. "Job Characteristics and the Form of Compensation Form of Compensation." In *Research in Labor Economics*. Emerald Group Publishing Limited.

MacLeod, W Bentley, and Miguel Urquiola. 2021. "Why Does the United States Have the Best Research Universities? Incentives, Resources, and Virtuous Circles." *Journal of Economic Perspectives*, 35(1): 185–206.

McCormack, John, Carol Propper, and Sarah Smith. 2014. "Herding Cats? Management and University Performance." *The Economic Journal*, 124(578): F534–F564.

Miklós-Thal, Jeanine, and Hannes Ullrich. 2016. "Career Prospects and Effort Incentives: Evidence from Professional Soccer." *Management Science*, 62(6): 1645–1667.

Mohr, Joachim. 2007. "Professoren, Die Vertreibung Der Weisen." *Spiegel Online*.

Oeffentlicher Dienst. 2004. "Gesetz Ueber Die Erhoehung Von Dienst- Und Versorgungsbezuegen in Bund Und Laendern 2003/2004." Last accessed 11-08.2015.

Oyer, Paul, and Scott Schaefer. 2011. "Personnel Economics: Hiring and Incentives." *Handbook of Labor Economics*, 4: 1769–1823.

Preissler, U. 2006. "Zwischenbilanz Professorenbesoldung - Sichtweise Und Beratung Des Deutschen Hochschulverbandes."

Arbeitsgruppe Fortbildung im Sprecherkreis der Deutschen Universitaetskanzler.

Pritchard, Rosalind. 2006. "Trends in the Restructuring of German Universities." *Comparative Education Review*, 50(1): 90–112.

Romer, Paul M. 1986. "Increasing Returns and Long-run Growth." *The Journal of Political Economy*, 94(5): 1002–1037.

Rubin, Jared, Anya Samek, and Roman M Sheremeta. 2018. "Loss Aversion and the Quantity–quality Tradeoff." *Experimental Economics*, 21(2): 292–315.

Shearer, Bruce. 2004. "Piece Rates, Fixed Wages and Incentives: Evidence from a Field Experiment." *The Review of Economic Studies*, 71(2): 513–534.

Silva, JMC Santos, and Silvana Tenreyro. 2006. "The Log of Gravity." *The Review of Economics and Statistics*, 88(4): 641–658.

Sun, Liyang, and Sarah Abraham. 2021. "Estimating Dynamic Treatment Effects in Event Studies with Heterogeneous Treatment Effects." *Journal of Econometrics*, 225(2): 175–199.

Universitaet Regensburg, Universitaetsleitung. 2016. "Grundsaetze Der Universitaet Regensburg Zur Vergabe Von Leistungsbezuege."

Von Proff, Sidonia, Guido Buenstorf, and Martin Hummel. 2012. "University Patenting in Germany before and After 2002: What Role Did the Professors' Privilege Play?" *Industry and Innovation*, 19(1): 23–44.

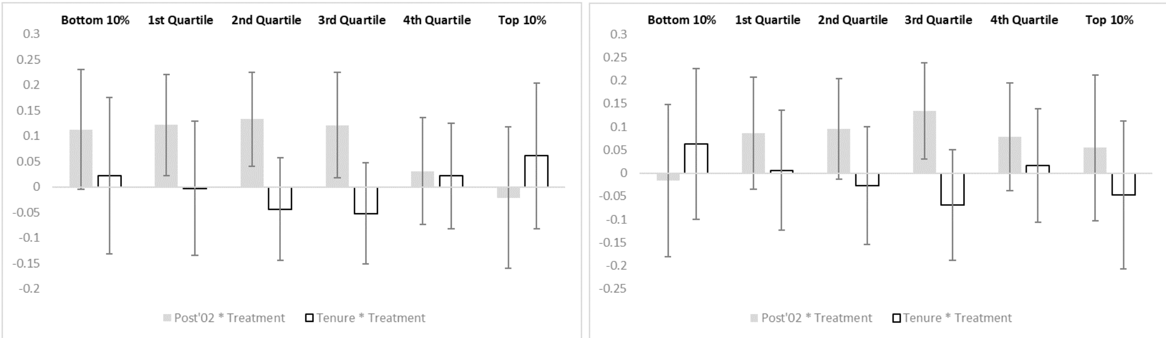
Waldinger, Fabian. 2012. "Peer Effects in Science: Evidence from the Dismissal of Scientists in Nazi Germany." *The Review of Economic Studies*, 79(2): 838–861.

Wissenschaftsrat. 2005. "Empfehlungen zur Ausgestaltung von Berufungsverfahren."

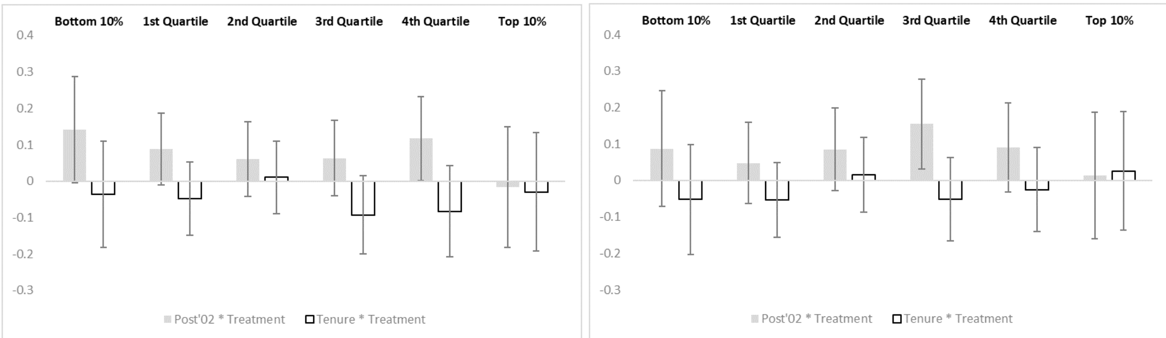
Wooldridge, Jeffrey M. 2010. *Econometric Analysis of Cross Section and Panel Data.* MIT press.

Wu, Yanhui, and Feng Zhu. 2022. "Competition, Contracts, and Creativity: Evidence from Novel Writing in a Platform Market." *Management Science*.

Xu, Lei, Tingting Nian, and Luis Cabral. 2020. "What Makes Geeks Tick? A Study of Stack Overflow Careers." *Management Science*, 66(2): 587–604.



(a) Quality Distributions: Number of Publications in Citation (L) and IFR Publication (R) Quantile Bins



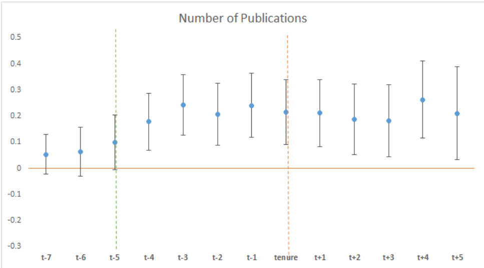
(b) Novelty and Impact Distributions: Number of Publications in Backward (L) and Forward (R) Cosine Similarity Quantile Bins

Figure 1: Quality and Cosine Similarity Distributions

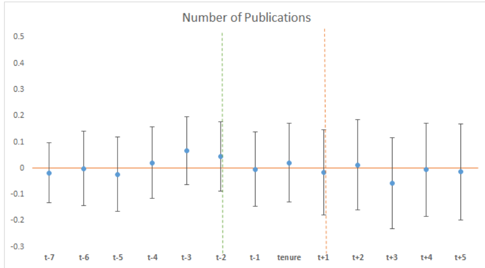
Notes: The histograms depict the coefficient estimates and corresponding 95% confidence intervals of the $post'02{*}Treatment$ (grey bars) and $Tenure{*}Treatment$ (white bars) interactions from separate regressions with the following dependent variables: citation quantile frequency variables (sub-figure a, left), impact factor rating quantile frequency variables (sub-figure a, right), backward cosine similarity quantile frequency variables (sub-figure b, left), forward cosine similarity quantile frequency variables (sub-figure b, right). To generate these dependent variables, percentiles of citations, impact factor ratings, and backward and forward cosine similarity metrics are calculated separately by field and publication year and the percentile cut-offs are used to generate quantile frequency variables for each author and publication year. All other specifications as in the baseline regression of the effort effect. See sections 4.2.3 and 4.2.8 for further details.

Treatment vs. Control

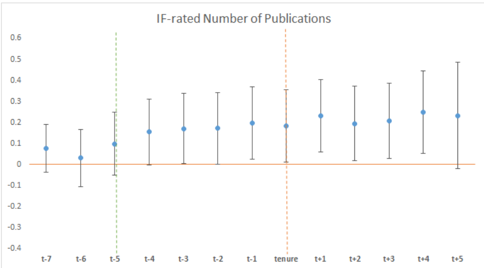
Placebo



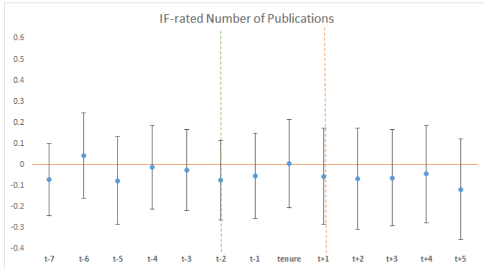
(a) Number of Publications



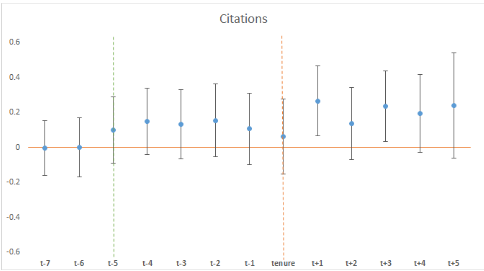
(b) Number of Publications



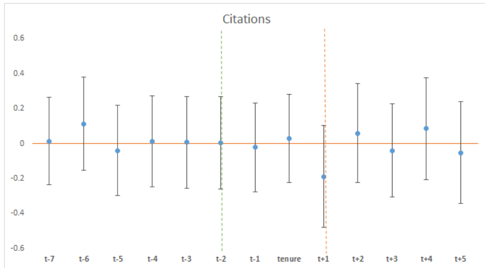
(c) Impact Factor-Rated Number of Publications



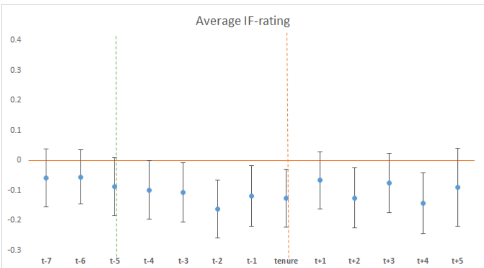
(d) Impact Factor-Rated Number of Publications



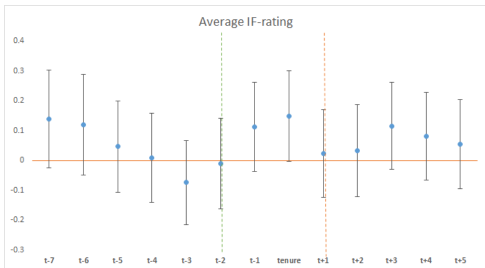
(e) Total Number of Citations



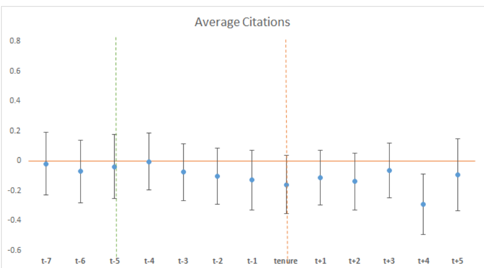
(f) Total Number of Citations



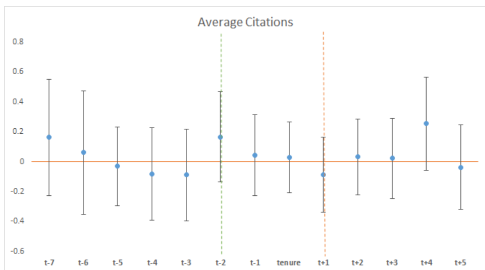
(g) Average Impact Factor Rating



(h) Average Impact Factor Rating



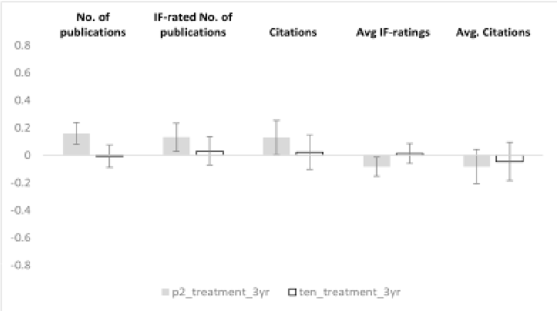
(i) Average Number of Citations



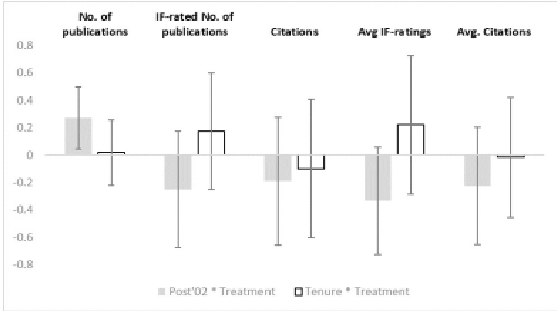
(j) Average Number of Citations

Figure 2: Pre-trends and Effect Dynamics

Notes: The plots 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, impact factor-rated number of publications, total number of citations, average impact factor rating, and average number of citations. All other specifications as in the baseline regression of the effort effect. See section 4.2.4 for further details.

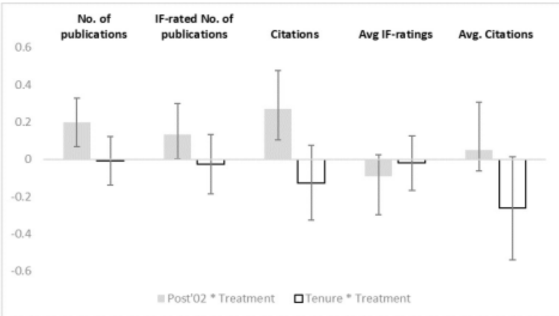


Natural and Applied Sciences

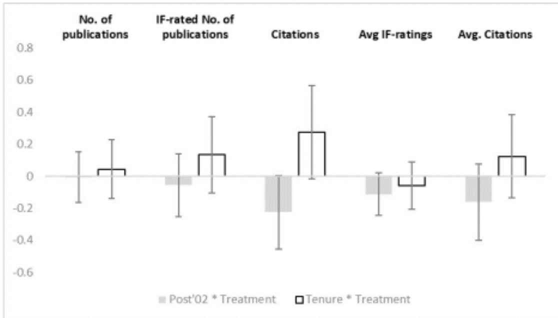


Social Sciences and Humanities

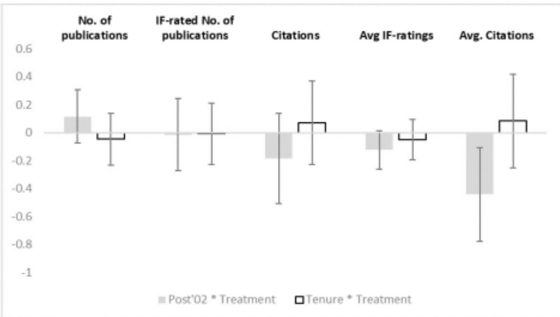
(a) Heterogeneous Effort Effect by Broad Academic Field



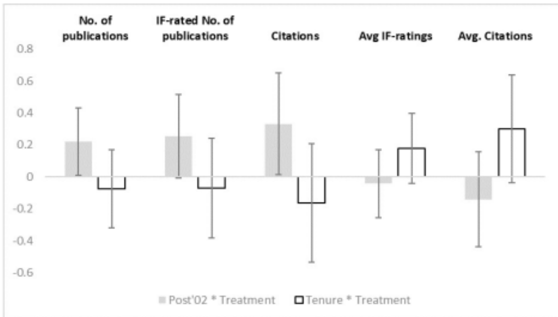
10th (top) decile



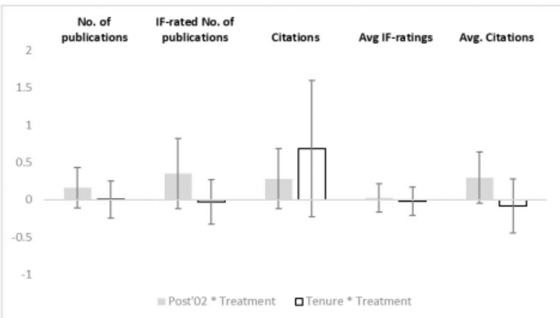
9th decile



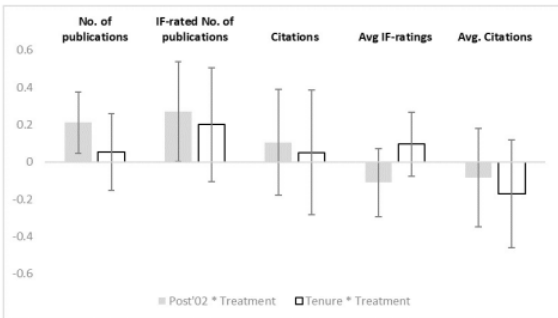
8th decile



7th decile



6th decile

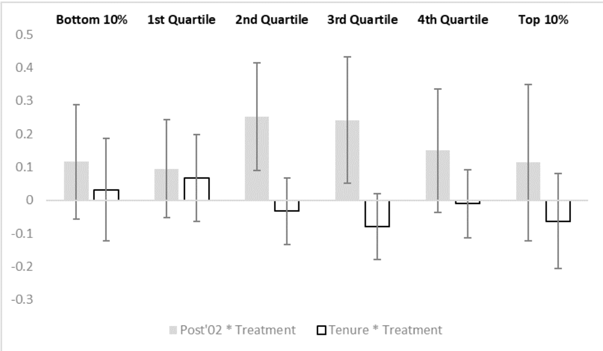


Below median

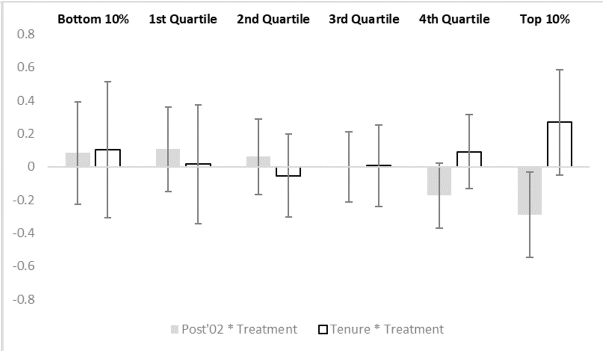
(b) Heterogeneous Effort Effect by Productivity Quantile

Figure 3: Heterogeneous Effort Effects by Broad Academic Field and Productivity Quantile

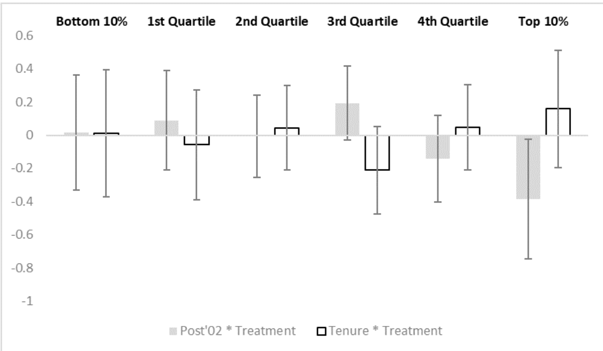
Notes: The histograms depict the coefficient estimates and corresponding 95% confidence intervals of the $post'02{*}Treatment$ (grey bars) and $Tenure{*}Treatment$ (white bars) interactions from separate regressions with the following dependent variables: number of publications, impact factor-rated number of publications, total number of citations, average impact factor rating, and average number of citations. In sub-figure a, the sample is restricted to academics in the *natural and applied sciences* in the histogram on the left, and academics in the *social sciences and humanities* in the histogram on the right. In sub-figure b, the samples are restricted to the top five productivity deciles and below median productive academics (from top left to bottom right). Productivity deciles are determined on the basis of the averages of the impact factor-rated number of publications over the three pre-announcement years 1999, 2000 and 2001, separately by academic field and treatment group. All other specifications as in the baseline regression of the effort effect. See sections 4.2.6 and 4.2.7 for further details.



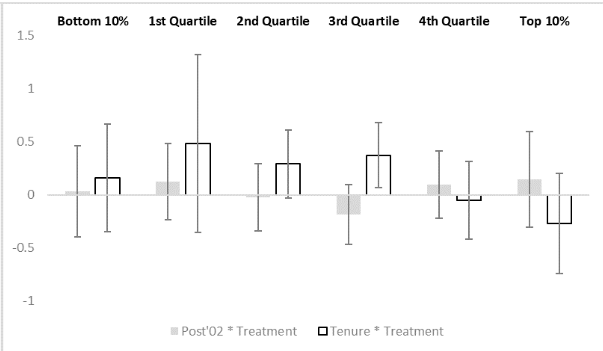
(a) 10th (top) Decile



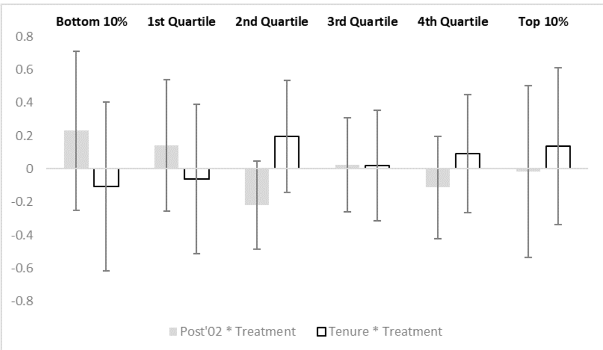
(b) 9th Decile



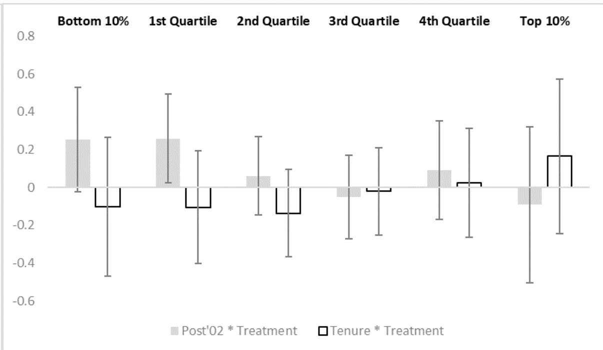
(c) 8thDecile



(d) 7th Decile



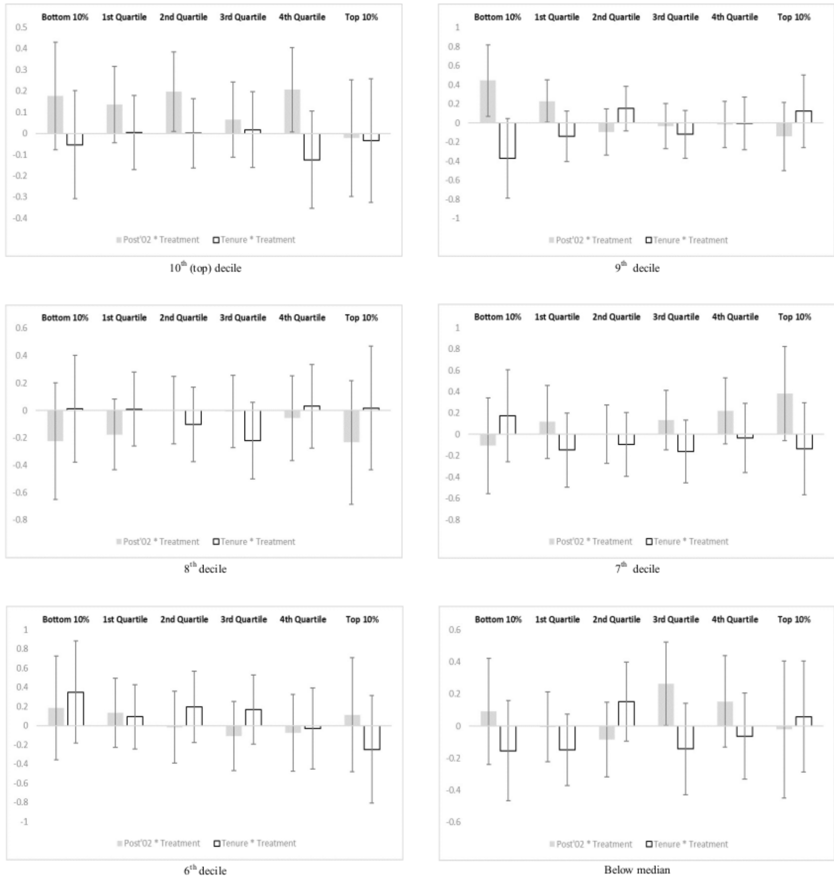
(e) 6thDecile



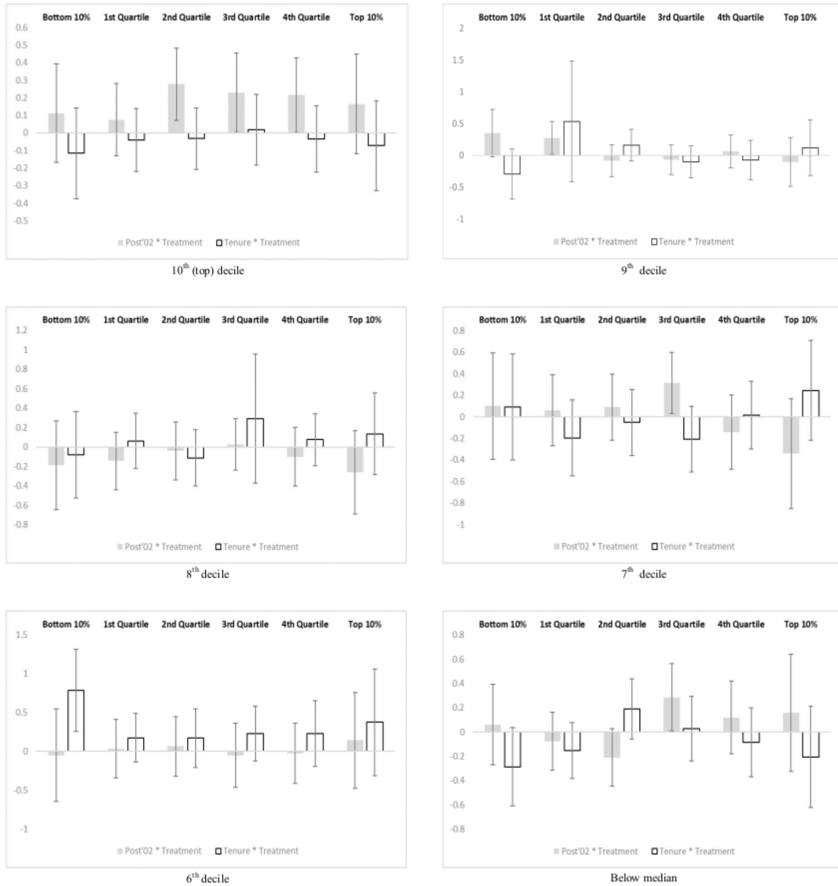
(f) Below Median

Figure 4: Heterogeneous Effort Effect - Quality distributions

Notes: The histograms depict the coefficient estimates and corresponding 95% confidence intervals of the $post'02{*}Treatment$ (grey bars) and $Tenure * Treatment$ (white bars) interactions from separate regressions with citation quantile frequency variables as dependent variable. Samples are restricted to the top five productivity deciles (sub-figures a-e) and below median productive academics (sub-figure f). Productivity deciles are determined on the basis of the averages of the impact factor-rated number of publications over the three pre-announcement years 1999, 2000 and 2001, separately by academic field and treatment group. All other specifications as in the baseline regression of the effort effect. See section 4.2.7 for further details.



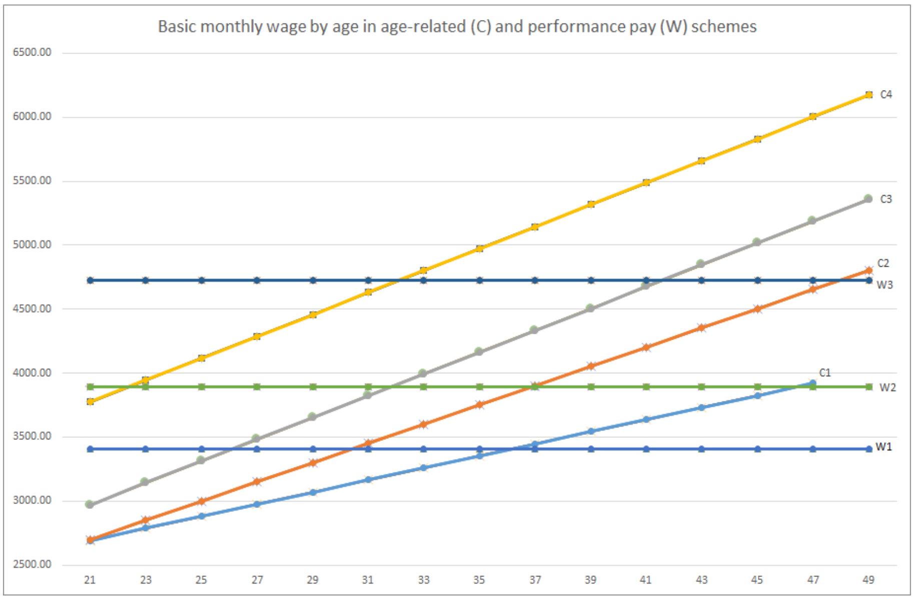
(a) Heterogeneous Results for Backward Cosine Similarity



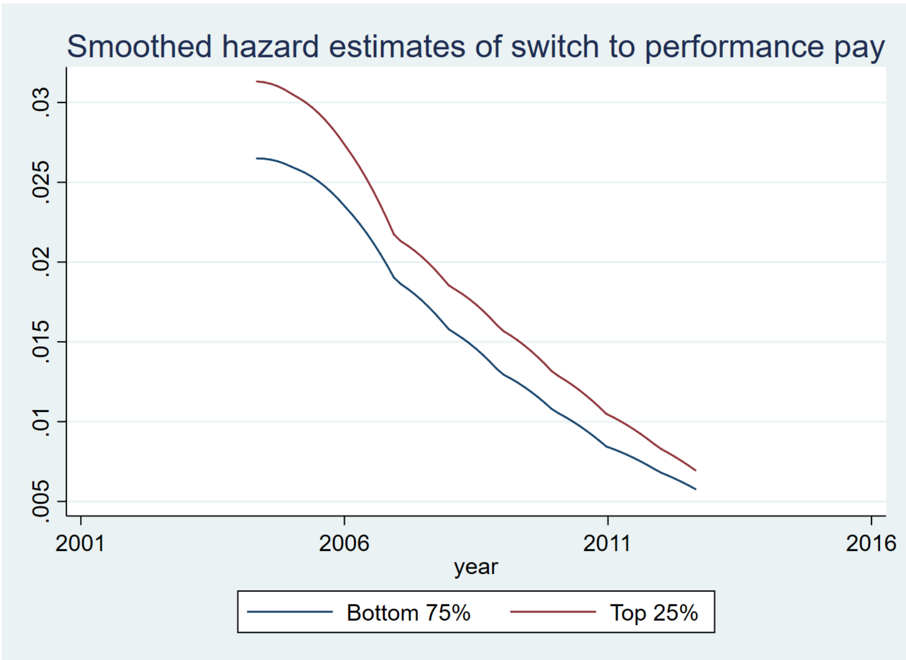
(b) Heterogeneous Results for Forward Cosine Similarity

Figure 5: Heteregenous Effort Effect - Cosine Similarity Distributions

Notes: The histograms depict the coefficient estimates and corresponding 95% confidence intervals of the $post'02 * Treatment$ (grey bars) and $Tenure * Treatment$ (white bars) interactions from separate regressions with backward cosine similarity quantile frequency variables (sub-figure a) and forward cosine similarity quantile frequency variables (sub-figure b) as dependent variable. In each sub-figure going from top left to bottom right, samples are restricted to the top five productivity deciles and below median productivity academics. All other specifications as in the baseline regression of the effort effect. See section 4.2.8 for further details.



(a) Comparison of Basic Wage Schedules



(b) Switching Hazard Functions by Productivity Quantiles

Figure 6: Basic Wage Schedules and Switching Hazard Functions

Notes: Subfigure a 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 monthly wages in former East-German states were 92.5% of these (Detmer and Preissler, 2006). Data source: Oeffentlicher Dienst (2004). Subfigure b shows 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. Quartiles are determined on the basis of the averages of the impact factor-rated number of publications over the three pre-implementation years (2002, 2003 and 2004), separately by academic field and tenure cohort. The sample is restricted to academics who held a tenured affiliation at a public university before 2005. See section 4.3.1 for further details.

Table 1: Baseline: Summary Statistics and Baseline Effort Effect Estimation

Panel A: Summary Statistics

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: Baseline Effort Effect Estimation

	# Publications	IF-rated Publications	Citations	Avg. IF-rating	Avg. 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

Notes: Panel A shows summary statistics for the sample used for the baseline effort effect estimation. 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 dropping academics once they pass away or retire) and includes data from 1993 until and including 2012. Panel B reports the coefficient estimates and standard errors of the $post'02 * Treatment$ and $Tenure * Treatment$ interactions from separate regressions with the following dependent variables: number of publications, impact factor-rated number of publications, total number of citations, average impact factor rating, and average number of citations. 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, $Tenured$ is 0 before an academic obtains their first tenured affiliation 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 clustered at the individual level. See section 4.2.2 for further details.

Table 2: Validity Checks

Panel A: Synthetic Cohorts	# Publications	IF-rated publications	Citations	Average IF-rating	Average citations
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
Panel B: 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 C: Publication Variables Weighted by Number of Authors					
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

Notes: The table reports the coefficient estimates and standard errors of the $post'02 * Treatment$ and $Tenure * Treatment$ interactions from separate regressions with the following dependent variables: number of publications, impact factor-rated number of publications, total number of citations, average impact factor rating, and average number of citations. In Panel A, assignment to the treatment and control cohorts is not based on the actual first tenured year, but on the average age at which academics start their first tenured affiliation. In Panel B, the control group is restricted to academics who do not switch to the performance pay scheme, where any first affiliation, position or contract renegotiation after implementation of the pay reform (as of 2005) is considered a switch. In panel C, the dependent variables are weighted by the number of authors on a publication. All other specifications as in the baseline regression of the effort effect. See sections 4.2.4, 4.2.5 and online appendix section 4.1 for further details.

Table 3: Tenure Probability Analysis

	1a	1b	2a	2b
Age	-0.262***	-0.142***	-0.262***	-0.142***
	(0.007)	(0.006)	(0.007)	(0.006)
Productivity	0.014***	0.015***	0.003***	0.004***
	(0.001)	(0.002)	(0.000)	(0.001)
Post'05		7.144***		7.160***
		(0.248)		(0.248)
Post'05 * Age		-0.172***		-0.172***
		(0.006)		(0.006)
Post'05 * Productivity		-0.001		-0.001
		(0.002)		(0.001)
Constant	4.917***	-0.193	4.933***	-0.19
	(0.228)	(0.238)	-0.227	-0.238
Number of Observations	213514	213514	213514	213514
Number of Subjects	26847	26847	26847	26847
Number of Tenure Starts	12749	12749	12749	12749
Log Likelihood	-21934.532	-21407.680	-21933.988	-21405.290
Chi-squared	3133.269	4340.346	3116.022	4329.72
Rho	2.618	2.644	2.619	2.646

Notes: The table reports estimation results of Weibull proportional hazard models of transitions into first tenured positions. The event of interest here (the "failure" event) is the start of the first tenured affiliation, for academics who do not hold a foreign affiliation immediately preceding this change. The time from the completion of the habilitation until the first tenured position is used as duration variable and 1998 is the entry date. Academics are "at risk" of obtaining a tenured position from habilitation onwards. In columns 1a and 1b, the Productivity variable used is the number of publications; in columns 2a and 2b it is the impact factor-rated number of publications. These productivity variables are lagged by two years. All models control for field fixed effects and are estimated for the years 1999-2013. Academics are dropped once they pass away or retire. Standard errors are robust and clustered by individual academic. See section A4.3 for further details.

Table 4: Selection Analysis

	Panel A: Treatment versus Control				Panel B: Placebo	
	1a	1b	2a	2b	3a	3b
Treatment	0.658	0.046	0.576	-0.207	-0.790	-1.653**
	(0.401)	(0.426)	(0.412)	(0.436)	(0.624)	(0.667)
Age	-0.126***	-0.299***	-0.131***	-0.313***	-0.165***	-0.449***
	(0.007)	(0.017)	(0.007)	(0.017)	(0.010)	(0.040)
Avg Productivity	-0.001	-0.004*	-0.038***	-0.056***	-0.003	-0.006
	(0.002)	(0.002)	(0.015)	(0.013)	(0.014)	(0.015)
Age * Treatment	-0.028***	-0.007	-0.026***	-0.001	0.013	0.034**
	(0.009)	(0.009)	(0.009)	(0.010)	(0.013)	(0.014)
Avg Productivity * Treatment	0.004**	0.007***	0.032**	0.051***	-0.008	-0.005
	(0.002)	(0.002)	(0.015)	(0.014)	(0.015)	(0.016)
Avg Productivity * Age			0.001***	0.001***	0.000	0.000
			(0.000)	(0.000)	(0.000)	(0.000)
Avg Productivity * Age * Treatment			-0.001*	-0.001***	0.000	0.000
			(0.000)	(0.000)	(0.000)	(0.000)
Age at Tenure		0.192***		0.200***		0.293***
		(0.015)		(0.016)		(0.038)
Constant	2.381***	1.599***	2.576***	1.924***	3.667***	2.721***
	(0.329)	(0.347)	(0.334)	(0.350)	(0.449)	(0.479)
Number of Observations	80131	80131	80131	80131	51431	51431
Number of Subjects	14972	14972	14972	14972	6960	6960
Number of Switches	2435	2435	2435	2435	1099	1099
Log Likelihood	-7545.484	-7404.981	-7541.479	-7394.569	-3232.134	-3178.716
Chi-squared	1365.450	1310.708	1378.937	1326.956	595.064	557.869
Rho	1.345	1.653	1.345	1.671	1.439	2.516

Notes: The table reports estimation results of Weibull proportional hazard models of contract changes (affiliation or position changes or contract renegotiations) by tenured professors. In Panel A, the treatment variable is 1 for academics who have made tenure before 2005 and 0 for those who make tenure afterwards. Panel B reports the results for a placebo experiment where the placebo-treatment group comprises academics who start their first tenured position before 2002, while academics who start their first tenured affiliation between 2002 and 2005 act as placebo-control group. "Avg Productivity" is calculated as three year pre-implementation averages (2002-2004) of the impact factor-rated number of publications. All specifications include academic field fixed effects and are estimated for years t>2004 and for academics who start their first tenured affiliation as of 1999 (excluding those with a foreign affiliation directly prior to this). The models in the "b" columns models also control for synthetic age at tenure. Standard errors are robust and clustered by individual academic. See section 4.3.2 for further details.

Online Appendix

A1 Model of Performance Pay and Multitasking

This Appendix provides a detailed exposition of the model outlined in the main text with proofs. The set-up of the model is provided in the main text and not repeated here. I first discuss the full insurance (flat wage) case as a benchmark before presenting the performance pay case. I then derive implications for the effort and selection effects of performance pay, by comparing equilibrium behavior under performance pay to that under flat wage pay. Finally, I compare the effort and selection effects of a performance pay system comprising career concerns and explicit incentives to one with career concerns only or explicit incentives only, and show that there are salient differences.

A1.1 Flat Wage

Consider the full insurance case first. Principals offer flat wage contracts and cannot tailor wages to their beliefs about agent ability or effort. This is the case in the German age-related pay system, in which principals (universities) do not have discretion over wage contracts. It applies equally to other markets in which contracts are similarly constrained, for instance due to government regulation or because output measures are neither verifiable nor publicly observable.. With a per period discount factor δ, the expected life-time utility of an agent is given by:

$$U = E\left\{-exp\left[-r\sum_{t=0}^{\infty}\delta^{t}\left(w_{t}-g\left(\vec{e}_{t}\right)\right)\right]\right\}$$

Since pay $w_t \geq 0$ does not depend on output or effort in any period, an agent's payoff in period t in certainty equivalents is simply:

$$u_t\left(w_t,\vec{e}_t\right) = E\left\{w_t - c\left(\vec{e}_t\right)\right\} = w_t - g\left(\vec{e}_t\right) \tag{9}$$

It follows that optimal effort in any period t equals minimum effort levels:

$$argmax_e\left[u_t\left(w_t,\vec{e}_t\right)\right] = argmax_e\left\{w_t - g\left(\vec{e}_t\right)\right\} = \vec{\underline{e}} \tag{10}$$

Any differences in effort therefore reflect differences in intrinsic motivation, minimum output requirements such as tenure requirements and other such mechanisms.

Suppose an agent's outside option yields per period utility $\underline{u}_t$. Principals then need to set the wage w_t such that $E\left\{w_t - g\left(\vec{e}_t^{*}\right)\right\} = w_t \geq \underline{u}_t$ to attract agents. In equilibrium therefore $w_t^{*} = \underline{u}_t$.

A1.2 Career Concerns and Incentive Contracts

Consider now the case of a perfectly competitive labor market in which output is verifiable and publicly observable and only short-term contracts are feasible.

Proof of Proposition 1:Because output is observable to all market participants, the market can use this information to update its beliefs about agent ability. Given the assumptions on ability and output noise, $\overrightarrow{y_t}$

is bivariate normal. The prior distribution of agent ability θ is normal as well, and hence so is the posterior distribution of θ. Using well-known formulas in De Groot (1970), the mean m_t and variance σ_t^2 of this posterior distribution of θ, given past output $(\vec{y}_0, .., \vec{y}_{t-1})$ and conjectured effort levels $(\vec{\hat{e}}_0, .., \vec{\hat{e}}_{t-1})$, are then given by:,

$$m_t := E\left[\theta|(\vec{y}_1, .., \vec{y}_{t-1}); (\vec{\hat{e}}_0, .., \vec{\hat{e}}_{t-1})\right] = \frac{\sigma_\varepsilon^2\sigma_\nu^2 m_0 + \sigma_0^2 \sum_{s=0}^{t-1} \left\{\sigma_\nu^2 \left(y_{p,s} - \hat{e}_{p,s}\right) + \sigma_\varepsilon^2 \left(y_{q,s} - \hat{e}_{q,s}\right)\right\}}{\sigma_\varepsilon^2\sigma_\nu^2 + t\sigma_0^2 \left(\sigma_\varepsilon^2 + \sigma_\nu^2\right)} \tag{11}$$

and

$$\sigma_t^2 := var\left(\theta_t\right) = \frac{\sigma_0^2\sigma_\varepsilon^2\sigma_\nu^2}{\sigma_\varepsilon^2\sigma_\nu^2 + t\sigma_0^2\left(\sigma_\varepsilon^2 + \sigma_\nu^2\right)} \tag{12}$$

Perfect competition in the labor market implies that agents are offered contracts that will earn them their expected productivity. That is:

$$E\left[1^T\vec{y}_t\right] = \left(m_t + \hat{e}_{p,t}\right) + \left(m_t + \hat{e}_{q,t}\right) = E\left[w_t\left(\vec{y}_t\right)\right] = c_t + \vec{b_t}^T\vec{y}_t \tag{13}$$

Here 1^T denotes a $1x2$ matrix of ones.[45] An agent's expected lifetime utility is then given by

$$\begin{aligned} U &= E\left\{-exp\left[-r\sum_{t=0}^{\infty}\delta^t\left(c_t + \vec{b_t}^T\vec{y}_t - g\left(\vec{e}_t\right)\right)\right]\right\} \\ &= E\left\{-exp\left[-r\sum_{t=0}^{\infty}\delta^t\left(\left(m_t + e_{p,t}\right) + \left(m_t + e_{q,t}\right) - g\left(\vec{e}_t\right)\right)\right]\right\} \end{aligned} \tag{14}$$

In any period t, effort affects the payoff that period through the explicit bonus $\vec{b_t}$, as well as future wage payments through updated beliefs about ability. If only career concerns incentives were present, we know, following Holmström (1999) and using (11) and (14), that optimal effort is given by the following first order conditions when $\vec{e}_t \gg 0$

$$\frac{\partial g\left(\vec{e}_t\right)}{\partial e_{p,t}} = \sigma_\nu^2 CC_t; \frac{\partial g\left(\vec{e}_t\right)}{\partial e_{q,t}} = \sigma_\varepsilon^2 CC_t \tag{15}$$

where $CC_t = \sum_{\tau=1}^{\infty}\delta^\tau \frac{\sigma_0^2}{\sigma_\varepsilon^2\sigma_\nu^2 + (t+\tau)\sigma_0^2(\sigma_\varepsilon^2+\sigma_\nu^2)}$. As noted in Macleod (2022), adding explicit performance incentives to career concerns in this model does not affect the career concerns incentives. Ability enters output additively and does not affect the marginal cost of effort, so the optimal bonus $\vec{b_t}$ does not depend on m_t, as will be shown below. Future income risk is therefore unaffected by effort, and career concerns incentives are unaffected by the introduction of explicit performance incentives. The first order conditions for optimal effort when $\vec{e}_t \gg 0$ are thus

$$\frac{\partial g\left(\vec{e}_t\right)}{\partial e_{p,t}} = \left(e_{p,t} - \bar{e}_{p,t}\right) + d\left(e_{q,t} - \bar{e}_{q,t}\right) = b_{p,t} + \sigma_\nu^2 CC_t \tag{16}$$

$$\frac{\partial g\left(\vec{e}_t\right)}{\partial e_{q,t}} = \left(e_{q,t} - \bar{e}_{q,t}\right) + d\left(e_{p,t} - \bar{e}_{p,t}\right) = b_{q,t} + \sigma_\varepsilon^2 CC_t \tag{17}$$

[45] This vector of ones implies that principals' return to the output quantity and quality (signal) is equal to this output (signal). It is straightforward to allow for other rates of return and all the model's results would continue to hold.

Substituting (16) in (17) and rearranging, optimal effort when $\vec{e}_t \gg 0$ is then:

$$e^*_{p,t} = \bar{e}_p + \frac{1}{1-d^2}\left(b_{p,t} + \sigma^2_\nu CC_t - d\left(b_{q,t} + \sigma^2_\varepsilon CC_t\right)\right) \tag{18}$$

$$e^*_{q,t} = \bar{e}_p + \frac{1}{1-d^2}\left(b_{q,t} + \sigma^2_\varepsilon CC_t - d\left(b_{p,t} + \sigma^2_\nu CC_t\right)\right) \tag{19}$$

To derive the optimal bonus $\overrightarrow{b^*_t}$, I again use that $\overrightarrow{b_t}$ affects only the effort and income risk in period t and is therefore chosen to maximize agent utility that period. This amounts to optimizing the certainty equivalent with respect to $\overrightarrow{b_t}$

$$\overrightarrow{b^*_t} = argmax_{b_t}\left\{\left(m_t + e^*_{p,t}\left(\overrightarrow{b_t}\right)\right) + \left(m_t + e^*_{q,t}\left(\overrightarrow{b_t}\right)\right) - g\left(\vec{e}^*_t\left(\overrightarrow{b_t}\right)\right) - \frac{r}{2}\left(\overrightarrow{b_t}^2 \Sigma^2_t\right)\right\}$$

where $\Sigma^2_t = \begin{bmatrix} \sigma^2_t + \sigma^2_\varepsilon & \sigma^2_t \\ \sigma^2_t & \sigma^2_t + \sigma^2_\nu \end{bmatrix}$. Substituting for $e^*_{p,t}$ and $e^*_{q,t}$ using (18) and (19), we can derive the following first order conditions for optimal bonus rates when $\vec{e}_t \gg 0$

$$\frac{1 - d - (b^*_{p,t} - db^*_{q,t} + CC_t(\sigma^2_\nu - d\sigma^2_\varepsilon))}{1-d^2} = r\left(b^*_{q,t}\sigma^2_t + b^*_{p,t}\left(\sigma^2_t + \sigma^2_\varepsilon\right)\right) \tag{20}$$

$$\frac{1 - d - (b^*_{q,t} - db^*_{p,t} + CC_t(\sigma^2_\nu - d\sigma^2_\varepsilon))}{1-d^2} = r\left(b^*_{p,t}\sigma^2_t + b^*_{q,t}\left(\sigma^2_t + \sigma^2_\nu\right)\right) \tag{21}$$

Rearranging, we get

$$b^*_{p,t} = \frac{(1-d) + b^*_{q,t}\left(d - (1-d^2)\, r\sigma^2_t\right) - CC_t(\sigma^2_\nu - d\sigma^2_\varepsilon)}{1 + (1-d^2)\, r\left(\sigma^2_t + \sigma^2_\varepsilon\right)}$$

$$b^*_{q,t} = \frac{(1-d) + b^*_{p,t}\left(d - (1-d^2)\, r\sigma^2_t\right) - CC_t(\sigma^2_\varepsilon - d\sigma^2_\nu)}{1 + (1-d^2)\, r\left(\sigma^2_t + \sigma^2_\varepsilon\right)}$$

Substituting one into the other and rearranging yields

$$b^*_{p,t} = \frac{1 + (1-d)\, r\sigma^2_\nu - CC_t\left(\sigma^2_\nu\left(1 + r\left(\sigma^2_\nu - d\sigma^2_\varepsilon\right)\right) - r\,(1+d)\,\sigma^2_t\left(\sigma^2_\varepsilon - \sigma^2_\nu\right)\right)}{1 + r\left(2\,(1+d)\,\sigma^2_t + \sigma^2_\varepsilon + \sigma^2_\nu\right) + (1-d^2)\, r^2\left(\sigma^2_\varepsilon\sigma^2_\nu + \sigma^2_t\left(\sigma^2_\varepsilon + \sigma^2_\nu\right)\right)} \tag{22}$$

$$b^*_{q,t} = \frac{1 + (1-d)\, r\sigma^2_\varepsilon - CC_t\left(\sigma^2_\varepsilon\left(1 + r\left(\sigma^2_\varepsilon - d\sigma^2_\nu\right)\right) - r\,(1+d)\,\sigma^2_t\left(\sigma^2_\nu - \sigma^2_\varepsilon\right)\right)}{1 + r\left(2\,(1+d)\,\sigma^2_t + \sigma^2_\varepsilon + \sigma^2_\nu\right) + (1-d^2)\, r^2\left(\sigma^2_\varepsilon\sigma^2_\nu + \sigma^2_t\left(\sigma^2_\varepsilon + \sigma^2_\nu\right)\right)} \tag{23}$$

It follows from the assumptions on $c\left(\vec{e}_t\right)$ and the fact that the right-hand sides depend on the model's parameters only that (22) and (23) define the unique optimal bonuses in each period when $\vec{e}_t \gg 0$.

Finally, substituting (22) and (23) into (18) and (19) and rearranging, we get the following expressions for optimal effort when $\vec{e}_t \gg 0$:

$$e^*_{p,t} = \bar{e}_{p,t} + \frac{1 + r\left(\sigma^2_\nu - d\sigma^2_\varepsilon\right)\left(1 + CC_t\,(1+d)\, CV_t\right)}{(1+d)\, D_t} \tag{24}$$

$$e^*_{q,t} = \bar{e}_{q,t} + \frac{1 + r\left(\sigma^2_\varepsilon - d\sigma^2_\nu\right)\left(1 + CC_t\,(1+d)\, CV_t\right)}{(1+d)\, D_t} \tag{25}$$

where $CV_t := r\left(\sigma_\varepsilon^2\sigma_\nu^2 + \sigma_t^2\left(\sigma_\varepsilon^2+\sigma_\nu^2\right)\right)$ and $D_t := 1 + r\left(2\left(1+d\right)\sigma_t^2 + \sigma_\varepsilon^2 + \sigma_\nu^2\right) + r\left(1-d^2\right)CV_t$. By the same reasoning as used for optimal bonuses above, it follows that (24) and (25) define the unique optimal effort levels. What remains to be shown, is that the market's conjectured effort levels are correct in equilibrium: $\vec{\tilde{e}}_t = \vec{e}_t^*$. It is immediate that this condition is met, since $\vec{e}_t^*$ does not depend on $\vec{\tilde{e}}_t$.

It follows from (24) that $e_{p,t}^* > 0$ always, since $0 < d < 1$, $\sigma_\nu^2 > \sigma_\varepsilon^2$, and agents are risk averse ($r > 0$). The right-hand side of (25) can be negative, however, since the limit of $e_{q,t}^* - \bar{e}_{q,t}$ is negative when $\sigma_\nu \to \infty$. For $e_{q,t}^* - \bar{e}_{q,t} < 0$ and $\left|e_{q,t}^* - \bar{e}_{q,t}\right| > \bar{e}_{q,t}$ then, the non-negativity constraint on quality effort binds and $e_{q,t}^* = 0$. When this happens, the optimal bonus rates and effort levels are given by (derived using the same steps as above):

$$b_{p,t}^* = -\frac{\left(\sigma_t^2+\sigma_\nu^2\right)\left(CC_t\sigma_\nu^2-1\right)}{\sigma_t^2+\sigma_\nu^2+CV_t}; b_{q,t}^* = \frac{\sigma_t^2\left(CC_t\sigma_\nu^2-1\right)}{\sigma_t^2+\sigma_\nu^2+CV_t} \tag{26}$$

$$e_{p,t}^* = \bar{e}_{p,t} + d\bar{e}_{q,t} + \frac{\sigma_t^2+\sigma_\nu^2\left(1+CV_tCC_t\sigma_\nu^2\right)}{\sigma_t^2+\sigma_\nu^2+CV_t}; e_{q,t}^* = 0 \tag{27}$$

This completes the proof of Proposition 1. ■

A1.3 Implications

A comparison of the performance pay equilibrium characterized in Proposition 1 to the flat wage equilibrium in subsection A1.1 yields a number of testable implications for output quantity and quality levels, heterogeneous responses and selection into performance pay. These implications are presented in order below.

A1.3.1 Effort Effect

We may assume, without loss of generality, that one of the output dimensions is more risky or less precisely measured. In academia, as in other knowledge work jobs, output quality is noisier than quantity; i.e. $\sigma_\nu^2 > \sigma_\varepsilon^2$. The proposition below then characterizes the effort response to performance pay:

Proposition 2 - Effort Response: *Quantity effort is unambiguously higher in the performance pay system than under flat wage pay ($e_{p,t}^* > \bar{e}_{p,t}$). Quality effort is higher for all ability classes iff risk aversion is low ($e_{q,t}^* > \bar{e}_{q,t}$ iff $0 < r < r^{pp}$). If risk aversion is high ($r > r^{pp}$), quality effort increases for high ability classes ($m > m^{pp}$) and decreases for low ability classes ($m < m^{pp}$).*

Corollary 1 - Effort Response: *If the quality measure is very noisy, quality effort may decrease in all ability classes; specifically if $d\left(m^{pp}\right) < d\left(\overline{m}\right)$.*

Here m^{pp} is the positive root that solves

$$1 + r\left(\sigma_\varepsilon^2 - d\left(m^{pp}\right)\sigma_\nu^2\right)\left(1 + CC_t\left(1+d\left(m^{pp}\right)\right)CV_t\right) = 0 \tag{28}$$

while r^{pp} is defined as

$$r^{pp} = \left(\frac{1}{2CC_t\left(1+d\left(\underline{m}\right)\right)\widetilde{CV_t}}\right)\left[\sqrt{\frac{\left(\sigma_\varepsilon^2 - d\left(\underline{m}\right)\sigma_\nu^2\right) - 4\left(CC_t\left(1+d\left(\underline{m}\right)\right)\widetilde{CV_t}\right)}{\left(\sigma_\varepsilon^2 - d\left(\underline{m}\right)\sigma_\nu^2\right)}} - 1\right] \tag{29}$$

and $\widetilde{CV_t} = \left(\sigma_\varepsilon^2\sigma_\nu^2 + \sigma_t^2\left(\sigma_\varepsilon^2+\sigma_\nu^2\right)\right)$ (i.e. CV_t without r).

Proof: Since $0 < d < 1$ and agents are risk averse ($r > 0$), the quantity effort result $e^*_{p,t} > \bar{e}_{p,t}$ follows immediately from (24) and (27) and the assumption that $\sigma_\nu^2 > \sigma_\varepsilon^2$.

To prove that quality effort is higher for all ability classes if risk aversion is low, note that (25) implies that $e^*_{q,t} > \bar{e}_{q,t}$ if and only if

$$1 + r\left(\sigma_\varepsilon^2 - d\sigma_\nu^2\right)\left(1 + CC_t\left(1+d\right)r\widetilde{CV_t}\right) > 0 \tag{30}$$

It is immediate that this inequality holds for all $r > 0$ when $d < \sigma_\varepsilon^2/\sigma_\nu^2$ (sufficient, not necessary). The left-hand side of (30) and, by extension, (25) decreases in r when $d > \sigma_\varepsilon^2/\sigma_\nu^2$. Hence, for any $\tilde{d} > \sigma_\varepsilon^2/\sigma_\nu^2$, (30) is fulfilled iff $0 < r < \tilde{r}\left(\tilde{d}\right)$, where $\tilde{r}$ is such that (30) holds with equality at $\tilde{r}, \tilde{d}$. The left-hand side of (30) is a quadratic equation in r, so $\tilde{r}$ is the positive root of this equation at $\tilde{d}$. Note that the coefficient of the linear and quadratic term in r are either both positive (when $d < \sigma_\varepsilon^2/\sigma_\nu^2$) or both negative (when $d > \sigma_\varepsilon^2/\sigma_\nu^2$). When both are positive, the left-hand side of (30) has no real roots, while there is only one positive real root when both are negative. This root is equal to the right-hand side of (29), evaluated at $\tilde{d}$. It is straightforward to show that $\partial\tilde{r}/\partial d < 0$, hence $r^{pp} = \tilde{r}\left(d\left(\underline{m}\right)\right) < \tilde{r}\left(d\left(m_0\right)\right), \forall m_0 : d\left(\underline{m}\right) > d\left(m_0\right) > \sigma_\varepsilon^2/\sigma_\nu^2$. But then (30) is fulfilled for all ability classes if $0 < r < r^{pp}$. This completes the proof of the second statement in the proposition.

To prove the third statement, note first that the above implies that $e^*_{q,t} < \bar{e}_{q,t}$ for some $d : 1 > d\left(\underline{m}\right) \geq d > \sigma_\varepsilon^2/\sigma_\nu^2$ when $r > r^{pp}$, while $e^*_{q,t} > \bar{e}_{q,t}, \forall r > 0$ when $0 < d < \sigma_\varepsilon^2/\sigma_\nu^2$. Furthermore, $e^*_{q,t} = \bar{e}_{q,t}$ when (30) holds with equality. Define the ability class and corresponding rate of substitution at which $e^*_{q,t} = \bar{e}_{q,t}$ as $d\left(m^{pp}\right)$ (see (28)). To prove that $e^*_{q,t} > \bar{e}_{q,t}$ is fulfilled if and only if $m > m^{pp}$ when $r > r^{pp}$, it then suffices to show that $\partial e^*_{q,t}/\partial d < 0$ when quality $e^*_{q,t} > 0$, since $\partial d/\partial m_0 < 0$. To do so, write $e^*_{q,t}$ in (25) as the sum of $\bar{e}_{q,t}$, a component that is driven by explicit incentives and another that is driven by career concerns:

$$\bar{e}_{q,t} + \frac{1 + r\left(\sigma_\varepsilon^2 - d\sigma_\nu^2\right)}{\left(1+d\right)D_t} + \frac{r\left(\sigma_\varepsilon^2 - d\sigma_\nu^2\right)CC_t\left(1+d\right)CV_t}{\left(1+d\right)D_t}$$

Taking the derivative with respect to d then yields:

$$\begin{aligned}\partial e^*_{q,t}/\partial d = \quad & -\frac{\left(1+2r\left(2(1+d)\sigma_t^2+\sigma_\varepsilon^2+\sigma_\nu^2\right)\right)+(1+d)r^2\left(\left(1-d+2d^2\right)\sigma_\nu^2+(1-3d)\sigma_\varepsilon^2\right)CV_t-R_{t,q}}{(1+d)^2D_t^2} \\ & -\frac{\left(CC_trCV_t\right)\left(\sigma_\nu^2+r\left(2\sigma_t^2+\sigma_\nu^2\right)\left(\sigma_\varepsilon^2+\sigma_\nu^2\right)+r\left(\left(1+d^2\right)\sigma_\nu^2-2d\sigma_\varepsilon^2\right)CV_t\right)}{D_t^2}\end{aligned} \tag{31}$$

where $R_{t,q} := \left(r^2\left(\sigma_\varepsilon^4 + \sigma_\nu^4 - \sigma_\varepsilon^2\sigma_\nu^2\left(2d + 3d^2 - 3\right) + \left(1+d\right)\sigma_t^2\left(\left(3-5d\right)\sigma_\nu^2 + \left(5-3d\right)\sigma_\varepsilon^2\right)\right)\right)$. Given the assumptions on the variances ($\left(\sigma_\nu^2, \sigma_\varepsilon^2, \sigma_0^2\right) \gg 0, \sigma_\nu^2 > \sigma_\varepsilon^2$), both derivative components are negative $\forall d : 0 < d < 1$, and hence $\partial e^*_{q,t}/\partial m_0 > 0$ when $e^*_{q,t} > 0$. This completes the proof of the proposition.

The corollary then follows from the fact that $\partial d\left(m^{pp}\right)/\partial\sigma_\nu$ is negative. To see this, apply the implicit function theorem to (28) to derive $\partial d\left(m^{pp}\right)/\partial\sigma_\nu$ as:

$$\begin{aligned}\frac{\partial d(m^{pp})}{\partial\sigma_\nu} = & \left[\frac{-1}{r\left(\left(\sigma_\varepsilon^2-d(m^{pp})\sigma_\nu^2\right)CC_tCV_t-\sigma_\nu^2\left(1+CC_t\left(1+d(m^{pp})\right)CV_t\right)\right)}\right]\left\{-2rd\left(m^{pp}\right)\sigma_\nu\left(1+CC_t\left(1+d\left(m^{pp}\right)\right)CV_t\right)+\right. \\ & \left. r\left(\sigma_\varepsilon^2 - d\left(m^{pp}\right)\sigma_\nu^2\right)\left(1+d\left(m^{pp}\right)\right)\sigma_0^4\sigma_\varepsilon^4\left(\sum_{\tau=1}^\infty\delta^\tau\frac{(\tau-1)\sigma_\varepsilon^4\sigma_\nu^4+tS_\tau+t^2SS_\tau+t^3\sigma_0^4\left(\sigma_\varepsilon^2+\sigma_\nu^2\right)^2}{\left(\sigma_\varepsilon^2\sigma_\nu^2+t\sigma_0^2\left(\sigma_\varepsilon^2+\sigma_\nu^2\right)\right)^2\left(t\sigma_0^2\left(\sigma_\varepsilon^2+\sigma_\nu^2\right)+\tau\sigma_0^2\sigma_\nu^2+\sigma_\varepsilon^2\left(\sigma_\nu^2+\tau\sigma_0^2\right)\right)^2}\right)\right\}\end{aligned}$$

Here $S_{\tau} = \left(\tau\sigma_0^4\sigma_\nu^4 + 2\tau\sigma_0^2\sigma_\varepsilon^2\sigma_\nu^2\left(\sigma_0^2+\sigma_\nu^2\right) + \sigma_\varepsilon^4\left(\sigma_\nu^4 + \tau\sigma_0^4 + 2\tau\sigma_0^2\sigma_\nu^2\right)\right)$ and $SS_{\tau} = \sigma_0^2\left(\sigma_\varepsilon^2+\sigma_\nu^2\right)$ $\left(2\sigma_\varepsilon^2\sigma_\nu^2 + (\tau+1)\,\sigma_0^2\left(\sigma_\varepsilon^2+\sigma_\nu^2\right)\right)$. It is immediate from the above that the denominator (in square brackets) is positive and the numerator (in curly brackets) is negative when $d\left(m^{pp}\right) > \sigma_\varepsilon^2/\sigma_\nu^2$. Since $d\left(m^{pp}\right) > \sigma_\varepsilon^2/\sigma_\nu^2$ for all applicable parameter values, $\partial d\left(m^{pp}\right)/\partial\sigma_\nu$ is negative and the threshold rate of substitution $d\left(m^{pp}\right)$ decreases as noise in the quality signal (σ_ν^2) increases. If, as a result, $d\left(m^{pp}\right) < d\left(\overline{m}\right)$, no agent ability class increases quality effort under performance pay. ■

A1.3.2 Heterogeneous Responses

Next, I characterize heterogeneity in the response to performance pay by ability type, starting with heterogeneity in the quantity effort response. As a building block for this characterization, I provide an overview of conditions under which quality effort is positive in Lemma 1.

Lemma 1 - **Positive Quality Effort**: *Quality effort is positive for all ability classes if*

$$\bar{e}_{q,t} \geq \bar{e}_{q,t}^{a,l}\ AND\ \bar{e}_{q,t} \geq \bar{e}_{q,t}^{b,l}\ OR\ \begin{cases} r < r^{q1,l} & if\ \bar{e}_{q,t} < \bar{e}_{q,t}^{a,l} \\ r < r^{q0,l}\ OR\ r > r^{q1,l} & if\ \bar{e}_{q,t} > \bar{e}_{q,t}^{a,l}\ AND\ \bar{e}_{q,t} < \bar{e}_{q,t}^{b,l} \end{cases} \tag{32}$$

Quality effort is positive for high ability classes and zero for low ability classes if

$$\begin{cases} r^{q1,h} > r \geq r^{q1,l} & if\ \bar{e}_{q,t}^{a,h} \leq \bar{e}_{q,t} < \bar{e}_{q,t}^{a,l} \\ r^{q0,h} > r \geq r^{q0,l}\ OR\ r^{q1,h} < r \leq r^{q1,l} & if\ \bar{e}_{q,t} > \bar{e}_{q,t}^{a,l}\ AND\ \bar{e}_{q,t}^{b,h} \leq \bar{e}_{q,t} < \bar{e}_{q,t}^{b,l} \end{cases} \tag{33}$$

Quality effort is zero for all ability classes if

$$\begin{cases} r \geq r^{q1,h} & if\ \bar{e}_{q,t} < \bar{e}_{q,t}^{a,h} \\ r^{q0,h} \leq r \leq r^{q1,h} & if\ \bar{e}_{q,t} > \bar{e}_{q,t}^{a,l}\ AND\ \bar{e}_{q,t} < \bar{e}_{q,t}^{b,h} \end{cases} \tag{34}$$

Here, r^{q0} is defined in (38) and r^{q1} in (37). $r^{q0,l}$ is equal to r^{q0} evaluated at $d\left(m_0\right) = d\left(\underline{m}\right)$, while $r^{q0,h}$ is r^{q0} evaluated at $d\left(m_0\right) = d\left(\overline{m}\right)$, and equivalently for $r^{q1,l}$ and $r^{q1,h}$. Furthermore, $\bar{e}_{q,t}^{a,h} = CC_t\frac{d(\overline{m})\sigma_\nu^2-\sigma_\varepsilon^2}{1-d(\overline{m})^2} < CC_t\frac{d(\underline{m})\sigma_\nu^2-\sigma_\varepsilon^2}{1-d(\underline{m})^2} = \bar{e}_{q,t}^{a,l}$ and $\bar{e}_{q,t}^{b,h} = \frac{d(\overline{m})\sigma_\nu^2-\sigma_\varepsilon^2}{(1+d(\overline{m}))(2(1+d(\overline{m}))\sigma_t^2+\sigma_\varepsilon^2+\sigma_\nu^2)} < \frac{d(\underline{m})\sigma_\nu^2-\sigma_\varepsilon^2}{(1+d(\underline{m}))(2(1+d(\underline{m}))\sigma_t^2+\sigma_\varepsilon^2+\sigma_\nu^2)} = \bar{e}_{q,t}^{b,l}$. The above conditions characterize the sign of quality effort for all possible values of risk aversion r and minimum quality effort level $\bar{e}_{q,t}$. They are thus necessary and sufficient.

Proof: From (25) we have that $e_{q,t}^* > 0$ if and only if

$$\bar{e}_{q,t} + \frac{1 + r\left(\sigma_\varepsilon^2 - d\sigma_\nu^2\right)\left(1 + CC_t\left(1+d\right)r\widetilde{CV_t}\right)}{\left(1+d\right)\left(1 + r\left(2\left(1+d\right)\sigma_t^2 + \sigma_\varepsilon^2 + \sigma_\nu^2\right) + r^2\left(1-d^2\right)\widetilde{CV_t}\right)} > 0 \tag{35}$$

It is immediate that this inequality holds for all $r > 0$ when $d < \sigma_\varepsilon^2/\sigma_\nu^2$ (sufficient, not necessary). It may not hold for all r when $d > \sigma_\varepsilon^2/\sigma_\nu^2$, since (25) decreases with d, $\forall d : 0 < d < 1$ (see proof of proposition

2). Define $\breve{r}\left(\breve{d}\right) > 0$ such that (35) holds with equality at $\left(\breve{r}, \breve{d}\right)$:

$$\begin{aligned}\breve{r}\left(\breve{d}\right) = r\left(\breve{d}\right) > 0 \; s.t. : 0 = r^2\left(\bar{e}_{q,t}\left(1+\breve{d}\right)\left(1-\breve{d}^2\right)\widetilde{CV}_t + \left(\sigma_\varepsilon^2 - \breve{d}\sigma_\nu^2\right)CC_t\left(1+\breve{d}\right)\widetilde{CV}_t\right) \\ + r\left(\bar{e}_{q,t}\left(1+\breve{d}\right)\left(2\left(1+\breve{d}\right)\sigma_t^2 + \sigma_\varepsilon^2 + \sigma_\nu^2\right) + \left(\sigma_\varepsilon^2 - \breve{d}\sigma_\nu^2\right)\right) + 1 + \bar{e}_{q,t}\left(1+\breve{d}\right)\end{aligned} \tag{36}$$

The right-hand side of (36) is a quadratic equation in r, and $\breve{r}\left(\breve{d}\right)$ are real, positive root(s) of this equation at $\breve{d}$. Denote by a the coefficient that multiplies r^2, b the coefficient that multiplies r and c the constant $1 + \bar{e}_{q,t}\left(1+\breve{d}\right)$. Since (36) is a quadratic equation in r and c is always positive, there are no positive real roots when both a and b are nonnegative, so $e^*_{q,t} > 0, \forall r, m_0$ in this case. There is only one positive real root when a is negative and b is positive or negative, and there are two positive real roots when a is positive and b is negative.

When a is negative, the positive root of (36) is equal to:

$$\begin{aligned}r^{q1} \equiv \left[2(1+d)\left(\left(1-d^2\right)\bar{e}_{q,t} + \left(\sigma_\varepsilon^2 - d\sigma_\nu^2\right)CC_t\right)\widetilde{CV}_t\right]^{-1}\left\{\bar{e}_{q,t}(1+d)\left(2(1+d)\sigma_t^2 + \sigma_\varepsilon^2 + \sigma_\nu^2\right) + \sigma_\varepsilon^2 - d\sigma_\nu^2\right. \\ \left. + \sqrt{\left(\bar{e}_{q,t}(1+d)\left(2(1+d)\sigma_t^2 + \sigma_\varepsilon^2 + \sigma_\nu^2\right) + \sigma_\varepsilon^2 - d\sigma_\nu^2\right)^2 - 4(1+d)\left(1 + \bar{e}_{q,t} + d\bar{e}_{q,t}\right)\widetilde{CV}_t\left(\bar{e}_{q,t}\left(1-d^2\right) + \left(\sigma_\varepsilon^2 - d\sigma_\nu^2\right)CC_t\right)}\right\}\end{aligned} \tag{37}$$

When a is positive and b is negative, the larger of the two positive roots is given by (37), and the smaller root is given by:

$$\begin{aligned}r^{q0} \equiv \left[2(1+d)\left(\left(1-d^2\right)\bar{e}_{q,t} + \left(\sigma_\varepsilon^2 - d\sigma_\nu^2\right)CC_t\right)\widetilde{CV}_t\right]^{-1}\left\{\bar{e}_{q,t}(1+d)\left(2(1+d)\sigma_t^2 + \sigma_\varepsilon^2 + \sigma_\nu^2\right) + \sigma_\varepsilon^2 - d\sigma_\nu^2\right. \\ \left. - \sqrt{\left(\bar{e}_{q,t}(1+d)\left(2(1+d)\sigma_t^2 + \sigma_\varepsilon^2 + \sigma_\nu^2\right) + \sigma_\varepsilon^2 - d\sigma_\nu^2\right)^2 - 4(1+d)\left(1 + \bar{e}_{q,t} + d\bar{e}_{q,t}\right)\widetilde{CV}_t\left(\bar{e}_{q,t}\left(1-d^2\right) + \left(\sigma_\varepsilon^2 - d\sigma_\nu^2\right)CC_t\right)}\right\}\end{aligned} \tag{38}$$

The limit of $e^*_{q,t}$ for $r \to 0$ is $\frac{1}{1+d}$, which is positive. Because (36) has only one positive root when a is negative, it follows that $e^*_{q,t} > 0$ if $r < r^{q1}$ and $e^*_{q,t} = 0$ otherwise in this case. Furthermore, because $\partial e^*_{q,t}/\partial d < 0, \forall d : 0 < d < 1$ for $e^*_{q,t}$ defined in (25) (see Proof of Proposition 2), r^{q1} is smallest for the lowest ability class and largest for the highest ability class when a is negative; i.e. $r^{q1,l} = r^{q1}\left(d\left(\underline{m}\right)\right) < r^{q1,h} = r^{q1}\left(d\left(\overline{m}\right)\right)$. But then, when a is negative, $e^*_{q,t} > 0, \forall m_0$ if $r < r^{q1,l}$, while $e^*_{q,t} = 0, \forall m_0$ if $r > r^{q1,h}$.

Similarly, when a is positive and b negative, $e^*_{q,t} > 0$ if $r < r^{q0}$ or $r > r^{q1}$ and $e^*_{q,t} = 0$ otherwise, because the limit of $e^*_{q,t}$ for $r \to 0$ is $\frac{1}{1+d} > 0$ and there are two positive roots in this case (i.e. $e^*_{q,t}(r)$ is U-shaped). Furthermore, as $\partial e^*_{q,t}/\partial d < 0, \forall d : 0 < d < 1$ for $e^*_{q,t}$ defined in (25), r^{q0} is smallest for the lowest ability class and largest for the highest ability class ($r^{q0,l} = r^{q0}\left(d\left(\underline{m}\right)\right) < r^{q0,h} = r^{q0}\left(d\left(\overline{m}\right)\right)$), while r^{q1} is largest for the lowest ability class and smallest for the highest ability class ($r^{q1,l} = r^{q1}\left(d\left(\underline{m}\right)\right) > r^{q1,h} = r^{q1}\left(d\left(\overline{m}\right)\right)$) in this case. Thus, when a is positive and b negative, $e^*_{q,t} > 0, \forall m_0$ if $r < r^{q0,l}$ and $r > r^{q1,l}$, while $e^*_{q,t} = 0, \forall m_0$ if $r \geq r^{q0,h}$ and $r \leq r^{q1,h}$.

Finally, it is straigthforward to show that a is negative if $0 < \bar{e}_{q,t} < \bar{e}^a_{q,t} = CC_t\frac{d\sigma_\nu^2 - \sigma_\varepsilon^2}{1-d^2}$, and b is negative if $0 < \bar{e}_{q,t} < \bar{e}^b_{q,t} = \frac{d\sigma_\nu^2 - \sigma_\varepsilon^2}{(1+d)(2(1+d)\sigma_t^2 + \sigma_\varepsilon^2 + \sigma_\nu^2)}$. Note that both $\bar{e}^a_{q,t}$ and $\bar{e}^b_{q,t}$ decrease with ability class, so that $\bar{e}^{a,h}_{q,t} = CC_t\frac{d(\overline{m})\sigma_\nu^2 - \sigma_\varepsilon^2}{1-d(\overline{m})^2} < CC_t\frac{d(\underline{m})\sigma_\nu^2 - \sigma_\varepsilon^2}{1-d(\underline{m})^2} = \bar{e}^{a,l}_{q,t}$ and $\bar{e}^{b,h}_{q,t} = \frac{d(\overline{m})\sigma_\nu^2 - \sigma_\varepsilon^2}{(1+d(\overline{m}))(2(1+d(\overline{m}))\sigma_t^2 + \sigma_\varepsilon^2 + \sigma_\nu^2)} < \frac{d(\underline{m})\sigma_\nu^2 - \sigma_\varepsilon^2}{(1+d(\underline{m}))(2(1+d(\underline{m}))\sigma_t^2 + \sigma_\varepsilon^2 + \sigma_\nu^2)} = \bar{e}^{b,l}_{q,t}$. Hence a and b are negative for all ability classes when $\bar{e}_{q,t} < \bar{e}^{a,h}_{q,t}$ and $\bar{e}_{q,t} < \bar{e}^{b,h}_{q,t}$, while a and b are nonnegative for all ability classes when $\bar{e}_{q,t} \geq \bar{e}^{a,l}_{q,t}$ and $\bar{e}_{q,t} \geq \bar{e}^{b,l}_{q,t}$.

Three regions can now be distinguished for the sign of $e^*_{q,t}$. The first region is such that $e^*_{q,t} > 0, \forall m_0$. Given the above, this region is defined by $\bar{e}_{q,t} \geq \bar{e}^{a,l}_{q,t}$ and $\bar{e}_{q,t} \geq \bar{e}^{b,l}_{q,t}$, and $0 < r < r^{q1,l}$ when $0 < \bar{e}_{q,t} < \bar{e}^{a,l}_{q,t}$, and $r < r^{q0,l}$ or $r > r^{q1,l}$ when $\bar{e}_{q,t} > \bar{e}^{a,l}_{q,t}$ and $\bar{e}_{q,t} < \bar{e}^{b,l}_{q,t}$ (i.e. (32) holds). The first set of conditions ensures that both a and b are nonnegative for all ability classes, so that $e^*_{q,t} > 0, \forall r, m_0$. The second set of conditions ensures that a is negative for at least the lowest ability classes and r small enough so that (35) holds for all ability classes. The third set ensures that a is positive for all ability classes and b is negative for at least the lowest ability classes, while r is such that (35) holds for all ability classes.

The second region is such that $e^*_{q,t} > 0$ for some (high) ability classes and $e^*_{q,t} = 0$ for other (low ability) classes. This region is defined by $r^{q1,h} > r \geq r^{q1,l}$ when $\bar{e}^{a,h}_{q,t} \leq \bar{e}_{q,t} < \bar{e}^{a,l}_{q,t}$; and $r^{q0,h} > r \geq r^{q0,l}$ or $r^{q1,h} < r \leq r^{q1,l}$ when $\bar{e}_{q,t} > \bar{e}^{a,l}_{q,t}$ and $\bar{e}^{b,h}_{q,t} \leq \bar{e}_{q,t} < \bar{e}^{b,l}_{q,t}$ (i.e. (33) holds). The first set of conditions ensures that a is negative for low ability classes only, and risk aversion is such that (35) does not hold for (some of) these low ability classes. The second set of conditions ensures that a is positive and b negative for low ability classes only, and risk aversion is such that (35) does not hold for (some of) these low ability classes.

The third region, finally, is such that $e^*_{q,t} = 0, \forall m_0$. Again using the above results, this region is defined by $r \geq r^{q1,h}$ when $0 < \bar{e}_{q,t} < \bar{e}^{a,h}_{q,t}$, and $r \geq r^{q0,h}$ and $r \leq r^{q1,h}$ when $\bar{e}_{q,t} > \bar{e}^{a,l}_{q,t}$ and $0 < \bar{e}_{q,t} < \bar{e}^{b,h}_{q,t}$ (i.e. (34) holds). The first set of conditions ensures that a is negative for all ability classes and r large enough so that (35) does not hold for any ability class. The second set of conditions ensures that a is positive and b negative for all ability classes and that r is such that (35) does not hold for any ability class. This completes the proof of the lemma.

I can now characterize heterogeneity in the quantity effort response.

Proposition 3a - **Quantity Effort Heterogeneity**: *Quantity effort increases monotonically with ability class if risk aversion is low, such that*

$$\begin{cases} r \leq \hat{r} & if\ \sigma_\nu \leq \sqrt{2}\sqrt{\sigma^2_\varepsilon} \\ r \leq \frac{1}{\sigma^2_\nu - \sigma^2_\varepsilon} & if\ \sigma_\nu > \sqrt{2}\sqrt{\sigma^2_\varepsilon} \end{cases} \tag{39}$$

and (32) holds. Quantity effort is U-shaped in ability if risk aversion is high, such that

$$\begin{cases} r > \hat{r} & if\ \sigma_\nu > \sqrt{2}\sqrt{\sigma^2_\varepsilon} \\ r > \frac{1}{\sigma^2_\nu - \sigma^2_\varepsilon} & if\ \sigma_\nu \leq \sqrt{2}\sqrt{\sigma^2_\varepsilon} \end{cases} \tag{40}$$

and (32) or (33) is satisfied. Quantity effort is monotonically decreasing in ability if (34) holds.

Here $\hat{r}$ is the third root of the polynomial in (43). Since $\hat{r} = \frac{1}{\sigma^2_\nu - \sigma^2_\varepsilon}$ when $\sigma_\nu = \sqrt{2}\sqrt{\sigma^2_\varepsilon}$, the proposition specifies heterogeneous quantity effort responses for all possible risk aversion values, quality output noise and minimum quality effort level $\bar{e}_{q,t}$. They are thus necessary and sufficient..

Proof: The proof of the first part of the proposition entails showing that the derivative $\partial e^*_{p,t}/\partial d$ is negative within the low risk aversion ranges specified in (39) and (32). Because $\partial d/\partial m_0 < 0$, $\partial e^*_{p,t}/\partial m_0$ is then positive. The proof of the third part involves showing that the derivative $\partial e^*_{p,t}/\partial d$ is positive across all ability classes if (34) holds. Finally, the second part of the proposition is proven by showing that $\partial e^*_{p,t}/\partial d$

is negative for high ability classes and positive for low ability classes if (40) and (33) or (32) are met.

It is immediate from (27) that $\partial e^*_{p,t}/\partial d > 0$ - and thus $\partial e^*_{p,t}/\partial m_0 < 0$ - across ability classes for which $e^*_{q,t} = 0$, since $\bar{e}_{q,t} > 0$[46]. But then, it follows from Lemma 1 that $\partial e^*_{p,t}/\partial m_0 < 0$ across all ability classes when (34) is met, and across ability classes for which $e^*_{q,t} = 0$ when (33) holds. The former completes the proof of the third statement in the proposition. Conversely, for $\partial e^*_{p,t}/\partial m_0$ to be positive across all ability classes, $e^*_{q,t} > 0, \forall m_0$ is a necessary condition. It then follows from Lemma 1, that $\partial e^*_{p,t}/\partial m_0$ can be positive across all ability classes if (32) is satisfied, and for those ability classes for which $e^*_{q,t} > 0$ when (33) holds. The exact sign of the derivative of $e^*_{p,t}$ in these latter regions depends on (24). I turn to this next.

Write the expression for $e^*_{p,t}$ in (24) as the sum of the minimum effort level $\bar{e}_{p,t}$, a component that is driven by explicit incentives, and another that is driven by career concerns:

$$\bar{e}_{p,t} + \frac{1 + r\left(\sigma^2_\nu - d\sigma^2_\varepsilon\right)}{(1+d)\, D_t} + \frac{r\left(\sigma^2_\nu - d\sigma^2_\varepsilon\right) CC_t\,(1+d)\, CV_t}{(1+d)\, D_t}$$

Taking the derivative with respect to d then yields:

$$\partial e^*_{p,t}/\partial d = -\frac{(1+2r(2(1+d)\sigma_t^2+\sigma_\varepsilon^2+\sigma_\nu^2))+(1+d)r^2((1-d+2d^2)\sigma_\varepsilon^2+(1-3d)\sigma_\nu^2)CV_t+R_{t,p}}{(1+d)^2 D_t^2}$$
$$-\quad\frac{(rCC_tCV_t)\left(2r\sigma_t^2\sigma_\nu^2(1-dr\sigma_\nu^2)+r\sigma_\varepsilon^4(1+(1+d^2)r(\sigma_t^2+\sigma_\nu^2))+\sigma_\varepsilon^2\left(1+r(2\sigma_t^2+\sigma_\nu^2)+r^2\sigma_\nu^2\left((1-d)^2\sigma_t^2-2d\sigma_\nu^2\right)\right)\right)}{D_t^2} \tag{41}$$

where $R_{t,p} := \left(r^2\left(\sigma^4_\varepsilon + \sigma^4_\nu - \sigma^2_\varepsilon\sigma^2_\nu\left(2d + 3d^2 - 3\right) + (1+d)\,\sigma^2_t\left((3-5d)\,\sigma^2_\varepsilon + (5-3d)\,\sigma^2_\nu\right)\right)\right)$. The derivative is negative if both numerators are positive, since both fractions enter negatively and both denominators are always positive. The numerator of the first fraction is positive if:

$$\begin{cases} \begin{cases} r \le \hat{r} & OR \\ r \ge \frac{1}{2}\sqrt{\frac{-7\sigma_\varepsilon^2-9\sigma_\nu^2}{(\sigma_\varepsilon^2-\sigma_\nu^2)(\sigma_\varepsilon^2+\sigma_\nu^2)^2}} + \frac{1}{2(\sigma_\varepsilon^2+\sigma_\nu^2)} & AND\ d < \hat{d} \end{cases} & if\ \sigma_\nu \le \sqrt{2}\sqrt{\sigma^2_\varepsilon} \\ \begin{cases} r \le \frac{1}{2}\sqrt{\frac{-7\sigma_\varepsilon^2-9\sigma_\nu^2}{(\sigma_\varepsilon^2-\sigma_\nu^2)(\sigma_\varepsilon^2+\sigma_\nu^2)^2}} + \frac{1}{2(\sigma_\varepsilon^2+\sigma_\nu^2)} & OR \\ r \ge \hat{r}\ AND\ d < \hat{d} \end{cases} & if\ \sigma_\nu > \sqrt{2}\sqrt{\sigma^2_\varepsilon} \end{cases} \tag{42}$$

Here $\hat{r}$ is the third root of the below polynomial in r:

$$1 + \left(2\sigma^2_\varepsilon + 2\sigma^2_\nu\right) r + \left(\sigma^4_\varepsilon - 2\sigma^2_\varepsilon\sigma^2_\nu + \sigma^4_\nu\right) r^2 + \left(4\sigma^4_\varepsilon\sigma^2_\nu - 4\sigma^2_\varepsilon\sigma^4_\nu\right) r^3 \tag{43}$$

and $\hat{d}$ is the second root of the following polynomial in d:

$$\begin{aligned} t\sigma_0^2\sigma_\varepsilon^2\left(1+r\sigma_\varepsilon^2\left(2+r\sigma_\varepsilon^2\right)\right)+t\sigma_0^2\sigma_\nu^2+\sigma_\varepsilon^2\sigma_\nu^2\left(1+r\sigma_\varepsilon^2\left(2+r\sigma_\varepsilon^2\right)\right)+r\sigma_0^2\sigma_\varepsilon^2\sigma_\nu^2\left(4\left(1+t\right)+r\sigma_\varepsilon^2\left(3+4t+r\sigma_\varepsilon^2\left(1+t\right)\right)\right)+ \\ +2r\sigma_\nu^4\left(t\sigma_0^2+\sigma_\varepsilon^2\right)+r^2\sigma_0^2\sigma_\varepsilon^2\sigma_\nu^4\left(5+4t\right)+r^2\sigma_\varepsilon^4\sigma_\nu^4\left(3+r\sigma_\varepsilon^2+2r\sigma_0^2\left(1+t\right)\right)+r^2t\sigma_0^2\sigma_\nu^6+r^2\sigma_\varepsilon^2\sigma_\nu^6\left(1+r\left(\sigma_\varepsilon^2+\sigma_0^2\left(1+t\right)\right)\right)+ \\ \left(2r\sigma_0^2\sigma_\varepsilon^2\sigma_\nu^2\left(2+r\left(\sigma_\nu^2\left(1-t\right)-\sigma_\varepsilon^2\left(1+t\right)\right)\right)-2r^2\sigma_\varepsilon^4\sigma_\nu^4\left(1-r\sigma_0^2\left(1+t\right)\right)-2r^3\sigma_\varepsilon^2\sigma_\nu^6\left(\sigma_\varepsilon^2+\sigma_0^2\left(1+t\right)\right)\right)d+ \\ \left(r^2\sigma_0^2\sigma_\varepsilon^4\sigma_\nu^2\left(r\sigma_\varepsilon^2\left(1+t\right)-\left(5+3t\right)\right)-r^2\sigma_\varepsilon^2\sigma_\nu^4\left(\sigma_\varepsilon^2\left(3-r\sigma_\varepsilon^2\right)+\sigma_0^2\left(1+t\right)\left(3+2r\sigma_\varepsilon^2\right)\right)-3r^3\sigma_\varepsilon^2\sigma_\nu^6\left(\sigma_0^2\left(1+t\right)+\sigma_\varepsilon^2\right)\right)d^2+ \\ \left(2r^3\sigma_\varepsilon^4\sigma_\nu^2\left(\sigma_0^2\left(\sigma_\varepsilon^2+\sigma_\nu^2\right)\left(1+t\right)+\sigma_\varepsilon^2\sigma_\nu^2\right)\right)d^3 \end{aligned} \tag{44}$$

[46] If the minimum effort level $\bar{e}_{q,t}$ increases with ability type, we would also need $\bar{e}_{q,t} + d\frac{\partial \bar{e}_{q,t}}{\partial d} > 0$ in order for $\partial e^*_{p,t}/\partial d > 0, \forall m_0$.

Since rCC_tCV_t is always positive, the numerator of the second fraction in (41) is positive if:

$$\begin{cases} \left\{ r \leq \frac{1}{\sigma_\nu^2 - \sigma_\varepsilon^2} \right\} & OR \\ \begin{cases} r > \frac{1}{\sigma_\nu^2 - \sigma_\varepsilon^2} & AND \\ d < \frac{\sigma_\nu^2}{\sigma_\varepsilon^2} - & \sqrt{\frac{M - t\sigma_0^2(\sigma_\varepsilon^2+\sigma_\nu^2)(\sigma_\varepsilon^2(1+r\sigma_\nu^2)(1+r\sigma_\varepsilon^2) - r^2\sigma_\nu^6)}{r^2\sigma_\varepsilon^4\sigma_\nu^2(\sigma_\varepsilon^2\sigma_\nu^2 + (1+t)\sigma_0^2(\sigma_\varepsilon^2+\sigma_\nu^2))}} \end{cases} \end{cases} \tag{45}$$

where $M = (\sigma_\nu^2 \left(r^2\sigma_0^2\sigma_\nu^6 - r\sigma_\varepsilon^6 \left(1 + r\left(\sigma_0^2 + \sigma_\nu^2\right)\right) - \sigma_\varepsilon^4 \left(1 + r^2\sigma_0^2\sigma_\nu^2 + r\left(2\sigma_0^2 + \sigma_\nu^2\right)\right) + r\sigma_\varepsilon^2\sigma_\nu^2 \left(r\sigma_\nu^4 + \sigma_0^2\left(r\sigma_\nu^2 - 2\right)\right)\right))$.

The upperbound constraint on r in the first line in (42) is smaller (more binding) than the upperbound in the first line in (45) for $\sigma_\nu \leq \sqrt{2}\sqrt{\sigma_\varepsilon^2}$. Hence $r \leq \hat{r}$ is a sufficient condition for $\partial e_{p,t}^*/\partial d$ to be negative when $\sigma_\nu \leq \sqrt{2}\sqrt{\sigma_\varepsilon^2}$. When $\sigma_\nu > \sqrt{2}\sqrt{\sigma_\varepsilon^2}$, the upperbound constraint on r in the first line in (45) is more binding than the upperbound in the third line in (42). Thus $r \leq \frac{1}{\sigma_\nu^2 - \sigma_\varepsilon^2}$ is a sufficient condition for $\partial e_{p,t}^*/\partial d$ to be negative - and hence $\partial e_{p,t}^*/\partial m_0 > 0$ - when $\sigma_\nu > \sqrt{2}\sqrt{\sigma_\varepsilon^2}$. Together with the conditions for $e_{q,t}^* > 0, \forall m_0$ specified in (32), this completes the proof of the first statement in the proposition.

When $\sigma_\nu > \sqrt{2}\sqrt{\sigma_\varepsilon^2}$, $\hat{r} > \frac{1}{\sigma_\nu^2 - \sigma_\varepsilon^2}$ and $\hat{r} > \frac{1}{2}\sqrt{\frac{-7\sigma_\varepsilon^2 - 9\sigma_\nu^2}{(\sigma_\varepsilon^2 - \sigma_\nu^2)(\sigma_\varepsilon^2+\sigma_\nu^2)^2}} + \frac{1}{2(\sigma_\varepsilon^2+\sigma_\nu^2)}$. A sufficient condition for $\partial e_{p,t}^*/\partial d > 0$ when both $\sigma_\nu > \sqrt{2}\sqrt{\sigma_\varepsilon^2}$ and $r > \frac{1}{2}\sqrt{\frac{-7\sigma_\varepsilon^2 - 9\sigma_\nu^2}{(\sigma_\varepsilon^2 - \sigma_\nu^2)(\sigma_\varepsilon^2+\sigma_\nu^2)^2}} + \frac{1}{2(\sigma_\varepsilon^2+\sigma_\nu^2)}$ is then:

$$r \geq \hat{r} \; AND \; d < \hat{d} \; AND \; d < \frac{\sigma_\nu^2}{\sigma_\varepsilon^2} - \sqrt{\frac{M - t\sigma_0^2\left(\sigma_\varepsilon^2 + \sigma_\nu^2\right)\left(\sigma_\varepsilon^2\left(1 + r\sigma_\nu^2\right)\left(1 + r\sigma_\varepsilon^2\right) - r^2\sigma_\nu^6\right)}{r^2\sigma_\varepsilon^4\sigma_\nu^2\left(\sigma_\varepsilon^2\sigma_\nu^2 + (1+t)\,\sigma_0^2\left(\sigma_\varepsilon^2 + \sigma_\nu^2\right)\right)}} \tag{46}$$

Similarly, when $\sigma_\nu \leq \sqrt{2}\sqrt{\sigma_\varepsilon^2}$, $\frac{1}{\sigma_\nu^2 - \sigma_\varepsilon^2} \geq \frac{1}{2}\sqrt{\frac{-7\sigma_\varepsilon^2 - 9\sigma_\nu^2}{(\sigma_\varepsilon^2 - \sigma_\nu^2)(\sigma_\varepsilon^2+\sigma_\nu^2)^2}} + \frac{1}{2(\sigma_\varepsilon^2+\sigma_\nu^2)} \geq \hat{r}$. A sufficient condition for $\partial e_{p,t}^*/\partial d > 0$ when both $\sigma_\nu \leq \sqrt{2}\sqrt{\sigma_\varepsilon^2}$ and $r \geq \hat{r}$ is then:

$$r \geq \frac{1}{\sigma_\nu^2 - \sigma_\varepsilon^2} \; AND \; d < \hat{d} \; AND \; d < \frac{\sigma_\nu^2}{\sigma_\varepsilon^2} - \sqrt{\frac{M - t\sigma_0^2\left(\sigma_\varepsilon^2 + \sigma_\nu^2\right)\left(\sigma_\varepsilon^2\left(1 + r\sigma_\nu^2\right)\left(1 + r\sigma_\varepsilon^2\right) - r^2\sigma_\nu^6\right)}{r^2\sigma_\varepsilon^4\sigma_\nu^2\left(\sigma_\varepsilon^2\sigma_\nu^2 + (1+t)\,\sigma_0^2\left(\sigma_\varepsilon^2 + \sigma_\nu^2\right)\right)}} \tag{47}$$

On the other hand, it can be shown that $\partial e_{p,t}^*/\partial d$ is positive if

$$d > \hat{d} \; AND \; d > \frac{\sigma_\nu^2}{\sigma_\varepsilon^2} - \sqrt{\frac{M - t\sigma_0^2\left(\sigma_\varepsilon^2 + \sigma_\nu^2\right)\left(\sigma_\varepsilon^2\left(1 + r\sigma_\nu^2\right)\left(1 + r\sigma_\varepsilon^2\right) - r^2\sigma_\nu^6\right)}{r^2\sigma_\varepsilon^4\sigma_\nu^2\left(\sigma_\varepsilon^2\sigma_\nu^2 + (1+t)\,\sigma_0^2\left(\sigma_\varepsilon^2 + \sigma_\nu^2\right)\right)}} \tag{48}$$

when either $\sigma_\nu > \sqrt{2}\sqrt{\sigma_\varepsilon^2}$ and $r > \hat{r}$ or when $\sigma_\nu \leq \sqrt{2}\sqrt{\sigma_\varepsilon^2}$ and $r > \frac{1}{\sigma_\nu^2 - \sigma_\varepsilon^2}$.

Note that the d thresholds specified in (46) are the same as those in (47) and (48). Taken together then, if risk aversion is high, such that $r > \hat{r}$ when $\sigma_\nu > \sqrt{2}\sqrt{\sigma_\varepsilon^2}$ or $r > \frac{1}{\sigma_\nu^2 - \sigma_\varepsilon^2}$ when $\sigma_\nu \leq \sqrt{2}\sqrt{\sigma_\varepsilon^2}$, $\partial e_{p,t}^*/\partial d$ is positive for low ability academics, for whom d is above the bounds specified in (48), while $\partial e_{p,t}^*/\partial d$ is negative for high ability academics, for whom d is below those same bounds. Hence quantity effort is U-shaped relative to ability class in this case. Together with the conditions for $e_{q,t}^* > 0$ for all ability classes or low ability classes only specified in (32) and (33) respectively, this completes the proof of the second statement in the proposition. ■

Proposition 3b - Quality Effort Heterogeneity: *If quality effort is positive for all ability classes under* *performance pay (i.e. if (32) holds), the change in quality effort relative to flat wage pay increases monotonically with ability. If quality effort is positive for high ability but zero for low ability classes (i.e. if (33) holds), the change in quality effort is either J-shaped or U-shaped in ability. If quality effort is zero for all ability classes (i.e. if (34) holds), the change in quality effort is either constant or decreasing in ability class.*

Corollary 2 - Heterogeneous Quality Effort Response: *Unless quality effort is zero across all ability classes, the most able academics increase quality effort the most or decrease it the least in response to performance pay, while lower ability academics display a smaller increase or greater decrease. The greatest decrease in quality effort occurs in intermediate ability classes ($\underline{m} < m_0 = \hat{m} < \overline{m}$) if (49) holds.*

Proof: Three cases, the first two with two sub-cases each, can be distinguished for the change in quality effort relative to flat wage pay:

1. Case 1: (32) holds so that $e^*_{q,t} > 0, \forall m_0$ (See Lemma 1). In this case, quality effort increases monotonically with ability. This follows from the sign of (31). In particular, since $\partial d/\partial m_0 < 0$ and $\partial \tilde{e}^*_{q,t}/\partial d < 0, \forall d : 0 < d < 1$, we have that $\partial e^*_{q,t}/\partial m_0 > 0$ when $e^*_{q,t} > 0$. But then, if the lowest ability class exerts positive quality effort, all other ability classes exert more quality effort. Two sub-cases can be distinguished, both of which follow directly from the result in Proposition 2 that $e^*_{q,t} > \bar{e}_{q,t}$ if $0 < r < r^{pp}$, or else for $m_0 > m^{pp}$.

   a) If $\underline{m} > m^{pp}$ or $0 < r < r^{pp}$, quality effort is higher than under flat wage pay for all ability classes. The quality effort increase is therefore largest for high ability and smallest for low ability academics.

   b) If $\overline{m} > m^{pp} > \underline{m}$ and $r > r^{pp}$, only high ability academics increase quality effort relative to flat wage pay, while low ability academics decrease quality effort.

2. Case 2: (33) holds, so that $e^*_{q,t} > 0$ for $\underline{m} < m_0 \leq \overline{m}$ (See proof of Lemma 1). The quality effort response is U-shaped in ability if the following condition holds:

$$\begin{aligned} 0 \leq &\bar{e}_{q,t}(\hat{m}) + \frac{1 + r\left(\sigma^2_\varepsilon - d(\hat{m})\,\sigma^2_\nu\right)\left(1 + CC_t\left(1 + d(\hat{m})\right)CV_t\right)}{\left(1 + d(\hat{m})\right)D_t\left(d(\hat{m})\right)} \; AND \\ \bar{e}_{q,t}(\underline{m}) &< -\left(\frac{1 + r\left(\sigma^2_\varepsilon - d(\hat{m})\,\sigma^2_\nu\right)\left(1 + CC_t\left(1 + d(\hat{m})\right)CV_t\right)}{\left(1 + d(\hat{m})\right)D_t\left(d(\hat{m})\right)}\right) \leq \bar{e}_{q,t}(\hat{m}) \end{aligned} \tag{49}$$

   Otherwise, the response is weakly J-shaped, with the lowest ability class(es) exerting no quality effort, and all higher ability classes exerting positive quality effort. This follows from the fact that $\partial e^*_{q,t}/\partial m_0 > 0$ when $e^*_{q,t} > 0$, and the assumption that effort cannot be negative. Because equilibrium quality effort of the lowest ability class is a corner solution under performance pay ($e^*_{q,t}(\underline{m}) = 0$) in this case, quality effort decreases by $\bar{e}_{q,t}(\underline{m})$ relative to flat wage pay for the lowest ability class. If (49) holds, there exist ability class(es) $\hat{m}$ for which quality effort under performance pay is an interior solution (first line), and for which the difference between quality effort under performance pay and flat wage pay is larger than the difference between quality effort under performance pay and under flat wage pay for the lowest ability agents (second line). When this is fulfilled, the greatest decrease in quality effort occurs in agents in intermediate ability classes

($\underline{m} < m_0 = \hat{m} < \overline{m}$) rather than those of lowest ability ($m_0 = \underline{m}$). Two sub-cases can be distinguished, which again follow directly from the result in Proposition 2 that $e^*_{q,t} > \bar{e}_{q,t}$ if $0 < r < r^{pp}$, or else for $m_0 > m^{pp}$:

   a) If $\overline{m} > m^{pp} > \underline{m}$ and $r > r^{pp}$, quality effort increases relative to flat wage levels for high ability academics only, and decreases for low ability academics.

   b) If $\overline{m} < m^{pp}$ and $r > r^{pp}$, all ability classes decrease quality effort relative to flat wage pay.

3. Case 3: (34) holds, so that $e^*_{q,t} = 0, \forall m_0$ (See Lemma 1). It is immediate that the quality effort decreases for all ability classes in this case. If $\bar{e}_{q,t}$ is constant across ability classes, the decrease in quality effort is the same for all abililty classes. If $\frac{\partial \bar{e}_{q,t}(m_0)}{\partial m_0} > 0$, the change in quality effort decreases (is larger negative) in ability class.

The proposition follows directly from the two cases, the corollary from the four sub-cases. ■

A necessary condition for the last line in (49) to be fulfilled is for $\bar{e}_{q,t}(\underline{m}) < \bar{e}_{q,t}(\hat{m})$. That is, the minimum quality effort level of the lowest ability class has to be smaller than the minimum quality effort level of intermediate ability class $\hat{m}$. This is fulfilled if, for instance, lower ability classes have lower tenure requirements or weaker intrinsic motivation for output quality.

A1.3.3 Selection Effect

Finally, I derive hypotheses about selection. I define positive and negative selection into performance pay in line with Lazear (2000).

Definition - Selection: *Selection into performance pay is positive if all agents with ability m_0 above some threshold level $m^{s,pp} : \overline{m} > m^{s,pp} > \underline{m}$ switch to performance pay from a flat wage, while agents with ability $m_0 \leq m_0^{s,pp}$ do not. Selection is negative if these inequalities are reversed.*

Proposition 4 - Selection Response: *Comparing the performance pay equilibrium characterized in Proposition 1 to the flat wage equilibrium derived in section A1.1, selection into performance pay is positive when risk aversion is low ($r \to 0$). Selection is ambiguous (positive, negative or neither) when risk aversion is high, such that $r \to \infty$ or*

$$\begin{cases} r > r^{q1,l} & if\ \bar{e}_{q,t} < \bar{e}^{a,l}_{q,t} \\ r > r^{q0,l} & if\ \bar{e}_{q,t} > \bar{e}^{a,l}_{q,t}\ AND\ \bar{e}_{q,t} < \bar{e}^{b,l}_{q,t} \end{cases} \tag{50}$$

Here, $r^{q0}, r^{q1}, \bar{e}^{[a,b],[l,h]}_{q,t}$ are defined as in Lemma 1.

Proof: Agents switch to performance pay if they prefer performance pay in t over a flat wage[47]. This happens if the certainty equivalent of the expected lifetime utility in the former system (CE^{pp}_t) exceeds that in the latter (CE^{fw}_t) as of period s, that is:

$$CE^{pp}_s(m_0) = \sum_{t=s}^{\infty} \left\{ \left(m_t + e^*_{p,t}\right) + \left(m_t + e^*_{q,t}\right) - c\left(\overrightarrow{e^*_t}\right) - \frac{r}{2}\left(\overrightarrow{b^*_t}^2 \Sigma^2_t\right) \right\} > \sum_{t=s}^{\infty} \{w_t\} = CE^{fw}_s\left([w_t]^{\infty}_{t=s}\right) \tag{51}$$

[47] I assume *wlog* that agents who are indifferent between performance pay and a flat wage remain in the flat wage system.

The derivative of the right-hand side of (51) with respect to ability class m_0 is 0. If there is an ability class $m_0 = m^{s,pp}$ such that agents in this ability class are indifferent between a flat wage and performance pay, then all agents with $m_0 \leq m^{s,pp}$ prefer a flat wage and all agents with $m_0 > m^{s,pp}$ prefer performance pay if the derivative with respect to ability class m_0 of the left-hand side of (51) is positive for all $m_0 \geq m^{s,pp}$ (sufficient, not necessary).

From Lemma 1, we have that $e^*_{q,t} > 0, \forall (m_0, t)$ when risk aversion is low, such that $r \leq r^{q1,l}$ and $r \leq r^{q0,l}$, or the minimum effort level is large, such that $\bar{e}_{q,t} > \bar{e}^{a,l}_{q,t}$ and $\bar{e}_{q,t} > \bar{e}^{b,l}_{q,t}$ (sufficient, not necessary, cf. (32)). Using (22)-(25), the certainty equivalent of the expected per period utility under performance pay when $e^*_{q,t} > 0$ approaches $2m_t + \bar{e}_{p,t} + \bar{e}_{q,t} + \frac{1+d}{(1+d)^2}$ for $r \to 0$. By the sum law, the limit of $CE^{pp}_s(m_0)$ for $r \to 0$ is therefore:

$$\lim_{r\to 0}\left[CE^{pp}_s(m_0)\,|e^*_{q,t} > 0\right] = \sum_{t=s}^{\infty}\left\{2m_t + \bar{e}_{p,t} + \bar{e}_{q,t} + \frac{1+d}{(1+d)^2}\right\}$$

The derivative of this limit with respect to ability class m_0, is:

$$\frac{\partial}{\partial m_0}\lim_{r\to 0}\left[CE^{pp}_s(m_0)\,|e^*_{q,t} > 0\right] = \sum_{t=s}^{\infty}\left\{\left(\frac{2\sigma^2_\varepsilon\sigma^2_\nu}{\sigma^2_\varepsilon\sigma^2_\nu + t\sigma^2_0\left(\sigma^2_\varepsilon + \sigma^2_\nu\right)}\right) - \left(\frac{1}{1+d^2}\right)\frac{\partial d}{\partial m_0}\right\} \tag{52}$$

Since $\partial d/\partial m_0$ is negative, (52) is positive. This proves the first statement of the proposition.

Quality effort is also positive for all ability types when risk aversion is large, such that $r > r^{q1,l}$ when $\bar{e}_{q,t} > \bar{e}^{a,l}_{q,t}$. This follows from (32) in Lemma 1. It is straightforward to show that the derivative with respect to m_0 of the limit of $CE^{pp}_s(m_0)\,|e^*_{q,t} > 0$ for $r \to \infty$ comprises two parts. The first part is always positive, while the second is positive only when $d \geq \left(\sigma^2_\varepsilon/\sigma^2_\nu\right)$ or $CC_t < \frac{(1-d^2)(\sigma^2_\varepsilon+\sigma^2_\nu)}{(1+d^2)\sigma^2_\varepsilon\sigma^2_\nu - d(\sigma^4_\varepsilon+\sigma^4_\nu)}$. The latter is fulfilled for δ, σ^2_0 sufficiently small. When $r \to \infty$ and $\bar{e}_{q,t} > \bar{e}^{a,l}_{q,t}$, $CE^{pp}_s(m_0)$ therefore always increases with ability for low ability classes, but decreases with ability for high ability classes ($d < \left(\sigma^2_\varepsilon/\sigma^2_\nu\right)$), if agents are sufficiently patient or there is much uncertainty about ability. $CE^{pp}_s(m_0)$ can thus be upward sloping or inverse U-shaped. Depending on if and where the CE^{fw}_s curve intersects with $CE^{pp}_s(m_0)$, selection can then be positive, negative or neither. This proves the first part of the second statement of the proposition for the case when $\bar{e}_{q,t} > \bar{e}^{a,l}_{q,t}$.

To prove the second part of the second statement of the proposition, as well as the first part of the second statement for the case when $\bar{e}_{q,t} < \bar{e}^{a,l}_{q,t}$, note the following. If risk aversion is large ($r \geq r^{q1,l}$) when $\bar{e}_{q,t} < \bar{e}^{a,l}_{q,t}$, or if $r^{q0,l} \leq r \leq r^{q1,l}$ when $\bar{e}^{a,l}_{q,t} < \bar{e}_{q,t} < \bar{e}^{b,l}_{q,t}$, it follows from the proof of Lemma 1 that $e^*_{q,t} = 0$ for ability classes with $\breve{d}$ such that $r \geq \breve{r}\left(\breve{d}\right) \geq r^{q1,l}$ when $\bar{e}_{q,t} \leq \bar{e}^{a,\breve{d}}_{q,t} < \bar{e}^{a,l}_{q,t}$, or $r \geq \breve{r}\left(\breve{d}\right) \geq r^{q0,l}$ and $r \leq \breve{r}\left(\breve{d}\right) \leq r^{q1,l}$ when $\bar{e}^{a,l}_{q,t} < \bar{e}^{a,\breve{d}}_{q,t} \leq \bar{e}_{q,t} \leq \bar{e}^{b,\breve{d}}_{q,t} < \bar{e}^{b,l}_{q,t}$. Here $\breve{r}\left(\breve{d}\right)$ is defined in (36) and $\bar{e}^{[a,b],\breve{d}}_{q,t}$ is $\bar{e}^{[a,b]}_{q,t}$ evaluated at $d = \breve{d}$. Using (26) and (27), the derivative of the left-hand side of (51) for these ability classes, is given by:

$$\frac{\partial}{\partial m_0}\left[CE^{pp}_s(m_0)\,|e^*_{q,t} = 0\right] = \sum_{t=s}^{\infty}\left\{\left(\frac{2\sigma^2_\varepsilon\sigma^2_\nu}{\sigma^2_\varepsilon\sigma^2_\nu + t\sigma^2_0\left(\sigma^2_\varepsilon + \sigma^2_\nu\right)}\right) + \bar{e}_{q,t}\left(1 + d\bar{e}_{q,t}\right)\frac{\partial d}{\partial m_0}\right\} \tag{53}$$

The second term on the right-hand side is negative, since $\partial d/\partial m_0$ is negative, and it decreases in both

$\bar{e}_{q,t}$ and d. When (34) holds, so that $e^*_{q,t} = 0$ for all ability classes, $CE^{pp}_s(m_0)$ can then be increasing, decreasing or U-shaped in ability class. If $\bar{e}_{q,t}$ is sufficiently small, such that (53) is positive for $m_0 = \underline{m}$, we have that $\partial CE^{pp}_s(m_0)/\partial m_0 > 0, \forall m_0$. But then, if $\underline{m} < m^{s,pp} < \overline{m}$, there is positive selection. Conversely, if $\bar{e}_{q,t}$ is so large that (53) is negative for $\overline{m}$ - and hence for all m_0 - $CE^{pp}_s(m_0)$ is monotonically decreasing in m_0. Consequently, there is negative selection, if there is an ability class $m^{s,pp}$. Finally, $CE^{pp}_s(m_0)$ is U-shaped if $\bar{e}_{q,t}$ is such that (53) is negative for d large, but positive for d small. Depending on if and where the CE^{fw}_s curve intersects with $CE^{pp}_s(m_0)$, selection can then be positive, negative or neither, and is thus ambiguous.

Lastly, when (33) holds, so that $e^*_{q,t} > 0$ for high ability classes, but $e^*_{q,t} = 0$ for low ability classes (see Lemma 1), selection is ambiguous as well. From the above, we have that (52) is positive for high ability classes if risk aversion is low ($r \to 0$), or if agents are sufficiently impatient or there is little uncertainty about ability when risk aversion is high ($r \to \infty$), and negative otherwise. At the same time, (53) is negative for low ability classes if $\bar{e}_{q,t}$ is sufficiently large, and positive otherwise. It follows that $CE^{pp}_s(m_0)$ can be monotonically increasing or decreasing, U-shaped or inverse U-shaped, and selection is therefore positive, negative or neither (i.e. ambiguous) if (33) holds. This completes the proof of the second part of the second statement. ■

These results carry over to general selection from risk-free outside options into academia or other labor markets featuring similar incentives. This is trivially true, since $w^*_t = \underline{u}_t$ and hence $CE^{fw}_t = \sum_{t=\tau}^{\infty}\left\{[w^*_t]_{t=\tau}^{\infty}\right\} = \sum_{t=\tau}^{\infty}\left\{[\underline{u}_t]_{t=\tau}^{\infty}\right\} = CE^o_t$, where CE^o_t denotes the certainty equivalent utility of an agent's outside option.

A1.4 Comparison with Implicit or Explicit Incentives Only

In this section, I compare the equilibrium behavior and comparative statics in the model with implicit, career concerns incentives plus explicit performance incentives (linear contracts) to the same in a model with career concern incentives only and a model with explicit incentives only.

A1.4.1 Career Concerns Only Model

In the absence of explicit incentive contracts, but all other assumptions as in section A1.2, optimal quantity and quality effort are defined by the following first order conditions when $\vec{e}_t \gg 0$ (cf. equations 15-17):

$$\frac{\partial g(\vec{e}_t)}{\partial e_{p,t}} = \left(e_{p,t} - \bar{e}_{p,t}\right) + d\left(e_{q,t} - \bar{e}_{q,t}\right) = \sigma^2_\nu CC_t \tag{54}$$

$$\frac{\partial g(\vec{e}_t)}{\partial e_{q,t}} = \left(e_{q,t} - \bar{e}_{q,t}\right) + d\left(e_{p,t} - \bar{e}_{p,t}\right) = \sigma^2_\varepsilon CC_t \tag{55}$$

Substituting one into the other and rearranging, equilibrium effort when $\vec{e}_t \gg 0$ is then:

$$e^*_{p,t} = \bar{e}_{p,t} + \frac{CC_t\left(\sigma^2_\nu - d\sigma^2_\varepsilon\right)}{1-d^2} \tag{56}$$

$$e^*_{q,t} = \bar{e}_{q,t} + \frac{CC_t\left(\sigma^2_\varepsilon - d\sigma^2_\nu\right)}{1-d^2} \tag{57}$$

It follows from the assumptions on $c(\vec{e}_t)$ and the fact that the right-hand sides depend on the model's parameters only that (56) and (57) define the unique optimal effort levels in each period when workers face career concerns only. Furthermore, since $\vec{e}_t^*$ does not depend on $\vec{\hat{e}}_t$, the market's conjectured effort levels are correct in equilibrium: $\vec{\hat{e}}_t = \vec{e}_t^*$.

It is immediate that (57) decreases with σ_ν^2 and converges to a finite negative limit for $\sigma_\nu^2 \to \infty$. But then, there exist $(\bar{e}_{q,t}, \sigma_\nu^2)$ such that at least the lowest ability classes exert no quality effort. Following the same steps as above, it is straightforward to show that $e_{p,t}^* = \bar{e}_{p,t} + d\bar{e}_{q,t} + CC_t\sigma_\nu^2$ when $e_{q,t}^* = 0$. It is immediate that these are the unique optimal effort levels when $e_{q,t}^* = 0$.

Define $\sigma_\nu^{q,l}$ and $\sigma_\nu^{q,h}$ as the quality noise levels at which the lowest and highest ability classes stop exerting quality effort, respectively:

$$\sigma_\nu^{q,l} : \left(1 - \left(d\left(\underline{m}\right)\right)^2\right)\bar{e}_{q,t} = \left(CC_t\left(\sigma_\nu^{q0}\right)\right)\left(\sigma_\varepsilon^2 - \left(d\left(\underline{m}\right)\right)\left(\sigma_\nu^{q0}\right)^2\right) \tag{58}$$

$$\sigma_\nu^{q,h} : \left(1 - \left(d\left(\overline{m}\right)\right)^2\right)\bar{e}_{q,t} = \left(CC_t\left(\sigma_\nu^{q1}\right)\right)\left(\sigma_\varepsilon^2 - \left(d\left(\overline{m}\right)\right)\left(\sigma_\nu^{q1}\right)^2\right) \tag{59}$$

Since the derivative with respect to d of $e_{q,t}^*$ in (57) is negative, $\sigma_\nu^{q,h} > \sigma_\nu^{q,l}$. It follows that $e_{q,t}^* > 0, \forall m_0$ if $\sigma_\nu < \sigma_\nu^{q,l}$, $e_{q,t}^* = 0, \forall m_0$ if $\sigma_\nu \geq \sigma_\nu^{q,h}$, and $e_{q,t}^* > 0$ for higher ability classes only ($m_0 > \underline{m}$) if $\sigma_\nu^{q,l} > \sigma_\nu > \sigma_\nu^{q,h}$. Equivalently, $e_{q,t}^* > 0, \forall m_0$ if $\bar{e}_{q,t} > \frac{CC_t(\sigma_\varepsilon^2 - d(\underline{m})\sigma_\nu^2)}{1-d(\underline{m})^2} = \bar{e}_{q,t}^{a,l}$, $e_{q,t}^* = 0, \forall m_0$ if $\bar{e}_{q,t} < \frac{CC_t(\sigma_\varepsilon^2 - (d(\overline{m}))\sigma_\nu^2)}{1-(d(\overline{m}))^2} = \bar{e}_{q,t}^{a,h}$, and $e_{q,t}^* > 0$ for higher ability classes only ($m_0 > \underline{m}$) if $\bar{e}_{q,t}^{a,h} < \bar{e}_{q,t} < \bar{e}_{q,t}^{a,l}$.

Given the assumptions that $0 < d < 1$ and $\sigma_\nu^2 > \sigma_\varepsilon^2$, it follows from (56) and the above that $e_{p,t}^* > \bar{e}_{p,t}, \forall m_0$. This is the same as in the model with career concerns and explicit performance incentives; in both models, quantity effort is larger than in the flat wage case for all ability types.

When it comes to quality effort, (57) implies that $e_{q,t}^* > \bar{e}_{q,t}$ for $m_0 > m^{cc}$, where m^{cc} is such that $d\left(m^{cc}\right) = \sigma_\varepsilon^2/\sigma_\nu^2$. That is, quality effort is larger for high ability classes only - regardless of risk aversion levels - in the model with career concerns only compared to flat wage pay. In contrast, we have from Proposition 2 that $e_{q,t}^* > \bar{e}_{q,t}, \forall m_0$ if $0 < r < r^{pp}$ in the model with career concerns plus explicit performance incentives. Furthermore, it is straightforward to show that $d\left(m^{pp}\right) > d\left(m^{cc}\right) = \sigma_\varepsilon^2/\sigma_\nu^2$ (cf. (28)). But then, if the quality measure is sufficiently noisy, so that $d\left(m^{cc}\right) = \sigma_\varepsilon^2/\sigma_\nu^2 < d\left(\overline{m}\right) < d\left(m^{pp}\right)$, no ability class increases quality effort in response to career concerns only relative to flat wage pay, while high ability classes do increase quality effort in response to a combination of career concerns and explicit performance incentives. Taken together, more ability classes increase quality effort in response to a combination of career concerns and explicit performance incentives compared to career concerns only, unless the quality measure is so noisy that no ability class increases quality effort in response to either incentive system (when $d\left(m^{cc}\right) < d\left(m^{pp}\right) < d\left(\overline{m}\right)$).

Turning to heterogeneity in effort responses, the derivatives of (56) and (57) with respect to d are, respectively:

$$\frac{\partial e_{p,t}^*}{\partial d} = \left[\bar{e}_{p,t}, \frac{2d\sigma_\nu^2 - (1+d^2)\,\sigma_\varepsilon^2}{(1-d^2)^2} CC_t\right] \tag{60}$$

$$\frac{\partial \tilde{e}_{q,t}^*}{\partial d} = \left[0, \frac{2d\sigma_\varepsilon^2 - (1+d^2)\,\sigma_\nu^2}{(1-d^2)^2} CC_t\right] \tag{61}$$

where the first argument in square brackets is the derivative of the respective effort dimension when $e_{q,t}^* = 0$. Output quantity decreases with ability type when $e_{q,t}^* = 0$, since $\bar{e}_{p,t} > 0$. When $e_{q,t}^* > 0$, $\partial e_{p,t}^*/\partial d$ is negative if and only if:

$$d < \frac{\sigma_\nu^2}{\sigma_\varepsilon^2} - \sqrt{\frac{\sigma_\nu^4}{\sigma_\varepsilon^4} - 1} \tag{62}$$

and positive otherwise. Hence quantity effort is smallest for academics of intermediate ability classes, for whom $d = (\sigma_\nu^2/\sigma_\varepsilon^2) - \sqrt{(\sigma_\nu^4/\sigma_\varepsilon^4) - 1}$ and higher for both higher and lower ability classes when $e_{q,t}^* > 0$.

Using (58) and (59), two cases can then be distinguished for heterogeneity in the quantity effort response. When $\sigma_\nu \geq \sigma_\nu^{q,h}$, quantity decreases monotonically with ability type, since $e_{q,t}^* = 0, \forall m_0$. When $\sigma_\nu < \sigma_\nu^{q,h}$, output quantity is U-shaped for either of two reasons. If $\sigma_\nu < \sigma_\nu^{q,l} < \sigma_\nu^{q,h}$, then $e_{q,t}^* > 0, \forall m_0$ and the U-shape follows from (62). If $\sigma_\nu^{q,l} < \sigma_\nu < \sigma_\nu^{q,h}$, then $e_{q,t}^* = 0$ and hence $\partial e_{p,t}^*/\partial d > 0$ for the lowest ability classes, while $e_{q,t}^* > 0$ for higher ability classes, for whom the sign of $\partial e_{p,t}^*/\partial d$ switches according to (62). Thus, while quantity effort responses increase monotonically with ability class under certain parameters (low risk aversion) in a system with career concerns and explicit performance incentives, there are no such regions under career concerns only.

It is immediate from the assumptions on $\sigma_\nu^2, \sigma_\varepsilon^2$ and d that the second argument in square brackets in (61) is always negative. This implies that $\partial \tilde{e}_{q,t}^*/\partial m_0 > 0$, since $\partial d/\partial m_0 < 0$. As in the model with career concerns and explicit performance incentives. three cases and four sub-cases can be distinguished for the change in quality effort relative to flat wage pay.

1. Case 1: $\sigma_\nu < \sigma_\nu^{q,l}$ so that $e_{q,t}^*(\underline{m}) > 0$, and hence $e_{q,t}^* > 0, \forall m_0$. In this case, quality effort increases monotonically with ability. This follows directly from the fact that $\partial \tilde{e}_{q,t}^*/\partial m_0 > 0$ when $e_{q,t}^* > 0$. Two sub-cases can be distinguished, both of which follow directly from the earlier result that $e_{q,t}^* > \bar{e}_{q,t}$ for $m_0 > m^{cc}$.

    a) If $\overline{m} > \underline{m} > m^{cc}$, quality effort is higher than under flat wage pay for all ability classes. The quality effort increase is therefore largest for high ability and smallest for low ability academics.

    b) If $\overline{m} > m^{cc} > \underline{m}$ on the other hand, only high ability academics increase quality effort relative to flat wage pay, while low ability academics decrease quality effort.

2. Case 2: $\sigma_\nu^{q,l} < \sigma_\nu < \sigma_\nu^{q,h}$ so that $e_{q,t}^*(\underline{m}) = 0$ but $e_{q,t}^*(\overline{m}) > 0$. The quality effort response is

U-shaped in ability if the following condition holds:

$$0 \leq \bar{e}_{q,t}\left(\hat{m}\right) + \frac{CC_t\left(\sigma_\varepsilon^2 - d\left(\hat{m}\right)\sigma_\nu^2\right)}{1-\left(d\left(\hat{m}\right)\right)^2}\ AND$$
$$\bar{e}_{q,t}\left(\underline{m}\right) \leq -\left(\frac{CC_t\left(\sigma_\varepsilon^2 - d\left(\hat{m}\right)\sigma_\nu^2\right)}{1-\left(d\left(\hat{m}\right)\right)^2}\right) \leq \bar{e}_{q,t}\left(\hat{m}\right) \tag{63}$$

Otherwise, the response is weakly J-shaped, with the lowest ability class(es) exerting no quality effort, and all higher ability classes exerting positive quality effort. This follows from the fact that $\partial\tilde{e}^*_{q,t}/\partial m_0 > 0$ when $e^*_{q,t} > 0$ and the assumption that effort cannot be negative, by the same logic as applied in the proof of proposition 3b. Two sub-classes can be distinguished, which again follow directly from the result that $e^*_{q,t} > \bar{e}_{q,t}$ for $m_0 > m^{cc}$:

a) If $\overline{m} > m^{cc} > \underline{m}$, quality effort increases relative to flat wage levels for high ability academics only, and decreases for low ability academics.

b) If $\overline{m} < m^{cc}$, all ability classes decrease quality effort relative to flat wage pay.

3. Case 3: $\sigma_\nu \geq \sigma_\nu^{q,h}$, so that $e^*_{q,t} = 0, \forall m_0$. The quality effort response is negative for all ability classes, with the change either constant or, if $\frac{\partial \bar{e}_{q,t}(m_0)}{\partial m_0} > 0$, decreasing in ability class.

The above cases and sub-cases, and the respective quality effort responses are the direct equivalent of those in the model with career concerns and explicit performance incentives.

Finally, I evaluate the selection effect. Agents prefer a career concerns only pay system in t over a flat wage if the certainty equivalent utility in the former system (CE_t^{cc}) exceeds that in the latter (CE_t^{fw}) as of period t:

$$CE_s^{cc}\left(m_0\right) = \sum_{t=s}^{\infty}\left\{\left(m_t + e^*_{p,t}\right) + \left(m_t + e^*_{q,t}\right) - c\left(\vec{e}_t\left(\vec{e}_t^*\right)\right)\right\} > \sum_{t=s}^{\infty}\left\{w_t\right\} = CE_s^{fw} \tag{64}$$

The right-hand side of the first equality follows from the assumption that there is perfect competition in the labor market and the fact that, without explicit performance contracts, there is no quadratic risk aversion cost. Using the same logic as in the proof of Proposition 4, it can be shown that selection can be positive, negative or ambiguous.

The derivative of CE_s^{fw} with respect to ability class m_0 is 0. If $\sigma_\nu < \sigma_\nu^{q,l}$, so that $e^*_{q,t} > 0, \forall m_0$, the derivative of CE_s^{cc} is:

$$\frac{\partial}{\partial m_0}\left[CE_s^{cc}\left(m_0\right)|e^*_{q,t} > 0\right] = \sum_{t=s}^{\infty}\left\{\left(\frac{2\sigma_\varepsilon^2\sigma_\nu^2}{\sigma_\varepsilon^2\sigma_\nu^2 + t\sigma_0^2\left(\sigma_\varepsilon^2+\sigma_\nu^2\right)}\right)\right. \tag{65}$$
$$\left.-\left\{\left(\frac{\left(CC_{tt}\left(1-d\left(2\left(1-d^2\right)+d^3\right)\right)\left(\sigma_\varepsilon^2+\sigma_\nu^2\right)\right) - CC_t^2\left(\left(1-d^4\right)\sigma_\varepsilon^2\sigma_\nu^2 - d\left(1-d^2\right)\left(\sigma_\varepsilon^4+\sigma_\nu^4\right)\right)}{\left(1-d^2\right)^3}\right)\right\}\frac{\partial d}{\partial m_0}\right]$$

Here we have used (56), (57), (60) and (61) to substitute for $e^*_{p,t}, e^*_{q,t}, \frac{\partial e^*_{p,t}}{\partial d}$ and $\frac{\partial e^*_{p,t}}{\partial d}$.

The term on the first line in (65) is always positive, while the second line is positive if $d \geq \sigma_\varepsilon^2/\sigma_\nu^2$ or $0 < CC_t < \left[\left(1-d^2\right)\left(\sigma_\varepsilon^2+\sigma_\nu^2\right)\right] / \left[\left(1+d^2\right)\sigma_\varepsilon^2\sigma_\nu^2 - d\left(\sigma_\varepsilon^4+\sigma_\nu^4\right)\right]$. The latter is fulfilled for δ, σ_0^2 small enough or t large enough, so there is positive selection into a career concerns-only pay system when agents are sufficiently impatient or there is little uncertainty about ability. Otherwise, $CE_s^{cc}(m_0)$ has an inverse-U shape, increasing with m_0 for low ability and decreasing for high ability classes. Depending on if and where the CE_s^{fw} curve intersects with $CE_s^{cc}(m_0)$, selection can then be positive, negative or neither. This is the same as selection into a system with career concerns and explicit performance incentives for the case when $e_{q,t}^* > 0, \forall m_0$ and risk aversion is high (cf. second part of Proposition 4), but contrasts with the unambiguously positive selection in the latter system when risk aversion is low ($r \to 0$).

Selection can be positive, negative or neither when $\sigma_\nu \geq \sigma_\nu^{q,l}$ as well. To see this, note that $e_{q,t}^* = 0$ for some $m_0 \geq \underline{m}$, when $\sigma_\nu \geq \sigma_\nu^{q,l}$. The left-hand side of (64) is then equal to:

$$\left[CE_s^{cc}(m_0)\,|e_{q,t}^* = 0\right] = \sum_{t=s}^{\infty}\left\{2m_t + \bar{e}_{p,t} + \bar{e}_{q,t}\left(d - \frac{1}{2}\bar{e}_{q,t}(1-d)\right) + \sigma_\nu^2 CC_t\left(1 - \frac{1}{2}\sigma_\nu^2 CC_t\right)\right\} \tag{66}$$

The corresponding derivative with respect to m_0 is:

$$\frac{\partial}{\partial m_0}\left[CE_s^{cc}(m_0)\,|e_{q,t}^* = 0\right] = \sum_{t=s}^{\infty}\left\{\left(\frac{2\sigma_\varepsilon^2\sigma_\nu^2}{\sigma_\varepsilon^2\sigma_\nu^2 + t\sigma_0^2\left(\sigma_\varepsilon^2 + \sigma_\nu^2\right)}\right) + \bar{e}_{q,t}\left(1 + d\bar{e}_{q,t}\right)\frac{\partial d}{\partial m_0}\right\} \tag{67}$$

This is the exact same as (53) - the derivative of $CE_s^{pp}(m_0)$ with respect to m_0 when $e_{q,t}^* = 0$ - and hence equivalent selection results ensue, but with conditions on σ_ν instead of r. In particular, when $\sigma_\nu \geq \sigma_\nu^{q,h}$, so that $e_{q,t}^* = 0, \forall m_0$, and $\bar{e}_{q,t}$ is sufficiently small, so that (67) is positive for all ability classes, selection is positive. When $\sigma_\nu \geq \sigma_\nu^{q,h}$ and $\bar{e}_{q,t}$ is sufficiently large, so that (67) is negative for all ability classes, selection is negative. When $\sigma_\nu \geq \sigma_\nu^{q,h}$ and $\bar{e}_{q,t}$ is such that (67) is negative for low ability and positive for high ability classes, $CE_s^{cc}(m_0)$ is U-shaped and selection is ambiguous. Finally, when $\sigma_\nu^{q,l} < \sigma_\nu < \sigma_\nu^{q,h}$, so that $e_{q,t}^* > 0$ for higher ability classes only, selection can be positive, negative or neither as well. To see this, note that the above implies that in this case (65) is either positive or - for δ, σ_0^2 large enough - negative for high ability classes, while (67) is either positive or - when $\bar{e}_{q,t}$ is sufficiently large - negative for low ability classes. Consequently, $CE_s^{cc}(m_0)$ can be monotonically increasing or decreasing, U-shaped or inverse U-shaped, and selection is therefore ambiguous.

All in all, while selection into a system with implicit and explicit incentives is unambiguously positive when risk aversion is low ($r \to 0$), and ambiguous otherwise, selection into a system with implicit performance incentives is ambiguous throughout.

The proposition below summarizes the differences in equilibrium behavior and comparative statics between the performance pay model with both explicit and career concerns incentives and the model with career concern incentives only.

Proposition 5 - **Comparison with Career Concerns Only**: *Comparing the performance pay equilibrium characterized in Proposition 1 to the career concerns only equilibrium derived above, the following differences can be seen:*

1. *Fewer ability classes increase quality effort relative to flat wage levels $\left(e_{q,t}^* > \bar{e}_{q,t}\right)$ in response to career concerns incentives only, unless no ability class increases quality effort in response to either incentive system (when $d(m^{cc}) < d(m^{pp}) < d(\overline{m})$).*

2. *There are no parameter value regions in which quantity effort increases monotonically with ability in response*

to career concerns incentives only.

3. *There are no parameter value regions in which selection into career concerns only is unambiguously positive.*

Proof: See derivation in section above. ■

Corollary 3 - Relative Prediction - Explicit Incentives Only: *Relative to a performance pay system with career concerns and explicit performance incentives, a system with career concerns incentives only is less likely to give rise to a positive quality effort effect or an unambiguously positive selection effect, and quantity effort is less likely to increase monotonically with ability.*

Proof: The corollary follows immediately from proposition 5, and the assumption that all parameter values that characterize the different regions for the effort effect, effort effect heterogeneity and selection effect in either pay system carry non-zero probability. ■

A1.4.2 Explicit Incentives Only Model

Now suppose there is no learning about ability. In particular, assume that ability is symmetrically known, while all other assumptions are as in the model with career concerns and explicit performance incentives (section A1.2). As a result, career concerns incentives are no longer at play, and only the explicit performance incentives of the performance contracts remain. Optimal quantity and quality effort are then defined by the following first order conditions when $\vec{e}_t \gg 0$ (cf. equations 15, 16 and 17):

$$\frac{\partial g\left(\vec{e}_t\right)}{\partial e_{p,t}} = \left(e_{p,t} - \bar{e}_{p,t}\right) + d\left(e_{q,t} - \bar{e}_{q,t}\right) = b_{p,t} \tag{68}$$

$$\frac{\partial g\left(\vec{e}_t\right)}{\partial e_{q,t}} = \left(e_{q,t} - \bar{e}_{q,t}\right) + d\left(e_{p,t} - \bar{e}_{p,t}\right) = b_{q,t} \tag{69}$$

Substituting one into the other and rearranging, optimal effort when $\vec{e}_t \gg 0$ is then:

$$e^*_{p,t} = \bar{e}_p + \frac{b_{p,t} - db_{q,t}}{1-d^2} \tag{70}$$

$$e^*_{q,t} = \bar{e}_p + \frac{b_{q,t} - db_{p,t}}{1-d^2} \tag{71}$$

To derive the optimal bonus $\overrightarrow{b^*_t}$, I again use that $\overrightarrow{b_t}$ affects only the effort and income risk in period t and is therefore chosen to maximize agent utility that period. This amounts to optimizing the certainty equivalent with respect to $\overrightarrow{b_t}$

$$\overrightarrow{b^*_t} = argmax_{b_t}\left\{\left(m_t + e^*_{p,t}\left(\overrightarrow{b_t}\right)\right) + \left(m_t + e^*_{q,t}\left(\overrightarrow{b_t}\right)\right) - g\left(\vec{e}^*_t\left(\overrightarrow{b_t}\right)\right) - \frac{r}{2}\left(\overrightarrow{b_t}^2 \Sigma^2_{t,ex}\right)\right\}$$

Here $\Sigma^2_{t,ex}$ is now given by $\Sigma^2_{t,ex} = \begin{bmatrix} \sigma^2_\varepsilon & 0 \\ 0 & \sigma^2_\nu \end{bmatrix}$. Substituting for $e^*_{p,t}$ and $e^*_{q,t}$ using (70) and (71), the first order conditions for optimal bonus rates when $\vec{e}_t \gg 0$ are:

$$\frac{1-d-\left(b^*_{p,t}-db^*_{q,t}\right)}{1-d^2}=rb^*_{p,t}\sigma^2_{\varepsilon} \tag{72}$$
$$\frac{1-d-\left(b^*_{q,t}-db^*_{p,t}\right)}{1-d^2}=rb^*_{q,t}\sigma^2_{\nu} \tag{73}$$

It is immediate that these optimal bonus rates do not vary with time, so the time subscripts can be dropped. Substituting one into the other and rearranging, we get

$$b^*_p=\frac{1+(1-d)\,r\sigma^2_{\nu}}{1+r\left(\sigma^2_{\varepsilon}+\sigma^2_{\nu}\right)+\left(1-d^2\right)r^2\sigma^2_{\varepsilon}\sigma^2_{\nu}} \tag{74}$$
$$b^*_q=\frac{1+(1-d)\,r\sigma^2_{\varepsilon}}{1+r\left(\sigma^2_{\varepsilon}+\sigma^2_{\nu}\right)+\left(1-d^2\right)r^2\sigma^2_{\varepsilon}\sigma^2_{\nu}} \tag{75}$$

It follows from the assumptions on $c\left(\vec{e}_t\right)$ and the fact that the right-hand sides depend on the model's parameters only, that (74) and (75) define the unique optimal bonuses in each period.

Finally, substituting (74) and (75) into (70) and (71) and rearranging, we get the following expressions for optimal effort when $\vec{e}_t \gg 0$:

$$e^*_{p,t}=\bar{e}_{p,t}+\frac{1+r\left(\sigma^2_{\nu}-d\sigma^2_{\varepsilon}\right)}{(1+d)\left(1+r\left(\sigma^2_{\varepsilon}+\sigma^2_{\nu}\right)+\left(1-d^2\right)r^2\sigma^2_{\varepsilon}\sigma^2_{\nu}\right)} \tag{76}$$
$$e^*_{q,t}=\bar{e}_{q,t}+\frac{1+r\left(\sigma^2_{\varepsilon}-d\sigma^2_{\nu}\right)}{(1+d)\left(1+r\left(\sigma^2_{\varepsilon}+\sigma^2_{\nu}\right)+\left(1-d^2\right)r^2\sigma^2_{\varepsilon}\sigma^2_{\nu}\right)} \tag{77}$$

These define the unique optimal effort levels by the same reasoning as used for optimal bonuses.

It is immediate from (77) that $e^*_{q,t}=0$ when

$$0=r^2\left(\bar{e}_{q,t}\left(1-d^2\right)\sigma^2_{\varepsilon}\sigma^2_{\nu}\right)+r\left(\bar{e}_{q,t}\,(1+d)\left(\sigma^2_{\varepsilon}+\sigma^2_{\nu}\right)+\left(\sigma^2_{\varepsilon}-d\sigma^2_{\nu}\right)\right)+(1+d)\,\bar{e}_{q,t}+1 \tag{78}$$

This quadratic equation in r has positive, real roots if and only if the coefficient of the linear term in r is negative, since the coefficient of the quadratic term and the constant are both always positive. The coefficient of the linear term in r is negative if $\bar{e}_{q,t}<\bar{e}^{ex}_{q,t}=\left(d\sigma^2_{\nu}-\sigma^2_{\varepsilon}\right)/\left(1+d\right)\left(\sigma^2_{\varepsilon}+\sigma^2_{\nu}\right)$. It is straightforward to show that $\frac{\partial\bar{e}^{ex}_{q,t}}{\partial d}>0$, so $\bar{e}^{ex,l}_{q,t}=\bar{e}^{ex}_{q,t}\left(d\left(\underline{m}\right)\right)>\bar{e}^{ex}_{q,t}\left(d\left(\overline{m}\right)\right)=\bar{e}^{ex,h}_{q,t}$. It follows that the coefficient of the linear term in r is negative for all ability classes if $\bar{e}_{q,t}<\bar{e}^{ex,h}_{q,t}$, positive for all ability classes if $\bar{e}_{q,t}\geq\bar{e}^{ex,l}_{q,t}$, and negative for low ability classes but positive for high ability classes if $\bar{e}^{ex,h}_{q,t}\leq\bar{e}_{q,t}<\bar{e}^{ex,l}_{q,t}$.

When the coefficient of the linear term in r is negative, (78) has two positive real roots:

$$r^{q0}_{ex}=\frac{d\sigma^2_{\nu}-\sigma^2_{\varepsilon}-(1+d)\,\bar{e}_{q,t}\left(\sigma^2_{\varepsilon}+\sigma^2_{\nu}\right)-\sqrt{\left(\sigma^2_{\varepsilon}-d\sigma^2_{\nu}+(1+d)\,\bar{e}_{q,t}\left(\sigma^2_{\varepsilon}+\sigma^2_{\nu}\right)\right)^2-4\bar{e}_{q,t}\left(1-d^2\right)\left(1+(1+d)\,\bar{e}_{q,t}\right)\sigma^2_{\varepsilon}\sigma^2_{\nu}}}{2\bar{e}_{q,t}\left(1-d^2\right)\sigma^2_{\varepsilon}\sigma^2_{\nu}} \tag{79}$$
$$r^{q1}_{ex}=\frac{d\sigma^2_{\nu}-\sigma^2_{\varepsilon}-(1+d)\,\bar{e}_{q,t}\left(\sigma^2_{\varepsilon}+\sigma^2_{\nu}\right)+\sqrt{\left(\sigma^2_{\varepsilon}-d\sigma^2_{\nu}+(1+d)\,\bar{e}_{q,t}\left(\sigma^2_{\varepsilon}+\sigma^2_{\nu}\right)\right)^2-4\bar{e}_{q,t}\left(1-d^2\right)\left(1+(1+d)\,\bar{e}_{q,t}\right)\sigma^2_{\varepsilon}\sigma^2_{\nu}}}{2\bar{e}_{q,t}\left(1-d^2\right)\sigma^2_{\varepsilon}\sigma^2_{\nu}} \tag{80}$$

It is immediate that $r^{q0}_{ex}<r^{q1}_{ex}$. The limit of $e^*_{q,t}$ for $r\to 0$ is $\frac{1}{1+d}$, which is positive. But then, given that (78) has two positive roots when the linear term in r is negative, $e^*_{q,t}>0$ if $r<r^{q0}_{ex}$ or $r>r^{q1}_{ex}$, while $e^*_{q,t} = 0$ otherwise. Because $\frac{\partial e^*_{q,t}}{\partial d} < 0$ for $e^*_{q,t}$ in (77), as is shown below, $r^{q0,l}_{ex} = r^{q0}_{ex}\left(d\left(\underline{m}\right)\right) < r^{q0}_{ex}\left(d\left(\overline{m}\right)\right) = r^{q0,h}_{ex}$. Similarly, $r^{q1,l}_{ex} = r^{q1}_{ex}\left(d\left(\underline{m}\right)\right) > r^{q1}_{ex}\left(d\left(\overline{m}\right)\right) = r^{q1,h}_{ex}$. Taken together, when the coefficient of the linear term in r in (78) is negative, $e^*_{q,t} > 0, \forall m_0$ if $r < r^{q0,l}_{ex}$ or $r > r^{q1,l}_{ex}$, while $e^*_{q,t} = 0, \forall m_0$ if $r^{q0,h} < r < r^{q1,h}$.

Using similar reasoning as in the proof of Lemma 1, three regions can now be distinguished for the sign of $e^*_{q,t}$. The first region is such that $e^*_{q,t} > 0, \forall m_0$. Given the above, this region is defined by

$$\bar{e}_{q,t} \geq \bar{e}^{ex,l}_{q,t} \; OR \; \left\{\left(r < r^{q0,l}_{ex} \; OR \; r > r^{q1,l}_{ex}\right) \; if \; 0 < \bar{e}_{q,t} < \bar{e}^{ex,l}_{q,t}\right\} \tag{81}$$

The second region is such that $e^*_{q,t} > 0$ for high ability classes, but $e^*_{q,t} = 0$ for low ability classes. This region is defined by

$$\left(r^{q0,h}_{ex} > r \geq r^{q0,l}_{ex} \; OR \; r^{q1,h}_{ex} < r \leq r^{q1,l}_{ex}\right) \; if \; \bar{e}^{ex,h}_{q,t} \leq \bar{e}_{q,t} < \bar{e}^{ex,l}_{q,t} \tag{82}$$

The third region is such that $e^*_{q,t} = 0, \forall m_0$. Using the above results, this region is defined by

$$r^{q0,h}_{ex} \leq r \leq r^{q1,h}_{ex} \; if \; 0 < \bar{e}_{q,t} < \bar{e}^{ex,h}_{q,t} \tag{83}$$

Comparing the above to Lemma 1, it follows that explicit performance incentives give rise to zero output quality across ability classes for more risk aversion levels than a system with implicit and explicit performance incentives. With explicit performance incentives, quality effort is zero across ability classes only when the level of minimum effort is low ($0 < \bar{e}_{q,t} < \bar{e}^{ex,h}_{q,t}$) and risk aversion falls in a limited, intermediate range ($r^{q0,h}_{ex} \leq r \leq r^{q1,h}_{ex}$). With explicit and implicit performance incentives, quality effort is zero across ability classes when both the level of minimum effort and risk aversion are intermediate ($\bar{e}_{q,t} > \bar{e}^{a,l}_{q,t}$ and $\bar{e}_{q,t} < \bar{e}^{b,l}_{q,t}$ and $r^{q0,h} \leq r \leq r^{q1,h}$), or when the level of minimum effort is low ($0 < \bar{e}_{q,t} < \bar{e}^{a,h}_{q,t}$) and risk aversion is sufficiently large ($r \geq r^{q1,h}$). There are thus more (high) risk aversion values for which $e^*_{q,t} = 0, \forall m_0$ in response to both explicit and career concerns incentives.

Following the same steps as above, it is straightforward to show that $b^*_{p,t} = \frac{1}{1+r\sigma^2_\varepsilon}$, $b^*_{q,t} = 0$, and $e^*_{p,t} = \bar{e}_{p,t} + d\bar{e}_{q,t} + \frac{1}{1+r\sigma^2_\varepsilon}$, when $e^*_{q,t} = 0$. It is immediate that these are the unique optimal bonus and effort levels when $e^*_{q,t} = 0$.

It follows from (76) and the assumptions that $0 < d < 1$ and $\sigma^2_\nu > \sigma^2_\varepsilon$ that $e^*_{p,t} > \bar{e}_{p,t}, \forall m_0$. Quantity effort increases relative to flat wage pay for all ability types, as in the model with career concerns and explicit performance incentives (cf. Proposition 2).

On the other hand, $e^*_{q,t} > \bar{e}_{q,t}$ for all ability classes if $r \leq r^{ex} = 1/\left(d\left(\underline{m}\right)\sigma^2_\nu - \sigma^2_\varepsilon\right)$, or else for $m_0 > m^{ex}$ only, where $m^{ex} : d\left(m^{ex}\right) = \left(1 + r\sigma^2_\varepsilon\right)/r\sigma^2_\nu$. The latter follows from the fact that (77) is zero at $d = d\left(m^{ex}\right)$ and decreases in d (shown below). The former derives from the fact that (77) is zero at $\tilde{r}^{ex}\left(d\right) = 1/\left(d\sigma^2_\nu - \sigma^2_\varepsilon\right)$, $\tilde{r}^{ex}\left(d\right)$ decreases in d, and (77) decreases in r when $d > \sigma^2_\varepsilon/\sigma^2_\nu$. If risk aversion is high, quality effort increases relative to flat wage pay for high ability classes, but decreases for low ability classes. When risk aversion is low, quality effort increases relative to flat wage pay for all ability types. This is equivalent to the quality effort response in the model with career concerns and explicit performance incentives.

The threshold risk aversion level below which $e^*_{q,t} > \bar{e}_{q,t}, \forall m_0$ is larger under explicit incentives than under career concerns plus explicit performance incentives ($r^{ex} > r^{pp}$). There are thus more risk aversion levels at which all ability classes increase quality effort in response to only explicit incentives. At the same time, for any $r > r^{ex}$, the threshold ability class above which quality effort increases is also smaller under explicit incentives only ($m^{pp} > m^{ex}$), provided output quality noise is not too small ($\sigma^2_\nu > 1$). This implies that more ability classes increase quality effort in response to explicit incentives only when risk aversion is high. By extension, there are fewer quality measure noise levels $\left(\sigma^2_\nu\right)$ at which no ability class increases quality effort relative to flat wage pay in response to only explicit incentives $\left(e^*_{q,t} < \bar{e}_{q,t}, \forall m_0\right)$.

Turning to heterogeneity in effort responses, $\frac{\partial e^*_{p,t}}{\partial d}$ and $\frac{\partial e^*_{q,t}}{\partial d}$ are, respectively:

$$\frac{\partial e^*_{p,t}}{\partial d} = \left[\bar{e}_{q,t}, -\frac{K + (1+d)\, r^3 \sigma^2_\varepsilon \sigma^2_\nu \left(\left(1 - d + 2d^2\right)\sigma^2_\varepsilon + (1-3d)\,\sigma^2_\nu\right)}{(1+d)^2 \left(1 + r\left(\sigma^2_\varepsilon + \sigma^2_\nu\right) + \left(1-d^2\right) r^2 \sigma^2_\varepsilon \sigma^2_\nu\right)^2}\right] \tag{84}$$

$$\frac{\partial e^*_{q,t}}{\partial d} = \left[0, -\frac{K + (1+d)\, r^3 \sigma^2_\varepsilon \sigma^2_\nu \left(\left(1 - d + 2d^2\right)\sigma^2_\nu + (1-3d)\,\sigma^2_\varepsilon\right)}{(1+d)^2 \left(1 + r\left(\sigma^2_\varepsilon + \sigma^2_\nu\right) + \left(1-d^2\right) r^2 \sigma^2_\varepsilon \sigma^2_\nu\right)^2}\right] \tag{85}$$

where $K = 1 + 2r\left(\sigma^2_\varepsilon + \sigma^2_\nu\right) + r^2\left(\sigma^4_\varepsilon + \sigma^4_\nu + \left(3 - 2d - 3d^2\right)\sigma^2_\varepsilon\sigma^2_\nu\right)$ and the first argument in square brackets is the derivative of the respective effort dimension when $e^*_{[p,q],t} = 0$.

It is immediate from (84) that $\frac{\partial e^*_{p,t}}{\partial d} > 0$ when $e^*_{q,t} = 0$, so that output quantity decreases with ability type in this case. When $e^*_{q,t} > 0$, it can be shown that $\frac{\partial e^*_{p,t}}{\partial d} < 0$, so that output quantity increases with ability type, if and only if:

$$\begin{gathered} 0 < d < \hat{d}^{ex} \; if \; r > \hat{r} \\ OR : 0 < d < 1 \; if \; r \leq \hat{r} \end{gathered} \tag{86}$$

Here $\hat{r}$ is the third root of the polynomial in r specified in (43), while $\hat{d}^{ex}$ is the second root of the following polynomial in d:

$$\begin{gathered} 1 + r\sigma^2_\varepsilon\left(2 + r\left(3\sigma^2_\nu + \sigma^2_\varepsilon\left(1 + r\sigma^2_\nu\right)\right)\right) + r\sigma^2_\nu\left(2 + r\sigma^2_\nu\left(1 + r\sigma^2_\varepsilon\right)\right) \\ -2r^2\sigma^2_\varepsilon\sigma^2_\nu\left(1 + r\sigma^2_\nu\right)d - r^2\sigma^2_\varepsilon\sigma^2_\nu\left(3 - r\sigma^2_\varepsilon + 3r\sigma^2_\nu\right)d^2 + 2r^3\sigma^4_\varepsilon\sigma^2_\nu d^3 \end{gathered} \tag{87}$$

The second line of (86) implies that quantity effort increases monotonically with ability class if risk aversion is sufficiently low. When risk aversion is high, quantity effort is smallest for academics of intermediate ability classes, for whom $d = \hat{d}^{ex}$, and larger for both higher and lower ability classes. The risk aversion threshold in the second line of (86) is the same as the risk aversion threshold in the model with career concerns plus explicit incentives for low levels of noise in the quality measure ($\sigma_\nu \leq \sqrt{2}\sqrt{\sigma^2_\varepsilon}$). With more noise in the quality measure, however, the risk aversion threshold is more binding under career concerns plus explicit incentives ($\frac{1}{\sigma^2_\nu - \sigma^2_\varepsilon} < \hat{r}$). Hence, when the quality measure is noisy ($\sigma_\nu > \sqrt{2}\sqrt{\sigma^2_\varepsilon}$), there are fewer risk aversion levels for which quantity effort increases monotonically with ability when there are career concerns and explicit incentives compared to explicit incentives alone.

Combining the characterization of the regions in which $e^*_{q,t} > 0, \forall m_0$, $e^*_{q,t} > 0$ for $\underline{m} < m_0 \leq \overline{m}$, and $e^*_{q,t} < 0, \forall m_0$ with the results for the sign of $\frac{\partial e^*_{p,t}}{\partial d}$ above, and using similar reasoning as in the model with career concerns and explicit performance incentives, three regions for heterogeneous quantity effort responses can be distinguished. First, quantity effort increases monotonically with ability class if risk aversion is low ($r \leq \hat{r}$) - so that $\frac{\partial e^*_{p,t}}{\partial d} > 0, \forall m_0$ - and (81) holds, so that $e^*_{q,t} > 0, \forall m_0$. Second, quantity effort is U-shaped in ability if risk aversion is high ($r > \hat{r}$) and either (81) or (82) hold as well, so that $e^*_{q,t} > 0, \forall m_0$ or $e^*_{q,t} > 0$ for $\underline{m} < m_0 \leq \overline{m}$. When $e^*_{q,t} > 0, \forall m_0$ and $r > \hat{r}$, the U-shape follows from (86): $\frac{\partial e^*_{p,t}}{\partial d} < 0$ for $m_0 > m\left(\hat{d}^{ex}\right)$, while $\frac{\partial e^*_{p,t}}{\partial d} > 0$ for $m_0 < m\left(\hat{d}^{ex}\right)$. When $e^*_{q,t} > 0$ for $\underline{m} < m_0 \leq \overline{m}$ and $r > \hat{r}$, the U-shape follows from the fact that $\frac{\partial e^*_{p,t}}{\partial d} < 0$ for large m_0, since $e^*_{q,t} > 0$ for high ability classes, and $\left[\frac{\partial e^*_{p,t}}{\partial d} | e^*_{q,t} > 0\right] < 0$ for $m_0 > m\left(\hat{d}^{ex}\right)$, while $\frac{\partial e^*_{p,t}}{\partial d} > 0$ for small m_0, either because $e^*_{q,t} = 0$ for low ability classes, and $\left[\frac{\partial e^*_{p,t}}{\partial d} | e^*_{q,t} = 0\right] > 0$, or because $e^*_{q,t} > 0$ and $\left[\frac{\partial e^*_{p,t}}{\partial d} | e^*_{q,t} > 0\right] > 0$ for $m_0 < m\left(\hat{d}^{ex}\right)$. Third, $e^*_{p,t}$ is monotonically decreasing if risk aversion is intermediate, such that (83) holds. This follows immediately from the fact that (83) ensures that $e^*_{q,t} = 0, \forall m_0$, and $\left[\frac{\partial e^*_{p,t}}{\partial d} | e^*_{q,t} = 0\right] > 0, \forall m_0$.

Comparing the above heterogeneous quantity response regions with those in the model with explicit and career concerns incentives, it follows that there are fewer risk aversion levels at which quantity effort is monotonically decreasing with ability class in response to explicit incentives only. This is a direct consequence of the earlier result that there are more (high) risk aversion values for which $e^*_{q,t} = 0, \forall m_0$ with explicit and career concerns incentives.

Turning to heterogeneous quality responses next, it is immediate from the assumptions on $\sigma^2_\nu, \sigma^2_\varepsilon$ and d that the second derivative in (85) is always negative. Thus, as in the model with career concerns plus explicit incentives or career concerns only, quality effort increases monotonically with ability when $e^*_{q,t}\left(\underline{m}\right) > 0$, the quality effort response is either U-shaped or J-shaped when $e^*_{q,t}\left(\underline{m}\right) = 0$ but $e^*_{q,t}\left(\overline{m}\right) > 0$, and the change in quality effort is negative for all classes and constant or decreasing in ability if $e^*_{q,t} = 0, \forall m_0$. Here too, four sub-cases can be distinguished, which follow directly from the earlier result that $e^*_{q,t} > \bar{e}_{q,t}$ if $0 < r < r^{ex}$, or else for $m_0 > m^{ex}$. These sub-cases are the direct equivalent to those arising in response to career concerns plus explicit incentives (cf. sub-section A1.3.2) and are therefore not repeated here.

Finally, when it comes to the selection effect, agents prefer a pay system with explicit incentives in t over a flat wage if the certainty equivalent utility in the former system (CE^{cc}_t) exceeds that in the latter (CE^{fw}_t) as of period t:

$$CE^{ex}_t\left(m_0\right) = \sum_{t=0}^{\infty}\left\{\left(m_t + e^*_{p,t}\right) + \left(m_t + e^*_{q,t}\right) - c\left(\overrightarrow{e^*_t}\right) - \frac{r}{2}\left(\overrightarrow{b^*_t}^2 \Sigma^2_{t,ex}\right)\right\} > \sum_{t=0}^{\infty}\left\{w_t\right\} = CE^{fw}_t \tag{88}$$

Recall that $\Sigma^2_{t,ex} = \begin{bmatrix} \sigma^2_\varepsilon & 0 \\ 0 & \sigma^2_\nu \end{bmatrix}$. The derivative of CE^{fw}_t with respect to ability class m_0 is 0, while the derivative of CE^{ex}_t is then:

$$\begin{aligned}\frac{\partial CE^{ex}_t\left(m_0\right)}{\partial m_0} = \sum_{t=0}^{\infty}\Bigg\{\left(\frac{2\sigma^2_\varepsilon\sigma^2_\nu}{\sigma^2_\varepsilon\sigma^2_\nu + t\sigma^2_0\left(\sigma^2_\varepsilon + \sigma^2_\nu\right)}\right) + \Bigg[\left(1 - \left(e^*_{p,t} - \bar{e}_{p,t}\right) - d\left(e^*_{q,t} - \bar{e}_{q,t}\right)\right)\frac{\partial e^*_{p,t}}{\partial d} \\ + \left(1 - \left(e^*_{q,t} - \bar{e}_{q,t}\right) - d\left(e^*_{p,t} - \bar{e}_{p,t}\right)\right)\frac{\partial e^*_{q,t}}{\partial d} - \left(e^*_{p,t} - \bar{e}_{p,t}\right)\left(e^*_{q,t} - \bar{e}_{q,t}\right) - r\left(\sigma^2_\varepsilon b^*_p\frac{\partial b^*_p}{\partial d} + \sigma^2_\nu b^*_q\frac{\partial b^*_q}{\partial d}\right)\Bigg]\frac{\partial d}{\partial m_0}\Bigg\}\end{aligned} \tag{89}$$

It can be shown that this derivative is unambiguously positive when $e^*_{q,t} > 0$, since the term in square brackets multiplying $\frac{\partial d}{\partial m_0}$ is negative. On the other hand, when $e^*_{q,t} = 0$, the derivative of CE^{ex}_t with respect to ability class m_0 is the same as that in the model with explicit and implicit performance incentives (53).

By a similar reasoning as used in the proof of Proposition 4, two regions can now be distinguished for the selection effect. When baseline quality effort is high $\left(\bar{e}_{q,t} \geq \bar{e}^{ex,l}_{q,t}\right)$ or risk aversion is low or high $\left(r < r^{q0,l}_{ex}\ OR\ r > r^{q1,l}_{ex}\right)$, so that $e^*_{q,t} > 0, \forall m_0$ (sufficient, not necessary, cf. (81)), CE^{ex}_t increases monotonically with ability class. The latter follows immediately from the sign of (89) when $e^*_{q,t} > 0, \forall m_0$. But then, if $\underline{m} < m^{s,pp} < \overline{m}$, there is positive selection.

If risk aversion is intermediate $\left(r^{q0,l}_{ex} < r < r^{q1,l}_{ex}\right)$ and baseline quality effort low $\left(\bar{e}_{q,t} < \bar{e}^{ex,l}_{q,t}\right)$, so that low ability classes exert no quality effort, but high ability classes might (cf. (82) and (83)), selection is ambiguous. To see this, note that $CE^{ex}_s(m_0)$ can be monotonically increasing, decreasing or U-shaped when $e^*_{q,t} = 0, \forall m_0$ by the same reasoning as provided under (53). When only low ability classes exert zero quality effort, $CE^{ex}_s(m_0)$ can be monotonically increasing or U-shaped. This follows, since (89) is positive for high ability classes, for whom $e^*_{q,t} > 0$. At the same time, (89) is either positive or negative for low ability classes, for whom $e^*_{q,t} = 0$, as shown under (53). Depending on if and where the CE^{fw}_s curve intersects with $CE^{ex}_s(m_0)$, selection can then be positive, negative or neither, and is thus ambiguous.

Comparing the selection effect of explicit incentives only to the selection effect of explicit and implicit incentives, it follows that selection is unambiguously positive for more risk aversion and minimum quality effort levels in the system with explicit incentives only. While selection into a system with career concerns and explicit performance incentives gives rise to unambiguously positive selection at low risk aversion levels only ($r \to 0$), selection into solely explicit performance incentives is positive when risk aversion is low or high $\left(r < r^{q0,l}_{ex}\ OR\ r > r^{q1,l}_{ex}\right)$, as well as at any risk aversion level when baseline quality effort is high $\left(\bar{e}_{q,t} \geq \bar{e}^{ex,l}_{q,t}\right)$.

The proposition below summarizes the differences in equilibrium behavior and comparative statics between the performance pay model with both explicit and career concerns incentives and the model with explicit performance incentives only.

Proposition 6 - Comparison with Explicit Incentives Only: *Comparing the performance pay equilibrium characterized in Proposition 1 to the explicit incentives only equilibrium derived above, the following differences can be seen:*

1. *There are fewer quality measure noise levels $\left(\sigma^2_\nu\right)$ at which quality effort is less than flat wage levels for all ability classes $\left(e^*_{q,t} < \bar{e}_{q,t}, \forall m_0\right)$ in response to explicit performance incentives only, and more risk aversion levels at which all ability classes increase quality effort. Furthermore, more ability classes increase quality effort in response to explicit incentives only when risk aversion is high (and $\sigma^2_\nu > 1$).*

2. *There are fewer (high) risk aversion levels at which quantity effort decreases monotonically with ability in response to explicit incentives only.*

3. *There are more risk aversion levels at which selection into explicit incentives only is unambiguously positive.*

Proof: See derivation in section above. ■

Corollary 4 - Relative Prediction - Explicit Incentives Only: *Relative to a performance pay system with career concerns and explicit performance incentives, a system with explicit performance incentives only is more likely to give rise to unambiguously positive selection as well as a positive quality effort effect, while quantity effort is less likely to decrease monotonically with ability.*

Proof: The corollary follows immediately from proposition 6, and the assumption that all parameter values that characterize the different regions for the effort effect, effort effect heterogeneity and selection effect in either pay system carry non-zero probability. ■

A2 Additional Institutional Details

A2.1 Performance Pay (W-Pay) and Tenure

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"). In order to qualify for a tenured affiliation - and thus performance pay bonuses in the performance pay scheme - academics need to have completed a PhD as well as, traditionally, a post-doctoral qualification ("habilitation"). The habilitation involves working as part of the research group of a full professor, and is completed with a postdoctoral thesis (Pritchard, 2006; Fitzenberger and Schulze, 2014).

In 2002 the German equivalent of assistant professorships ("Juniorprofessur") was introduced as an alternative path to tenured professorships (Pritchard, 2006). Junior professorships can last up to six years and grant aspiring academics more independence than the habilitation (Fitzenberger and Schulze, 2014). In either case, aspiring professors would need to apply for tenured professorships after the completion of the habilitation or Junior professorship, because during the time period covered by the data for the empirical analysis, tenure-track positions generally did not exist. Furthermore, universities were commonly not allowed to hire their own habilitands or Junior professors as tenured professors due to the so-called "home-hiring ban" ("Hausberufungsverbot") (Academics.de, 2016). Hence, aspiring professors would normally have to move to another university for a first tenured position, and they would start on a new employment contract.

Attraction bonuses are, in principle, available for first-time tenured professors if their qualifications and expected academic success warrants such bonuses. Some states, however, stipulate that first-time tenured professors are to be offered the basic wage only, except in exceptional circumstances (Detmer and Preissler, 2006). In response, many professors who just acquired their first tenured position would immediately apply for other tenured positions so as to be able to negotiate an attraction or retention bonus (Detmer and Preissler, 2006). Professors are also generally required to show proof of another offer in order to be able to negotiate a retention bonus (Detmer and Preissler, 2005). The attraction and retention bonuses are thus implicit, market-based incentives. The president or rector of a university or the dean of the relevant faculty usually decides on the attraction or retention bonus (cf. e.g. Bergsdorf (2005); Huber (2005); Leitungsgremium (2005); Universitaet Regensburg (2016)).

Junior Professors, can earn a (non-pensionable) supplement of 260 euro per month upon positive evaluation in the W-pay scheme, plus an additional supplement - not to exceed 10% of a junior professor's basic wage - in special circumstances (Detmer and Preissler, 2005). For comparison, tenured professors can earn performance bonuses of up to 5241,48 euro per month - more than the highest basic wage in the performance pay scheme - or more in special circumstances[48] (BMBF, 2002; Detmer and Preissler, 2005). Performance incentives are thus much more high-powered for tenured professors, which is why I restrict attention to those in the empirical analyses.

A2.2 On-the-job Performance Bonuses

Most universities distinguish several performance levels that are associated with increasing on-the-job performance bonuses (the so-called "Leistungsstufen") (Harbring, Irlenbusch and Kräkel, 2004; Kräkel, 2006; Lünstroth, 2011). The first performance level of the Regensburg University for instance requires performance to exceed fulfillment of (normal) professorial duties, the second level requires achievements that help further the national standing of the university, while the third level is reserved for achievements that have improved the international reputation of the university (Universitaet Regensburg, 2016). The lowest performance level generally requires achievements that lie (substantially) above those in line with the ordinary fulfillment of professorial duties (cf. Huber (2005); Universitaet Regensburg (2016); Gien (2017)). The on-the-job performance bonuses thus constitute explicit performance incentives. These incentives should affect only academics that fall under the performance pay scheme, and only from the moment that the academics enter into the performance pay scheme.

There is substantial variation in the number of performance 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). Generally speaking, the university president, rector or council announces either both the number of levels and associated bonus pay or only the total number of bonuses (if the bonus pay amounts are specified in the university's statutes) to be awarded in a given year at the beginning of that year (Lünstroth, 2011). It is therefore difficult for academics to know, ex ante, at what university they may have a higher chance of earning on-the-job bonuses.

A2.3 Cost Neutrality

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[49] (benchmark year 2001) (BMBF, 2002). 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 going forward (Handel, 2005). The law does allow for this benchmark to be exceeded by, on average, 2% per year, though not exceeding 10% in total, as long as the state's budget allows (BMBF, 2002).

The budget-neutrality stipulation was explicitly introduced to prevent the reform from leading to cost cutting or a cost explosion (Handel, 2005; Detmer and Preissler, 2006). Because the basic wage in the

[48] For instance, when the academic already earns bonuses that exceed this limit and a higher bonus is necessary to attract the academic or prevent them from leaving to another univerity in Germany or abroad.

[49] States are allowed to raise their target average professorial pay level to - at most - the highest average professorial pay at state or federal level.

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[50].

A3 Further Details on Data

In this appendix, I provide further information about the three core input data sets, and describe the preparation and matching procedures used to generate the panel data for the empirical tests in this paper. All data handling was done using Python, unless indicated otherwise.

A3.1 Further Details on DGK

Kuerschners Deutscher Gelehrten Kalender (DGK) is a bibliographic and bibliometric encyclopedia of academics affiliated with German, Austrian and Swiss universities. All people who have passed the "venia legendi" and are both actively teaching and researching at a relevant university in Germany, Austria and Switzerland are included in DGK. The "venia legendi" encompasses the "habilitation" and a qualification to teach at university level (the "Lehrbefugnis"). The habilitation is a post-doctoral qualification that is acquired through publication of a habilitation thesis after up to six years of research as part of a full professor's research group ("Lehrstuhl"). An exception to the venia legendi rule for inclusion in DGK are honorary professorships (Honorarprofessoren) and junior professorships (Juniorprofessoren). Universities considered relevant for DGK are generally those that can reward doctoral degrees ("Promotionsrecht"). This includes all public universities that I restrict attention to. Academics who move to a university outside of Germany, Austria or Switzerland are generally dropped from the encyclopedia, unless they personally request to remain included. People whose academic affiliation can no longer be verified are dropped from the encyclopedia too.

The information in DGK stems from academic calendars, course rosters/teaching schedules, announcements of appointments by universities and in academic and professional journals, surveys, university websites, etc. (De Gruyter, 2006, 2008). De Gruyter Publishers, the current publishers of the DGK, have kindly supplied me with the editorial database underlying the online DGK edition (current up to 13 July 2013), as well as a copy of the exports from this database taken on 10-11-2006, 17-11-2008 and 27-09-2010. This database and its past exports contain all the information of published DGK editions from the same years: all records of people complying with the DGK inclusion criteria set out above, inactive records of people who left (German, Austrian or Swiss) academia, as well as records of people who passed away or could no longer be traced. The database also contains activation dates of records (the date when a person first complied with the DGK inclusion criteria and was included in the database) and inactivation dates, where applicable[51].

[50] 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.

[51] The DGK editorial database was started in 1996, when the DGK data were migrated from the previous publisher to De Gruyter. The earliest activation dates in the database however appear to be 1999, and the affiliation histories used in this paper therefore start as of that year. The De Gruyter database is updated continuously and is used to generate the online version of the DGK. The DGK has an online version since 2010.

From each DGK edition, I extract the affiliation information and other career and personal information provided by the editors of DGK, as well as supplementary self-reported career information provided by academics and listed in one of the DGK data fields. To extract usable data from the self-reported career information field, I exploit regularities in the structure of the text in this field with an extensive set of regular expressions. I discard information snippets that have no valid (start) date, university name or title. I also discard information about temporary positions, such as visiting positions. Section A3.6 in this online appendix explains how the DGK information is used to construct an individual-level affiliations panel.

A3.2 Further Details on FuL

Forschung und Lehre (FuL) is Germany's largest higher education and research magazine that has been published monthly by the German higher education association (Deutscher Hochschulverband) since 1994 (DHV, 2014). Every magazine contains a section titled "Habilitationen und Berufungen" with announcements of habilitations, acquisitions of the Lehrbefugnis (an authorization to lecture), professorial appointments and the receipt, acceptance or rejection of academic (professorial) offers. These notifications are based on information from press releases from universities, newspapers and professional magazines as well as from readers/individual scientists (DHV, 2002). Electronic copies of past Forschung und Lehre magazines from 1996 onwards can be downloaded from the "archive" section of the magazine's website (DHV, 1999-2013). I use Forschung und Lehre magazines from 1999 to 2013, to align with the years for which I have data from DGK.

I first extract the text from the pdfs (electronic copies) of the Forschung und Lehre magazines and subsequently use an extensive set of regular expressions to extract relevant pieces of information from the "Habilitationen und Berufungen" announcements. For Habilitation or Lehrbefugnis announcements, I extract the name and current title of the academic, the current affiliation, the university at which the qualification was obtained (if different from the current affiliation), the field in which the qualification was acquired, as well as the subject category under which the announcement was made in the FuL magazine. I take the month and year of the FuL issue in which the announcement was made to be the time when the qualification was obtained, backdated by four months to correct for the average printing lag.

In the case of a professorial offer ("Berufung") announcement, I extract whether the offer was obtained ("erhalten"), accepted ("angenommen"), or rejected ("abgelehnt), or if the professor was appointed ("ernannt"). I also record the name and current title of the academic, the current affiliation, offer university, offered position and field in which the position is offered, as well as the subject category under which the announcement was made in the FuL magazine. In case of multiple offers, I always record accepted or appointed offers first, followed by offers that are obtained. I record rejected offers last. In case of only obtained offers, I record offers from German universities first, otherwise the order is random. I take the month and year of the FuL issue in which the announcement was made to be the time when the offer was accepted, rejected or appointed, backdated by four months to correct for the average printing lag.[52] The announcement of an appointment in e.g. February 2003 is thus interpreted as the appointment tak-

[52] Offers that were only reported as being extended/obtained are not backdated, because there are usually 2-3 months between offer extension ("Ruf") and acceptance or rejection, and 6 to 7 months between Ruf and appointment (Wissenschaftsrat, 2005). With an average printing lag of 4 months, the FuL announcement of a Ruf therefore likely falls between its acceptance or rejection and possible appointment.

ing place in October 2002. Section A3.6 in this online appendix explains how this information is used to record contract renegotiations and help construct an individual-level affiliations panel.

A3.3 Further Details on ISI

The ISI Web of Science database (hereafter: ISI) is compiled and maintained by Clarivate Analytics (and, before that, by Thomson Reuters) and can be accessed via the website webofknowledge.com (Clarivate Analytics, 1993-2012*a*). From this database, I restrict attention to publications from the following databases: Science Citation Index Expanded (SCI-Expanded), Social Sciences Citation Index (SSCI), Arts and Humanities Citation Index (AHCI), Conference Proceedings Citation Index - Science (CPCI-S) and the Conference Proceedings Citation Index - Social Sciences & Humanities (CPCI-SSH). I downloaded all records of publications with at least one author with a German (work) address and published between 1993 and 2012 from the ISI website.

A3.4 Matching DGK and FuL

I make the information from FuL and DGK compatible by replacing university names in the FuL and DGK databases with unique identifiers, mapping titles and positions to a unified list of existing titles and positions, and classifying a title or position as being tenured or non-tenured. The following are tenured professorial positions: C3-Professor, W2-Professor, Ausserordentliche Professor and Associate Professor as well as C4-Professor, W3-Professor, Ordentliche Professor and U(niversitaets)-Prof. Furthermore, I classify all subject areas distinguished in DGK under 12 broad field categories. These are the fields distinguished in the 'Habilitationen und Berufungen section' of FuL: 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. For example, the DGK subject areas 'Rechtswissenschaft' and 'Rechtsgeschichte' are mapped into FuL-field 'law', while 'Immunologie' and 'Molekulare Medizin" are mapped into the FuL-field 'medicine'[53].

Subsequently, I distill a list of unique academics from both the FuL and DGK records. In order to do so, I match academics appearing in FuL with academics in DGK on the basis of their last name, subject area and initials, and subsequently deduplicate the list of academics on the basis of these same criteria. As last name, I use the name after the last space in the FuL name field, with potential hyphens of composite last names deleted (so e.g. Schmidt-Angel becomes SchmidtAngel). Composite first names are separated first and the first letter of all name components are taken to be initials (e.g. Anna-Maria has initials A, M). I match on initials rather than first names, because the publication records in ISI list initials only. Matching on initials, last name and subject area is therefore the best I can do across all three main input data sets. I require at least one of the FuL-field codes for the field or subject areas listed for an academic in DGK to be the same as the FuL-field code an academic is classified under in FuL. If an academic does not have a subject area listed in DGK or if this subject area cannot be classified under one of the FuL-field codes, a match is attempted on the basis of last name and initials only.

[53]The full mapping is available from the author upon request.

To improve matching accuracy, I discard a potential match if:

a) a person's last (most recent) announcement in FuL is made while a potential match in DGK is over 67 years old[54] (based on birth year given in DGK)

b) a potential DGK match has a date of passing that falls before the last (most recent) announcement year in FuL

c) a potential DGK match is reported to have retired in DGK-year-x, while there are FuL announcements after year x

d) a potential DGK match is reported as having a tenured position before the habilitation year reported in FuL[55]

As mentioned in the main text, 83% of academics who appeared as having a tenured affiliation at a German university in FuL can be matched to academics listed in DGK. Failed matches are mostly due to spelling inconsistencies or errors, which I resolve manually where possible.

A3.5 Matching Academics with Publications

I match publications from ISI to the deduplicated list of academics appearing in FuL and DGK on the basis of last name, initials and subject area. To enable matching on subject area, I map the Web of Science categories listed for journals to the aforementioned 12 FuL-field codes. The Web of Science categorizes 'Ethics' and 'History" are mapped into the FuL-field 'philosophy and history' for instance, while the categories 'Economics' and 'Industrial Relations & Labor' are mapped to the FuL-field 'economics'[56]. Furthermore, to deal with differences in the way last names are represented across sources, I abstract from common prefixes such as 'von' and 'von der'.

I match publications to academics using subsequent sets of criteria. I first try to match publications to academics on the basis of the last name, all initials and field or subject area. If multiple academics can be matched to a publication on the same criteria, the publication is attributed to all matched academics. If there are no matches on last name, initials and field, I attempt a match on a slightly different set of criteria: last name with any spaces and hyphens removed, initials and field or subject. If no matches are found still, I move on to the next set of criteria, and so on. The subsequent sets of criteria are: last name, first initial only and field or subject; last name without spaces and hyphens, first initial only and field or subject; first last name (if composite last name such as "Gross Herzenberg", for male academics only), all initials and field or subject; first last name (if composite last name such as "Gross Herzenberg", for male academics only), first initial only, and field or subject; second last name (if composite last name such as "Schmidt-Bauer", for female academics only), all initials and field or subject; second last name (if composite last name such

[54] German law mandates that academics retire at the age of 65 (Mohr, 2007; Bundesgesetzblatt, 1985), so unless an academic moves abroad around the time of mandated retirement in Germany (cf. Mohr, 2007), I should not observe FuL announcements regarding new affiliations for an academic who is past the age of retirement. I use 67 as cut-off instead of 65 to allow for some delay in a possible move abroad or FuL's reporting thereof.

[55] Where I allow for the habilitation announcement to occur in the year after tenure, to accommodate cases in which an academic obtains a tenured position immediately upon passing the habilitation while the habilitation announcement is delayed.

[56] The full mapping is available from the author upon request.

as "Schmidt-Bauer", for female academics only), first initial only and field or subject. If still no match has been found, I attempt to match the publication to academics who do not have a field or subject code[57], on the basis of (in order): last name and all initials; last name without spaces and hyphens and all initials; last name and first initial only; last name without spaces and first initial only.

A3.6 Creating an Individual-Level Affiliations Panel

For any professorial offer announcement in FuL, starting from the first (oldest) FuL volume going forward, the current university of a person, their current position (title) and whether this concerns a tenured position is filled back in time from the year before the FuL announcement to whichever one of the following is later: 1) the year that FuL reported as the year in which the person passed their Habilitation or Lehrbefugnis, 2) the year after the start of the current position, or 3) the start year of the panel[58]. If the FuL announcement concerns an accepted offer, the new university, new position (title) and whether the position is tenured or not is filled forward from the year of the FuL announcement to the last year of the panel[59]. If the FuL announcement concerns an appointment ("ernannt"), an academic's current university is taken to be the offer university (i.e. an appointment to a different position within the same university) unless the offer university is specifically stated to be different from an academic's current university. The offer university, offered position (title) and whether the position is tenured is filled forward in the same way as with an announcement of an accepted offer. If the FuL announcement states that an offer was rejected, the current university, current position and whether the position is tenured or not is filled forward as above. Finally, for an announcement of a received offer ("erhalten"), the information regarding the offer university, position and whether the position is tenured is filled forward tentatively, with the option to overwrite if new information arrives that supersedes it. This process is repeated and the affilliations updated in chronological order until the last FuL volume is processed. I also record announcements of rejections and offer extensions as contract renegotiations.

Next, I cross-check and combine the affiliation information in DGK with that from FuL as follows, going from oldest to newest DGK edition. Starting from the last (most recent) piece of self-reported career information in a DGK edition going backward, I fill out the affiliation information contained in this piece from the start year provided to one of the following: 1) the listed end year of the affiliation (if provided), 2) the start date of a new (more recent) position, or 3) the publication date of the current DGK edition. I do so only if one of the following conditions is met: 1) the information is the last entry in the self-reported career information field and the information matches the affiliation information provided by DGK editors, 2) there is already affiliation information for a particular year and that information is derived from a self-reported piece in the current DGK edition, or it concerns affiliation information provided by DGK editors in a previous DGK edition, or 3) there is already affiliation information for a particular year and that information stems from an FuL announcement made before the reported end or start date of the

[57] This is the case if the subject area recorded in DGK could not be mapped to an FuL field code or if no subject area was recorded in DGK

[58] If an affiliation is already filled out in a year before the offer announcement, the current position is checked for consistency with the affiliation already recorded in the panel. If the recorded affiliation is incomplete (e.g. contains only a title or university) but matches the new information, the partial affiliation information is supplemented with the new information. If the two do not match up, the current FuL information is filled out for the year immediately prior to the FuL announcement only, an error message is created and the case is left for further, case-by-case evaluation.

[59] Note that this information is overwritten when new information, from a later announcement, arrives. There is thus chronological updating.

self-reported affiliation[60].

Subsequently, I fill out the editor-provided affiliation information, starting from the publication year of the DGK edition and going backward. I do not overwrite affiliation information that stems from an FuL announcement in the same year as the DGK edition or the year thereafter, but flag those for further inspection. I also do not overwrite affiliation information derived from an FuL announcement in the same year when the DGK affiliation matches the previous affiliation listed in the FuL announcement (this suggests DGK is not up-to-date for the academic). I use the editor-provided information to correct any mislabeling in FuL of honorary or temporary professorial positions as tenured positions.

In a last step, I make sure that any earliest affiliations are filled backward until an academic received their habilitation or Lehrbefugnis, or otherwise became active in German academia and I delete any affiliation information before these starting points. Similarly, I fill any latest affiliations forward until the affiliation's end year (if listed in DGK), year of passing of the academic, or year in which the academic otherwise left German academia, and I delete any affiliation information filled out beyond these years. Finally, I drop any affiliation information for years that fall outside of the period for which I have affiliation information from DGK and FuL (1999-2013).

The affiliations, publications and renegotiations data are stored in a database, from which the data sets used for the various analyses in the paper are extracted.

A4 Further Empirical Analyses

A4.1 Synthetic Cohorts

As a further 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 performance incentives introduced with the pay reform, I estimate the baseline regression with synthetic cohorts, which are defined by 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, which is equal to an academic's birth year if known, and a synthetic birth year otherwise. In turn, the synthetic birth year is derived 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[61].

Panel D in Appendix Table A.1 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 citations of 12.4% (all significant at 5%) as of the moment the implicit, career concerns incentives of the performance pay scheme take effect.

[60] FuL information announced after the end date of self-reported information will get overwritten by DGK information in later DGK editions or, if announced after the end date of the affiliation information in the lastest DGK edition, flagged for further inspection.

[61] These average ages are 37 and 30, respectively.

A4.2 Instrumental Variables Approach

An alternative way to estimate the effort effect is to focus on switchers instead: estimating the baseline specification for academics who hold a tenured affiliation before the reform and labeling those that switch to performance pay as the treated group. The assignment to this treatment group is however endogenous, and I therefore use an instrumental variables approach. Figure 6a shows that older academics earn a higher basic wage under the age-related pay system, specifically when they are 33 or 43 years of age or older. Age and age-related variables may therefore have explanatory power for selection into performance pay and could thus potentially serve as instruments.

I estimate an instrumented version of the baseline specification, with $Post'02$ and $Contract\_Change$ interacted with a $Treatment$ dummy as the interaction variables of interest. The $Treatment$ dummy is one for academics who held a tenured affiliation at a public university before implementation of the pay reform, and who change affiliation, position or contract after implementation. $Contract\_Change$ is one in the year in which a position, affiliation or contract change happens, as well as all following years. If this contract change happens after 2005, it coincides with the moment an academic enters into the performance pay scheme. As such, it is the equivalent to the $Tenured$ variable in the baseline specification, and coincides with the moment the explicit incentives of the performance pay scheme take effect. Both the $Treatment$ dummy and $Contract\_Change$ variable are possibly endogenous.

I use two sets of age-related variables as instruments: age and age-squared (Panel A in Table A.7), and indicator variables that are equal to one if an academic is 33 and 43 years of age or older (Panel B in Table A.7). Because I do not have age information for all academics, I construct a synthetic age variable for which I impute unknown ages using the average age at PhD, habilitation or tenure. I follow Aghion, Howitt and Mayer-Foulkes (2005) in instrumenting for the endogenous terms with interacted instruments. That is, I use $Synthetic\,Age$ and $Synthetic\,Age\,Squared$, interacted with the $Post'02$ and $Post'05$ variables, as instruments for $Post'02 * Treatment$ and $Contract\_Change * Treatment$. I estimate the resulting instrumented panel fixed effects model using the two-stage efficient GMM estimator with robust standard errors clustered at the individual level.[62,63]

The resulting estimates of the effort effect are qualitatively similar to the baseline results presented in the main text; with increases in total (quality-adjusted) research output and impact, and - imprecisely estimated - decreases in average impact.[64] The magnitude of the effect estimates is much larger than in the equivalent panel fixed effects model (Panel C in Table A.7). The latter is likely at least in part because the instruments appear not to be valid in the age and age-squared instrumented total quantity and total impact regressions: the Hansen J statistic suggests the instruments are not uncorrelated with the error term in columns 1 through 3 of Panel A. Indeed, it is entirely plausible that age affects productivity not just through its effect on the likelihood of switching to performance pay, but directly as well. Furthermore, the instrumented

[62] I use the 2-step GMM estimator so as to derive efficient estimates in the presence of arbitrary heteroskedasticity and clustering (Wooldridge, 2010). I estimate the model as a linear IV (panel fixed effects) model to be able to perform a number of IV diagnostic tests, even though the first stage would be most appropriately estimated as a hazard rate model and the second stage as a Poisson model.

[63] To align the sample with that used for the effort effect estimation in the main text, I restrict attention to academics who started their first tenured affiliation in 2002, 2003 or 2004, and who did not hold a foreign affiliation immediately preceding this. Resuls are however similar when extending this inclusion window to 1999-2004 (results available from the author on request).

[64] Results are similar when substituting $Post'05 * Treatment$ for $Contract\_Change * Treatment$ in the second stage. Results are also similar when estimating as 2SLS instead of GMM (Results are available from the author upon request).

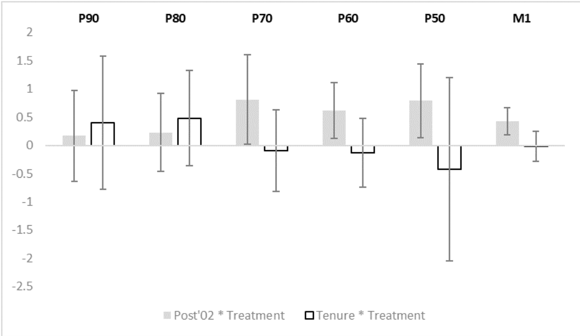
interaction terms appear not to be endogenous in the average quality or impact regressions throughout and in the over-32 and over-42 instrumented quality adjusted and total impact regressions. Because of these misspecification concerns, the instrumental variables estimation results should be interpreted with extreme caution, and taken to be indicative, at best.

A4.3 Tenure Probability

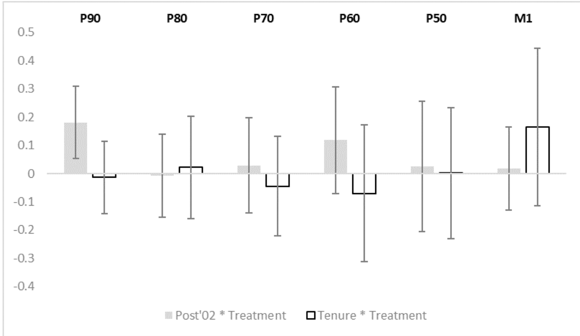
The effort effect estimates may be biased if the probability of obtaining a tenured affiliation changes with the reform. In particular, I have to rule out that tenure requirements go up, as this would lead to more productive academics obtaining tenured positions after implementation of the reform. To this end, I estimate the tenure probability using hazard rate analysis, much like the analysis of switching rates in the selection effect section in the main text of the paper (Section 4.3.3). Here the event of interest is the start of the first tenured affiliation[65], and academics are "at risk" of obtaining a tenured affiliation after completion of the habilitation. For academics for whom the habilitation year is unknown, I impute it using the average age at completion of the habilitation. I estimate Weibull proportional hazard models of the tenure probability as a function of synthetic age (defined as above) and productivity and the interactions of those variables with the $Post'05$ trend-break variable, controlling for academic field. As productivity variables I use the two-year lag of the number of publications (Table 3 columns 1a and 1b) and the impact factor-rated number of publications (columns 2a and 2b).

As expected, the probability of obtaining tenure increases over time on average, as evidenced by $\rho > 1$ and the positive significant $Post'05$ coefficient estimate (the latter picks up on any average change in tenure probability over time, as it is the only time-related variable included in the specification). This general increase in tenure probability is, however, tempered by a negative effect of age on tenure probability. As expected, a higher productivity increases the probability of obtaining a first tenured affiliation in general, but there is no additional increase after implementation of the reform (compare the coefficient estimates of the $Productivity$ and $Post'05 * Productivity$ variables). That is, there is no evidence that the requirements for becoming a tenured professor increase with the reform, and hence no evidence that academics who start their first tenured position after the reform are more productive. Results are robust to weighting the productivity variables by number of authors, using productivity averages over the pre-reform years 2002-2004, or lagging productivity variables by one year (results available from the author upon request).

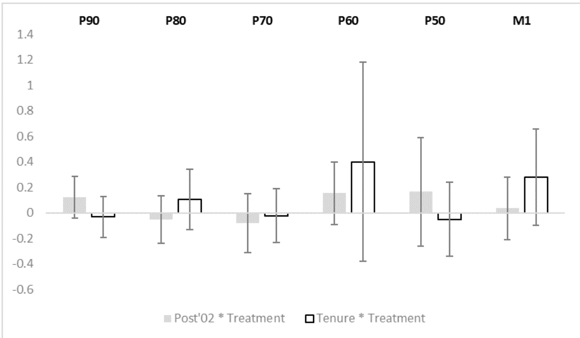
[65] To align with the preceding analyses, I restrict attention here to first tenured affiliations when the preceding affiliation is not foreign, though results are robust to including all first observations of tenured positions (results available from the author upon request).



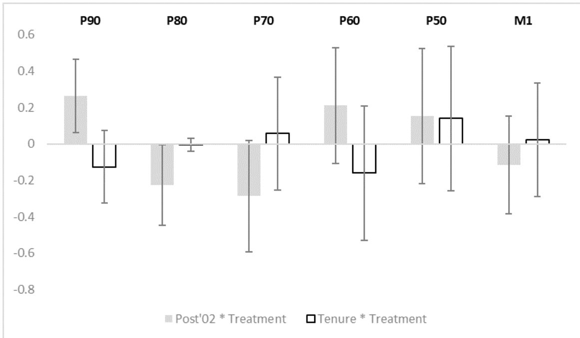
(a) Extensive Margin Response



(b) Intensive Margin Response - Number of Publications



(c) Intensive Margin Response - Total Impact Factor Rating



(d) Intensive Margin Response - Total Number of Citations

Figure A.1: Extensive and Intensive Margin Responses

Notes: The histograms depict the coefficient estimates and corresponding 95% confidence intervals of the $post'02{*}Treatment$ (grey bars) and $Tenure{*}Treatment$ (white bars) interactions from separate logit regressions of the probability that an academic has at least one publication in a given year in subfigure a, and from separate regressions with the following conditional dependent variables (conditional on having at least one publication) in sub-figures b-d: number of publications (b), impact factor-rated number of publications (c), total number of citations to all publications published in a given year (d). Samples are restricted to, respectively, the top five productivity deciles (P90-P50) and below median productivity academics (M1). Productivity deciles are determined on the basis of the averages of the impact factor-rated number of publications over the three pre-announcement years 1999, 2000 and 2001, separately by academic field and treatment group. All other specifications as in the baseline regression of the effort effect. See section 4.2.7 for further details.

Table A.1: Robustness Checks

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: Publication Variables Restricted to Articles Only					
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 C: Inverse Hyperbolic Sine Transform specification					
Post'02 * Treatment	0.082***	0.074***	0.138***	-0.045**	-0.040
	(0.020)	(0.028)	(0.049)	(0.022)	(0.049)
Tenure * Treatment	-0.021	-0.008	0.013	0.006	0.041
	(0.024)	(0.031)	(0.054)	(0.023)	(0.055)
Number of Observations	108363	108363	108363	48552	48552
Number of Individuals	6039	6039	6039	4671	4671
Log Likelihood	-95546.352	-129072.048	-192127.937	-32323.396	-73172.001
Chi-squared	2927.707	2987.478	2115.081	1454.155	540.056
Panel D: 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

Table A.1: Robustness Checks

Panel E: With Post'05*Treatment Interaction					
	# Publications	IF-rated publications	Citations	Average IF-rating	Average citations
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: 4-Year Treatment and Control Groups					
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 G: 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
Panel H: Controlling for 7 Years Before to 8 Years after Tenure, Less Pre-Tenure Year					
Post'02 * Treatment	0.147***	0.115**	0.107	-0.108***	-0.115*
	(0.043)	(0.059)	(0.070)	(0.038)	(0.070)
Tenure * Treatment	-0.016	0.026	0.017	0.033	-0.013
	(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	-136646.313	-338239.390	-4508059.852	-70680.751	-736200.571
Chi-squared	2865.820	2854.092	1098.897	607.815	475.542

Notes: The table reports the coefficient estimates and standard errors of the $post'02 * Treatment$ and $Tenure * Treatment$ interactions from separate regressions with the following dependent variables: number of publications, impact factor-rated number of publications, total number of citations, average impact factor rating, and average number of citations. Panel A reports the estimation results of the baseline effort effect specification as a panel fixed effects model. In Panel B, the dependent variables are based on a restricted set of publications, comprising only journal articles and proceedings papers. Panel C shows the results of the estimation of the baseline specification as a fixed effects panel data model with the inverse hyperbolic sine transformation of the dependent variables as dependent variables. In panel D, the time-to-tenure fixed effects in the baseline specification are replaced by an absolute time-to-tenure variable, which is 0 in the first year of the first tenured position, 1 both in the year before and after, and so on. In Panel E the $Tenure * Treatment$ interaction is substituted for a $Post'05 * Treatment$ interaction to capture the effect of the explicit performance incentives, where $Post'05$ is 1 as of 2005. In Panel F, the treatment group is comprised of academics who start their first tenured affiliation in 2004-2008 and the control group of academics who start their first tenured affiliation in 2001-2004. In Panel G, the control group starts their first tenured affiliation in 2003 or 2004, and the treatment group in 2005 and 2006. In Panel H the time-to-tenure fixed effect of the tenure year is dropped, while a time-to-tenure fixed effect for the eight year after tenure is added. All other specifications as in the baseline regression of the effort effect. See sections 4.2.3-4.2.5 for further details.

Table A.2: Additional Baseline Results

	Panel A: Maximum and Minimum Number of Citations		Panel B: Average Number of Co-authors and Pages	
	Maximum citations	Minimum citations	Co-authors	Pages
Post'02 * Treatment	0.001	-0.176	0.054	-0.051
	(0.066)	(0.136)	(0.155)	(0.042)
Tenure * Treatment	-0.063	-0.058	-0.213	0.029
	(0.078)	(0.171)	(0.183)	(0.036)
Number of Observations	47789	45928	47092	48097
Number of Individuals	4110	3967	3933	4216
Log Likelihood	-2366039.795	-525485.210	-268258.275	-131295.152
Chi-squared	430.387	457.933	426.031	176.011

Notes: The table reports the coefficient estimates and standard errors of the $post'02 * Treatment$ and $Tenure * Treatment$ interactions from separate regressions with the following dependent variables: the maximum and minimum number of citations to the publications of an academic in a given year (panel A), and the average number of co-authors on publications and the average number of pages of publications, where the average is taken over all publications of an academic in a given year (panel B). All other specifications as in the baseline regression of the effort effect. See sections 4.2.3 and 4.2.5 for further details.

Table A.3: Heterogeneous Response by Broad Academic Field

	# Publications	IF-rated publications	Citations	Average IF-rating	Average citations
Post'02 * Treatment	0.260**	-0.127	0.027	-0.272***	-0.185
	(0.102)	(0.168)	(0.196)	(0.101)	(0.167)
Post'02 * Treatment * Nat. Apl Sci.	-0.131	0.238	0.080	0.188*	0.085
	(0.114)	(0.184)	(0.211)	(0.111)	(0.181)
Tenure * Treatment	-0.019	0.168	-0.080	0.187*	-0.119
	(0.112)	(0.165)	(0.185)	(0.102)	(0.143)
Tenure * Treatment * Nat. Apl Sci.	0.016	-0.133	0.102	-0.172*	0.090
	(0.098)	(0.148)	(0.172)	(0.095)	(0.135)
Number of Observations	80953	71540	75504	45588	46324
Number of Individuals	4505	3981	4201	3771	3966
Log Likelihood	-132161.984	-326341.640	-4342991.519	-68490.957	-710222.214
Chi-squared	2910.339	2968.235	1113.439	669.440	480.945

Notes: The table reports results from the estimation of the baseline effort effect model augmented with a dummy variable for *natural and applied sciences* (Nat. Apl Sci.), and its double and triple interactions with $Post'02$, $Tenure$ and $Treatment$. The omitted category is *social sciences and humanities*. All other specifications as in the baseline regression of the effort effect. See section 4.2.6 for further details.

Table A.4: Heterogeneous Responses - Interactions

Panel A: By Productivity Quantile	# Publications	IF-rated publications	Citations	Average IF-rating	Average citations
Post'02 * Treatment	0.280***	0.296**	0.249*	-0.116	-0.107
	(0.071)	(0.116)	(0.129)	(0.077)	(0.110)
Post'02 * Treatment * Top Decile	-0.152*	-0.199	-0.063	0.038	0.161
	(0.087)	(0.132)	(0.152)	(0.091)	(0.140)
Post'02 * Treatment * 9th Decile	-0.331***	-0.443***	-0.391**	-0.016	0.014
	(0.095)	(0.140)	(0.163)	(0.096)	(0.147)
Post'02 * Treatment * 8th Decile	-0.152	-0.290*	-0.497***	-0.008	-0.270
	(0.101)	(0.152)	(0.182)	(0.094)	(0.181)
Post'02 * Treatment * 7th Decile	-0.022	-0.019	-0.053	0.004	-0.222
	(0.107)	(0.155)	(0.176)	(0.125)	(0.171)
Post'02 * Treatment * 6th Decile	-0.094	0.134	0.252	0.263**	0.470***
	(0.124)	(0.217)	(0.223)	(0.110)	(0.177)
Tenure * Treatment	-0.032	0.093	0.046	0.090	-0.042
	(0.079)	(0.115)	(0.134)	(0.071)	(0.109)
Tenure * Treatment * Top Decile	-0.035	-0.150	-0.199	-0.098	-0.143
	(0.080)	(0.114)	(0.139)	(0.083)	(0.131)
Tenure * Treatment * 9th Decile	0.118	0.030	0.152	-0.124	0.064
	(0.096)	(0.134)	(0.172)	(0.083)	(0.136)
Tenure * Treatment * 8th Decile	0.071	0.020	0.170	-0.059	0.102
	(0.092)	(0.130)	(0.165)	(0.082)	(0.152)
Tenure * Treatment * 7th Decile	0.021	-0.107	-0.086	0.014	0.226
	(0.101)	(0.139)	(0.191)	(0.114)	(0.166)
Tenure * Treatment * 6th Decile	0.110	-0.049	-0.004	-0.217**	-0.194
	(0.111)	(0.172)	(0.199)	(0.096)	(0.165)
Number of Observations	80953	71540	75504	45588	46324
Number of Individuals	4505	3981	4201	3771	3966
Log Likelihood	-131129.798	-321279.569	-4287805.762	-68430.475	-708085.051
Chi-squared	3329.827	3418.293	1569.162	687.096	509.727

Panel B: Without Switchers	# Publications	IF-rated publications	Citations	Average IF-rating	Average citations
Post'02 * Treatment	0.302***	0.331***	0.309**	-0.086	-0.014
	(0.078)	(0.128)	(0.141)	(0.088)	(0.120)
Post'02 * Treatment * Top Decile	-0.167*	-0.196	-0.109	0.012	0.021
	(0.094)	(0.145)	(0.166)	(0.111)	(0.151)
Post'02 * Treatment * 9th Decile	-0.327***	-0.461***	-0.387**	-0.049	-0.099
	(0.105)	(0.156)	(0.187)	(0.108)	(0.169)
Post'02 * Treatment * 8th Decile	-0.113	-0.271*	-0.556***	-0.051	-0.407**
	(0.109)	(0.162)	(0.198)	(0.106)	(0.201)
Post'02 * Treatment * 7th Decile	0.036	0.028	-0.068	-0.075	-0.323*
	(0.115)	(0.167)	(0.195)	(0.146)	(0.195)
Post'02 * Treatment * 6th Decile	0.040	0.274	0.301	0.272**	0.331*
	(0.142)	(0.282)	(0.278)	(0.124)	(0.188)
Tenure * Treatment	-0.025	0.059	0.021	0.064	-0.087
	(0.084)	(0.123)	(0.148)	(0.083)	(0.121)
Tenure * Treatment * Top Decile	-0.006	-0.105	-0.171	-0.087	-0.070
	(0.086)	(0.123)	(0.155)	(0.103)	(0.148)
Tenure * Treatment * 9th Decile	0.131	0.063	0.147	-0.077	0.134
	(0.108)	(0.152)	(0.198)	(0.096)	(0.156)
Tenure * Treatment * 8th Decile	0.078	0.058	0.216	-0.037	0.166
	(0.098)	(0.136)	(0.181)	(0.093)	(0.170)
Tenure * Treatment * 7th Decile	-0.012	-0.160	-0.139	0.062	0.239
	(0.111)	(0.151)	(0.214)	(0.136)	(0.193)
Tenure * Treatment * 6th Decile	0.047	-0.074	-0.009	-0.214**	-0.092
	(0.129)	(0.228)	(0.254)	(0.108)	(0.177)
Number of Observations	70414	62260	65702	39741	40369
Number of Individuals	3919	3465	3656	3281	3441
Log Likelihood	-113340.774	-279883.526	-3761432.559	-60249.603	-629560.855
Chi-squared	2896.039	2925.687	1302.279	584.031	435.866

Notes: The table reports results from the estimation of the baseline effort effect model augmented with a dummy variable for productivity deciles, and their double and triple interactions with $Post'02$, $Tenure$ and $Treatment$. The omitted category is below median productivity academics. Productivity deciles are determined on the basis of the averages of the impact factor-rated number of publications over the three pre-announcement years 1999, 2000 and 2001, separately by academic field and treatment group. In Panel B, the control group is restricted to academics that do not switch to the performance pay scheme, where any first affiliation, position or contract renegotiation after implementation of the pay reform (as of 2005) is considered a switch. All other specifications as in the baseline regression of the effort effect. See section 4.2.7 for further details.

Table A.5: Selection Analysis - Robustness Checks

	Above vs Below Median	Cox Proportional Hazard Model	Baseline with Field Strata
Treatment	-0.673	1.214***	-0.114
	(0.537)	(0.415)	(0.438)
Age	-0.323***	-0.120***	-0.314***
	(0.016)	(0.020)	(0.017)
Age * Treatment	0.008	-0.031***	-0.003
	(0.012)	(0.009)	(0.010)
Productivity	-2.456***	-0.036***	-0.055***
	(0.683)	(0.012)	(0.013)
Productivity* Treatment	1.938**	0.031**	0.050***
	(0.880)	(0.013)	(0.014)
Productivity* Age	0.056***	0.001***	0.001***
	(0.015)	(0.000)	(0.000)
Productivity* Age * Treatment	-0.039**	-0.001**	-0.001***
	(0.019)	(0.000)	(0.000)
Age at Tenure	0.195***	0.029	0.201***
	(0.015)	(0.019)	(0.015)
Constant	2.586***		1.549***
	(0.401)		(0.435)
Number of Observations	80131	80131	80131
Number of Switches	2435	2435	2435
Log Likelihood	-7394.647	-21430.395	-7385.161
Chi-squared	1404.077	710.192	1162.541
Rho	1.655		

Notes: The table reports estimation results of Weibull and Cox (second column only) proportional hazard models of contract changes (affiliation or position changes or contract renegotiations) by tenured professors. In the first column, the "Productivity" variable is a dummy 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. In the last two columns, "Productivity" is calculated as the average of the impact factor-rated number of publications over the three pre-implementation years (2002-2004). In the last column, the Weibull model is estimated with strata for academic fields. All other specifications as in Table 4. See section 4.3.2 for further details.

Table A.6: Switching Analysis

	Weibull Model		Cox PH Model	
	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

Notes: The table reports estimation results of Weibull (columns 1a-1b) and Cox (columns 2a-2b) proportional hazard models of contract changes (affiliation or position changes or contract renegotiations) by professors tenured before 2005. "Post" is 1 as of 2005. All other specifications as in Table 4. See sections 4.3.2 and 4.3.3 for further details.

Table A.7: Instrumental Variables Estimation of Effort Effect

Panel A: Age and Age-squared IV	# Publications	IF-rated publications	Citations	Average IF-rating	Average citations
Post'02 * Treatment	2.399***	7.578***	86.587***	-0.207	-23.968
	(0.533)	(2.373)	(29.272)	(0.440)	(18.353)
Contract Change * Treatment	0.725	12.541**	65.253	0.559	-0.642
	(1.061)	(5.655)	(49.292)	(0.737)	(17.236)
Number of Observations	57374	57374	57374	25721	25721
Number of Individuals	3197	3197	3197	2228	2228
Log Likelihood	-157177.424	-249583.010	-392578.018	-54032.603	-139540.827
Test Statistics for Under- and Weak Identification of Instruments, Overidentifying Restrictions and Endogeneity of Regressors					
Kleibergen-Paap rk LM statistic	100.817	100.817	100.817	37.442	37.442
Chi-squared p-value	0.000	0.000	0.000	0.000	0.000
Kleibergen-Paap rk Wald F statistic	28.179	28.179	28.179	9.872	9.872
Stock-Yogo Critical values	[5%; 10%]	[5%; 10%]	[5%; 10%]	[10%; 20%]	[10%; 20%]
Hansen J-statistic	13.799	20.988	17.462	0.204	1.991
Chi-squared p-value	0.001	0	0	0.9029	0.3695
Endogeneity Test	11.552	6.689	9.1	0.767	2.266
Chi-squared p-value	0.003	0.035	0.011	0.6813	0.3221
Panel B: Over-32 and Over-42 IV					
Post'02 * Treatment	1.910***	7.147**	50.690	0.381	-27.113
	(0.608)	(2.933)	(34.866)	(0.496)	(16.718)
Contract Change * Treatment	-0.260	2.704	-29.215	0.099	-2.356
	(1.219)	(6.427)	(65.107)	(0.955)	(26.063)
Number of Observations	57374	57374	57374	25721	25721
Number of Individuals	3197	3197	3197	2228	2228
Log Likelihood	-156924.517	-249105.190	-392319.928	-54031.927	-139576.869
Test Statistics for Under- and Weak Identification of Instruments, Overidentifying Restrictions and Endogeneity of Regressors					
Kleibergen-Paap rk LM statistic	65.318	65.318	65.318	20.401	20.401
Chi-squared p-value	0.000	0.000	0.000	0.000	0.000
Kleibergen-Paap rk Wald F statistic	17.241	17.241	17.241	5.559	5.559
Stock-Yogo Critical values	[5%; 10%]	[5%; 10%]	[5%; 10%]	[20%; >25%]	[20%; >25%]
Hansen J-statistic	3.342	1.37	2.637	3.248	1.179
Chi-squared p-value	0.188	0.504	0.268	0.197	0.555
Endogeneity Test	4.699	2.012	0.903	0.575	2.989
Chi-squared p-value	0.095	0.366	0.637	0.75	0.224

Notes: Panels A and B report the results of 2-step GMM estimation of an instrumented baseline regression. In panel A, the $Post'02 * Treatment$ and $Contract\ Change * Treatment$ interactions are instrumented for by $Post'02 * Synthetic\ Age$, $Post'02 * Synthetic\ Age - Squared$, $Post'05 * Synthetic\ Age$ and $Post'05 * Synthetic\ Age - Squared$ variables. In Panel B, the $Post'02 * Treatment$ and $Contract\ Change * Treatment$ interactions are instrumented for by $Post'02 * Over - 32$, $Post'02 * Over - 42$, $Post'05 * Over - 32$ and $Post'05 * Over - 42$ variables. The instruments are based on synthetic age variables for which unknown ages are imputed using the average age at PhD, habilitation or tenure. The $Over - 32$ and $Over - 42$ variables are 1 whenever the synthetic age of the academic is larger than 32, respecively 42, and 0 otherwise. $ContractChange$ is 1 starting from the year in which an academic who already holds a tenured affiliation, changes their position, affiliation or contract (receives an outside offer) and every year thereafter. $Treatment$ is 1 for academics who experience a contract change as of 2005 and 0 otherwise. All other specifications as in the baseline regression of the effort effect. See online appendix section A.4.2 for further details.