Gender Differences in the Response to Incentives: Evidence from Academia

Erina Ytsma[*][†]

Knowledge work and other high-skilled jobs commonly feature both explicit performance incentives and market-based career-concern incentives. Gender differences in mobility, negotiation attitudes, and performance assessments could cause women to respond less to career-concern incentives. This could have implications for gender gaps in performance, pay and female representation, as well as innovation and knowledge creation. This paper studies whether men and women respond to different types of incentives, and how this affects knowledge output and female representation. Exploiting the introduction of performance pay in German academia as a natural experiment and using a data set that encompasses the affiliations and publication records of the universe of German academics, I provide evidence of gender differences in the causal response to explicit performance incentives and career-concern incentives. I find that, on average, men have a significant effort response to career-concern incentives only, of 16% to 19%, while women have a significant effort response to explicit performance incentives only, of 36% to 40%. Exploring implications for knowledge creation and representation, I find that women increase work that ends up being most highly cited in response to performance pay, while men do not. Yet, I find that more productive women are not more likely to receive academic job offers. This could limit female representation in academia and may negatively impact innovation. Finally, I explore various mechanisms and show that a lack of credit for work in larger teams and consequent signal noise can explain the absence of a response to career concern incentives in women.

Keywords: gender gap, effort and selection effects, career concerns, explicit performance incentives. JEL: J16, M52.

PRELIMINARY AND INCOMPLETE. PLEASE DO NOT CITE OR DISTRIBUTE WITHOUT PERMISSION.

[*]Tepper School of Business, Carnegie Mellon University, Pittsburgh PA 15213, USA. +1-412-268-1117. eytsma@andrew.cmu.edu.

[†]*Acknowledgements*: I would like to thank Philippe Aghion, De Gruyter, Baran Duzce, Alex Eble, Eric French, Ina Ganguli, Bob Gibbons, Ben Jones, Axel Schniederjuergen, Heather Sarsons, Monika Schnitzer, Fabian Waldinger, the ministries of education of the German states and seminar participants at CSWEP @ AEA, LME '25, NBER SI SSF '25, Relational Contracts Workshop '25, BEDI '25 and INSEAD/PSE/College de France for helpful comments, information or data.

Latest version: https://gitlab.com/eytsma/research_papers/-/blob/master/GenderDiffs-IncentiveResponse-Academia_ErinaYtsma.pdf
This version: 26 March 2026

1 INTRODUCTION

Women are underrepresented in academia, as in other high-skilled professions, and this underrepresentation increases at higher ranked positions (Ginther and Kahn, 2004; Lundberg and Stearns, 2019; Auriol et al., 2022; Sherman and Tookes, 2022; Bagues et al., 2024). Several factors have been found to contribute to the underrepresentation of women in academia. The work of female scholars fares differently through the peer review process (Card et al. (2020); Hengel (2022)), and there are differences in the recognition men and women receive for work done in teams (Sarsons et al., 2021). Furthermore, gender-neutral tenure clock stopping policies have a differential impact on men and women (Antecol et al., 2018), there is a gender difference in mobility (Azoulay et al., 2017), as well as in promotion and tenure rates (Ginther and Kahn (2004; 2021)). This paper studies a potentially contributing factor that is common to many high-skilled sectors: gender differences in the response to different kinds of incentives.

Academia and other knowledge work and high-skilled jobs often feature both explicit performance incentives, such as publication bonuses, as well as implicit incentives, such as career concerns (Gibbons and Murphy, 1992; Chevalier and Ellison, 1999; Hong et al., 2000; Coupé et al., 2006; Franzoni et al., 2011; Aghion et al., 2013; Ferrer, 2016). This paper focuses on career concerns in the form of implicit market-based incentives specifically, such as attraction or retention bonuses negotiated in contract talks. The response to explicit incentives has been studied extensively, particularly in more routine tasks (Lazear, 2000; Nagin et al., 2002; Shearer, 2004; Bandiera et al., 2005; Dohmen and Falk, 2011; Leuven et al., 2011) and there is evidence that men and women respond similarly to explicit incentives Bandiera et al. (2021). Career concerns, on the other end, are understudied, especially empirically, and there are reasons to believe that women may respond less strongly to career-concern incentives.

First, women tend to be less mobile (Azoulay et al., 2017). This would weaken their bargaining position in the market (Caldwell and Danieli, 2024). Second, women are less likely to negotiate (Babcock and Laschever, 2009; Biasi and Sarsons, 2022). Both factors would reduce market-based career-concern incentives that derive from bonuses determined in contract negotiations. Thirdly, career concerns are implicit. The performance assessment on the basis of which attraction or retention bonuses are determined could therefore be more subjective. A more subjective performance assessment may be more biased, disadvantaging women (Bagues et al., 2017; Lundberg and Stearns, 2019; Card et al., 2020). For all these reasons, women might respond less strongly to career-concern incentives.

In turn, this may give rise to gender differences in performance and female representation. Companion paper Ytsma (2022) shows, theoretically, that career-concern incentives increase the likelihood of quality crowding and decrease the likelihood of positive selection effects. If women respond (more) to explicit incentives, and men (more) to career-concern incentives, their effort, output and sorting patterns may therefore diverge. This could affect female representation, through hiring decisions and self-selection.

In this paper, I study gender differences in the response to explicit incentives and implicit market-based career-concern incentives in academia. To do so, I exploit the introduction of performance pay in German academia as a natural experiment. This is a useful setting for three reasons. First, the incentives introduced with the reform have differential incidence by tenure and age cohort. This allows me to causally identify both the effort and selection effects of the incentives using difference-in-differences frameworks. Second, the new pay scheme features both explicit and career-concern incentives, and these incentives take effect at different times. This enables me to identify the response to these different types of incentives separately. Third, the reform was nationwide: it applied to all professors at public universities in Germany. Using a data set that encompasses the affiliations, publications and related information of the universe of academics in Germany, I can analyze the effect of the reform's incentives across genders, across academic fields and across productivity quantiles.

To estimate causal effort effects, I use the framework introduced in Ytsma (2022). I exploit 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 are appointed to their first tenured affiliation after the reform are paid according to the performance pay scheme, but academics appointed to their first tenured affiliation before the reform start out in the age-related pay scheme. If the timing of appointment to the first tenured affiliation is exogenous, any differential output change from before to after the reform between academics appointed to their first tenured position just before and just after the reform can be interpreted as the causal effort effect of performance pay.

The effort effect estimation shows that men and women respond to different kinds of performance incentives. In men, I find a positive and significant effort response to career-concern incentives only. This amounts to an increase in research output of 16 to 19%. Women, on the other end, only have a positive and significant effort response to explicit incentives, causing their research output to increase by 36 to 40%. I find no evidence of pre-existing trends and show that responses align closely with the time when the different types of incentives take effect. Furthermore, there are no significant output differences in equivalent placebo D-i-Ds, while results are robust to excluding control academics who switch to performance pay at a later date or weighting research output by the number of contributing authors.

Next, I study implications of this gender difference in the response to different types of incentives for the quality, novelty and impact of the research produced. I find that women produce more of the most highly cited work in response to the performance incentives they respond to. Men do not. I unpack this finding by productivity quantile, and find that more productive women produce more of the highest quality work. The most productive men, on the other hand, increase only low to medium quality work, but not the highest quality work. Medium productive men even produce fewer of the most highly cited papers. All in all, this leads to a decrease in average quality of the research output that men produce in response to career-concern incentives. There is no such decrease for women. Finally, using textual analysis techniques to study the incentive effect on research novelty and impact, I find that women increase work that is not very novel, but which does garner a lot of follow-on work. In contrast, men increase work that is more novel, but which has only medium impact.

The paper proceeds with an analysis of implications for female representation. To study this, I analyze the selection effect of performance pay using the framework in Ytsma (2022). I exploit the fact that 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. I find that more productive men are more likely to switch to performance pay than less productive men. However, more productive women are not more likely to switch to performance pay. In order to examine whether these switching patterns are driven by academics' self-selection, or by the selection of candidates by hiring departments, I test for positive self-selection in another difference-in-differences framework. To do so, I exploit variation along two dimensions: tenure cohort and age. Only academics who already hold a tenured affiliation before the reform can select into the performance pay scheme, by changing position or renegotiating their contract. Accordingly, the treated-control designation for the selection analysis is the opposite of what it was for the effort effect analysis. The second dimension of variation is age, because 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.

I find that there is positive self-selection into performance pay in both men and women: both more productive men and more productive women are more willing to sort into performance pay. To the extent that incentives and pay determine whether someone sorts into academia, this suggests that academia can attract more productive men and women if it features career-concern and explicit performance incentives. However, the fact that more productive women are not more likely to switch to performance pay, at least not a few years into the reform, suggests that more productive women are not more likely to get (better) job offers. That is, even though more productive women appear more willing to self-select into performance pay and - by extension - academia, they may not be given the opportunity to do so, because they don't receive (better) job offers. This could limit female representation.

Finally, I explore possible mechanisms for the differential response to incentives. I find no evidence that this is driven by reduced mobility in women or more biased assessment of female performance. Instead, the absence of a significant effort response to career-concern incentives in women may be a rational response to more productive women not being more likely to get (better) job offers. I dig into possible micro-foundations for this by examining female effort responses to the different types of incentives in fields with large or small team sizes. Women may receive less credit for work in large teams (e.g. Sarsons et al., 2021), rendering their ability signal more noisy and weakening career-concern incentives. Indeed, I find that women do not have a significant effort response to career-concern incentives in fields with large teams - responding to explicit performance incentives instead - but they do respond to career-concern incentives in small team fields. This suggest that credit attribution and signal noise are important drivers for the lack of average effort response to career-concern incentives in women.

The paper makes four main contributions. First, it shows that men and women respond to different types of incentives. When both career-concern incentives and explicit incentives are present, men respond to the former and women to the latter. Second, this difference in responsiveness has implications for innovation. In response to explicit incentives, women produce more work of the highest quality. Men do not produce more of the highest quality work in response to career-concern incentives. This manifests itself as follows: women increase work that is not very novel, but which garners a lot of follow-on work. In contrast, men increase work that is more novel, but which has only medium impact. Third, even though there is positive self-selection by both men and women in response to career-concern and explicit incentives, respectively, more productive women are not more likely to end up in performance pay or (better positions in) academia - ostensibly because they are not more likely to be hired. Fourth, the paper provides causal empirical evidence of different responses to career-concern incentives and explicit incentives, with the former causing quality crowding and non-monotonic responses in ability.

The paper contributes to the literature on gender differences in general, and the position of women in academia in particular (Ginther and Kahn, 2004; Bertrand, 2018; Lundberg and Stearns, 2019; Auriol et al., 2022; Sherman and Tookes, 2022; Cortés and Pan, 2023; Bagues et al., 2024). The paper also contributes to the literature on performance incentives. 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), and mostly in routine tasks with precise output measures (e.g. Lazear, 2000; Shearer, 2004; Bandiera et al., 2005). The contribution of the current paper therefore lies in providing causal evidence of effort and selection effects of explicit incentives as well as implicit career-concern incentives, in a highly complex task with noisy output measures (knowledge creation). As such, the paper contributes to the literature on career concerns as well (e.g. Gibbons and Murphy (1992); Miklós-Thal and Ullrich (2016); Xu et al. (2020); Camargo et al. (2021). To the best of my knowledge, this paper is the first to present causal empirical evidence of differences in the response to career-concern and explicit incentives.

Furthermore, the paper 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 et al., 2014; MacLeod and Urquiola, 2021) and the process and organization of knowledge creation (e.g. Jones, 2009; Waldinger, 2012; Azoulay et al., 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 et al., 2009; Hall and Harhoff, 2012; Hvide and Jones, 2018). In doing so, the academic fields that can be studied are generally restricted to the (applied) sciences. Moreover, it is frequently difficult to differentiate between changes in the production of commercializable knowledge 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 et al., 2011; Chan et al., 2014; Borjas and Doran, 2015; Ghirelli et al., 2023). 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 two common types of performance incentives in knowledge creation: explicit and career-concern incentives. Moreover, the paper studies differences in responses to these incentives in men and women, across all academic fields and productivity levels.

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 et al., 2011, 2016; Graff Zivin and Lyons, 2021), with mixed results. Gibbs et al. (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 differences in the effects of long-term, explicit incentives as well implicit, career-concern incentives, on broad scope innovation, effected over a long time. Moreover, the paper studies gender differences in the response to these incentives.

The rest of the paper is structured as follows: section 2 provides information on the institutional background, section 3 presents the empirical analysis, and section 4 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.[1] Accordingly, I restrict attention to tenured professors in the empirical analyses of the paper. 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 the performance pay scheme, professors can earn performance bonuses in two ways: as attraction or retention bonus, or for on-the-job performance (BMBF, 2002).[2] 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

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

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

students. To be able to negotiate an attraction or retention bonus, professors often need to show proof of another offer (Detmer and Preissler, 2005).

Because past achievements and performance determine if and how much of an attraction or retention bonus academics can negotiate, they have an incentive to improve their performance. As such, the attraction and retention bonuses constitute implicit market-based incentives in the form of career concerns.[3] 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-concern 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; Coupé et al., 2006; Lerner and Wulf, 2007; Aghion et al., 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 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

[3]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 an academic's ability and chances of success, these grant pay supplements give rise to career-concern incentives as well.

and Murphy, 1992; Hong et al., 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-concern 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.

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.[4] 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 9) (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 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 rather than negotiated 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 (BMBF, 2002; BMI, 2007).[5] Combined

[4]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.

[5]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).

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 introduced substantially higher-powered career-concern and explicit performance incentives compared to the age-related pay system. Moreover, the incentives of the performance pay system 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.[6]

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 pay. I exploit such switches to estimate the selection effect of the performance incentives.

[6]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.

3 EMPIRICAL ANALYSIS

3.1 Data Description

To estimate 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. This data is also used in Ytsma (2022), but here gender information from one of the input data files (DGK) is leveraged.

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 output as outcome variable (BMBF, 2002). There were 89 public universities in the years spanned by the panel (HRK, 2014). This constitutes the vast majority of universities in Germany, since there were only a few - small - private universities during this period. 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 (De Gruyter, 2006, 2008, 2010, 2013).[7] 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.).

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 output and use journal impact factors from the ISI journal citation report

[7]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.

(JCR) of the year of publication to construct impact factor-rated publication measures (Clarivate Analytics, 2000-2012, 1993-2012).[8]

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.

To construct an individual-level affiliations panel from the information in DGK and FuL, I take the affiliation information in DGK and use an iterative procedure to fill affiliations over time. I cross-check this information with offer and appointment information from FuL, after backdating by the average printing lag of FuL announcements.[9] 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 successful 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

[8]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.

[9]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.

pay for the selection effect analysis.[10,11]

3.2 Effort Effect

In order to identify effort effects of the career concerns and explicit performance incentives introduced with the pay reform in Germany, I use the framework in Ytsma (2022). I exploit 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.[12] Accordingly, academics who are appointed to their first tenured affiliation after 2004 switch to the performance pay scheme upon their appointment, while academics who are appointed to their first tenured affiliation before 2005 continue to fall under the age-related pay scheme. If the timing of appointment to 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 the performance incentives on knowledge creation by comparing the change in research output 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. 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 (2145 men, 571 women), the control cohort 3,197 (2471 men, 557 women). I construct several measures of research output 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 output measures for (at least) 18 years, from 1993 through 2010. The output measures take all available publication types into account, including journal articles, books, book chapters

[10] 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 of contract switches.

[11] Appendices A3.4-6 in Ytsma (2022) provide supplementary information about data matching procedures and construction of the affiliations panel.

[12] 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.

and conference proceedings.[13] Table 1 provides summary statistics. Appendix figures 14 and 15 depict supplemental descriptive statistics: the percentage of male and female newly tenured professors by year (Fig 14) and the percentage of male and female professors across fields (Fig 15).

3.2.1 Baseline Difference-in-Differences

As in Ytsma (2022), I estimate effort effects in the following difference-in-differences model:

$$E\left[Y_{i,f,t-x_f}|X_{i,f,t}\right] = exp[\alpha_i + \beta_1 Post'02 * Treatment_i + \beta_2 Tenure_{i,j} * Treatment_i + \sum_{j=-7}^{7} ttt_{i,j} + \gamma_t] \tag{1}$$

The dependent variable, $Y_{i,f,t-x_f}$ is an output 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 (2002), and before and after commencement of the first tenured affiliation. As discussed in section 2.1, the career-concerns component of performance pay take effect from the moment the reform is announced in 2002. From that moment, academics who anticipate falling under the performance pay scheme in the future, have an incentive to beef up their CV, to increase their chances of receiving a higher attraction or retention bonus in the future. These career-concern incentives 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 affiliation in the performance pay system thus face strong career concerns, since 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-concern incentives from the differential change in output of about-to-be-tenured academics in the treatment and control cohort from before to after announcement of the reform ($Post'02 * Treatment_i$).

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

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-concern incentives, which remain in effect, and unchanged.[14] I can therefore identify the effort effect of the explicit incentives component of performance pay from the differential change in output in the treatment and control cohort from before to after the start of the first tenured affiliation ($Tenure * Treatment_i$).[15]

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 effect of the career concerns and explicit incentives components of performance pay separately. Because $Post'02$ and $Tenure$ are defined as trend-break variables, the interactions capture persistent differential changes in research output.

The $ttt_{i,j}$ variables are relative time fixed effects in the form of time-to-tenure dummies. They control flexibly for output patterns in the run up to and after tenure that are common to treatment and control cohorts.[16] 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 time-to-tenure (relative time) fixed effects for each cohort, from seven years before to seven years after tenure.[17] If the timing of the start of the first tenured affiliation is exogenous, it follows from Athey and Imbens (2022) that estimation of (1) 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.

[14] The theoretical model in Ytsma (2022) shows that adding explicit incentives to career-concern incentives does not affect the career-concern 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.

[15] 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.

[16] 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 et al., 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-concern incentives take effect at the same calendar time (2002) for all cohorts.

[17] 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 et al., 2021).

I estimate the model as a conditional quasi-maximum likelihood (QML) fixed-effect Poisson model,[18] because the dependent variables are non-negative and highly skewed with a large mass at zero and long right tail. I estimate robust standard errors, clustered at the individual level.

3.2.2 Baseline Results

Figure 1 shows the coefficient estimates and 95% confidence intervals of the $Post'02 * Treatment$ and $Tenure * Treatment_i$ interactions in the baseline regression with different dependent variables. In each panel, the dependent variables are, from left to right: the number of publications, the impact factor-weighted number of publications, the total citations to publications, the average impact factor of publications, and the average number of citations. The blue bars depict the $Post'02 * Treatment$ estimate, the orange bars the $Tenure * Treatment_i$ estimate. Hence the blue bars show the effort response to the career-concern incentives and the orange bars the response to the explicit performance incentives. Panel A shows these estimation results for men, panel B for women.

The figure shows that men have a significant effort response only to career-concern incentives, while women have a significant effort response to explicit performance incentives only. In response to career-concern incentives, men increase the number of publications by 19%, the impact-factor weighted number of publications goes up by 16% and total citations increase 17%.[19] Women increase the number of publications by 36% and the impact-factor weighted number of publications by 40% in response to explicit performance incentives. Furthermore, the average impact factor of the research produced by men decreases. There is no significant decline in the average quality of the research produced by women.

<Figure 1 about here>

3.3 Validity of Identifying Assumptions

Interpretation of the above results as the causal effort effect of the higher-powered incentives of performance pay - specifically career-concern incentives - requires the parallel trends assumption to be met and the timing of the start of the first tenured appointment to be exogenous. Potential threats to identification are other events or reforms happening around the same time or endogenous assignment or selection into treatment.

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 "Excel-

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

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

lence Initiative") as of late 2006/early 2007 (DFG, 2016), the abolition of the professor's privilege in 2002 (Von Proff et al., 2012),[20] and the introduction of the "Junior Professorship" in 2002 as an alternative path to professorships from the habilitation (Lutter and Schröder, 2016). The introduction of the Junior Professorship cannot be driving the results, because the first Junior Professors became eligible for a tenured position by 2008/9 and thus fall outside the treatment and control cohort. The abolition of the professor's privilege applied to all professors at the same time, and hence cannot explain the *differential* output changes between the treated and control cohort either. The Excellence Initiative funded a select set of universities, graduate schools and 'clusters of excellence'. It therefore impacted different universities differentially, but does not affect the treatment and control cohorts differentially either. Besides, the initiative was announced in mid-2005 and the first funding decisions in the initiative were made in late 2006. For both reasons, the initiative cannot be driving the differential output changes between the treated and control cohorts as of 2002. Pre-trend tests and placebo D-i-Ds provide further support for the identifying assumptions, using the same models as in Ytsma (2022).

3.3.1 Pre-trends

To test for pre-trends, I estimate the following conditional QML Poisson fixed effects model of dynamic output differences between the treated and control cohort for men and women:

$$E\left[Y_{i,f,t-x_f}|X_{i,f,t}\right] = exp[\alpha_i + \sum_{k=1}^{15} \beta_k ttt_{i,k-8} * Treatment_i + \sum_{j=-7}^{7} ttt_{i,j} + \gamma_t] \tag{2}$$

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-tenure) for different tenure cohorts, and allows me to estimate differences in output between the treated and control cohorts when they are at the same point in their career trajectory, 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 Figures 2 (men) and 3 (women). 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 the career concern incentives of the performance pay scheme 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.

[20] 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.

The absence of pre-existing trends lends support to the parallel trends assumption. Figure 2 shows that the research output of treated men starts to exceed that of control men once the career-concern incentives take effect (top three figures), while the average quality of this output declines after this point (bottom two figures). For women, the difference in research output becomes significant only once the explicit performance incentives take effect (first two figures). The clear alignment of the output response with the time at which career-concern incentives, respectively explicit performance incentives, take effect - after controlling for field-specific publication lags - supports the interpretation of the baseline results as the causal effort effect of these incentives. It also underlines that men respond to career-concern incentives, while women respond to explicit performance incentives.

<Figure 2 about here>

<Figure 3 about here>

3.3.2 Placebo D-i-D

There are no significant output differences between a placebo treatment and control cohort (see right-hand side figures of Figure 2 and 2 as well as Panel A in Tables 3 and 4 for placebo baseline regression results). 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 cohorts start out in the age-related pay system. If tenure itself were driving the baseline results, and if the time-to-tenure fixed effects in (1) and (2) did not control for all output changes in the run up to and after tenure, we would expect the interactions of the placebo treatment indicator with time-to-tenure fixed effects to show similar, significant output differences as in the baseline dynamic effects plots in the figures on the left in Figures 2 and 3. The right-hand side figures of Figures 2 and 3 show no evidence of this.

The placebo D-i-D also lends support to the exogeneity of tenure timing. If academics in the placebo treatment 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 with a temporary drop after tenure). I find no evidence of this.

Finally, the placebo DiD results also provide evidence that the incentives of the performance pay scheme do not give rise to assignment of academics with different innate ability to different tenure cohorts. The publication record of the '03/'04 cohort (the placebo treatment cohort) at the time of their tenured appointment may have been affected by the higher-powered career-concern incentives of the performance pay scheme, which take effect in 2002.[21] In contrast, the tenured appointment of the '01/'02 tenure cohort (the placebo treatment cohort) cannot have been affected by the (career-concern) incentives of the performance pay scheme, because the appointment decisions will have been made before any output produced in response to these incentives could have been published. If the incentives of the performance pay scheme had resulted in tenure cohorts with differential academic ability distributions, on different output growth paths, this should have showed as output differences between the two placebo cohorts. There is no evidence of this.

<Table 3 about here>

3.3.3 Excluding Switchers

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.[22] 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. 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 control groups that exclude 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 3.1 for details). Results are robust to this alternative specification (see Panel B in Tables 3 and 4).

3.3.4 Number of authors and pages

There is no evidence that the documented changes are the result of strategic co-authorship behavior either. The average number of co-authors on papers does not increase. and the results of the baseline regressions with dependent variables weighted by number of authors (so a paper with three authors counts for one-third) are very similar to the baseline results (see Panel C in Tables 3 and 4). If anything, there is some evidence that women increase the number of single-authored papers in response to career-concern incentives. Papers also do not become significantly shorter. Taken together, these results provide evidence that the response to incentives is a real output response and not the results of strategic

[21]Recall that the data span all academic fields, and publication lags are (much) shorter and publication rates higher than in economics in many other fields.

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

behavior in terms of adding names of befriended academics to papers or writing shorter papers.

4 Implications

The fact that men and women respond to different kinds of incentives may have implications for innovation. Ytsma (2022) shows that career-concern incentives increase the likelihood of quality crowding and decrease the likelihood of positive selection effects. The intuition stems from the fact that market-based career-concern incentives are not set (optimally) by any one firm, but arise from belief updating based on observed output signals. So, while firms can optimally adjust explicit incentives to limit crowding out of more noisily measured output dimensions - such as quality - career-concern incentives are smaller for such dimensions, since noisier output measures are relied on less for belief updating. This is particularly true in low ability classes, when quality and quantity effort are stronger substitutes at lower abilities. Furthermore, since belief updating does not regard disutility from effort or risk, career-concerns can be inefficiently high, especially for high ability academics, making positive selection less likely.

4.1 Implications for Quality, Novelty and Impact

The baseline results already showed that the average quality of the research output of men declines in response to career-concern incentives. I unpack this finding here, by studying quality distributions, response heterogeneity by ability, and novelty and impact of the research produced in response to the incentives.

4.1.1 Quality distributions

To get a better idea of the quality of the work produced in response to performance pay, I analyze the distribution of citations to papers published by treated and control cohorts. I calculate the percentiles of citations 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 receives 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 plots in Figure 4 depict the resulting $Post'02{*}Treatment_i$ (blue bars) and $Tenure{*}Treatment_i$ (orange bars) coefficient estimates and 95% confidence intervals for men (subfigure A) and women (subfigure B).

In response to career-concern incentives, treated men only produce more papers that are low to medium highly cited. They don't produce more of the most highly cited papers. Moreover, treated men even reduce medium-highly cited papers slightly in response to explicit performance incentives. The response is the opposite for women: treated women increase medium, high and most highly cited papers. And, as before, theirs is a significant response to explicit performance incentives only. That is, whereas men do not produce more of the highest quality work in response to the performance incentives they respond to - career-concern incentives, women do produce more high quality work in response to the performance incentives they respond to - explicit performance incentives.

<Figure 4 about here>

4.1.2 Heterogeneous responses by productivity quantile

To unpack the quality distributions further, I analyze quality responses for different academic productivity quantiles. The decrease in average quality observed in men and the quality distribution results shown above could be due to compositional changes in output (e.g. due to high productivity quantiles only producing more low to medium quality output) and/or due to decreases in output quality (quality crowding out) in one or more productivity quantiles. Only the latter would potentially have consequences for female representation.

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.[23] Because the productivity distributions are highly right-skewed, with the median academic having no publications in an average year (cf. Tables 1 and 2), I look at the top three deciles (P90-P70) and below-median academics as one group (M1). The plots in Figures 5 and 6 depict the citation quantile bin distributions separately for each productivity quantile and gender.

Figure 5 shows that the top productivity decile of men does not produce more of the most highly cited papers. In response to career-concern incentives, the most productive men only increase papers in the second and third citation quartiles, by 31% and 27% respectively. Medium/medium-high productive men (P80 and P70) even reduce papers in the top citation decile, by 26% and 34%, respectively.

In contrast, the most productive and medium productive women do produce more of the most highly cited papers in response to performance pay, and this response is large. Although the most productive women decrease top citation decile publications in response to career-concern incentives, by 47%, women in the top and 8th productivity

[23] Using pre-announcement years precludes assignment to productivity quantiles from being affected by (differential) responses to the (career-concern) incentives introduced with the performance pay scheme.

decile produce more papers in the top citation quartile and decile in response to explicit performance incentives, by 69% to 244% (see Figure 6). Overall, this yields more of the most highly cited papers (cf. Figure 4). There is no significant increase in output produced by low productivity women in any citation bin, while medium productive women (P70) also produce more medium-cited papers. Together, this explains why there is no decrease in average quality of the work produced by women in response to performance pay.

<Figures 5 and 6 about here>

4.1.3 Novelty and Impact

To provide evidence on the novelty and impact of the work produced in response to performance incentives by men and women, I perform textual analysis of paper abstracts as in Ytsma (2022). 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 et al., 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 $c_{d,s<t}$ denotes the count of documents that contain the term w and that were published in period $s < t$. 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. After filtering out common stop words, 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 et al., 2019).[24]

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

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

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 $(t,\tilde{t})$.[25] If the focal and comparison publication have abstracts that have no bigrams in common, the cosine similarity is 0. The more bigrams the focal and comparison publication have in common, the closer the cosine similarity is to 1.

As in Kelly et al. (2021), I calculate two different cosine similarities: backward and forward similarity. The backward similarity of focal publication d published in t is calculated as the sum of the cosine similarities between the focal document and comparison publications published in a three year window *before* the focal document was published. 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.[26]

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.

Figure 7-A shows that men produce more papers that are dissimilar to papers published before in the same field in response to career-concern incentives: papers in the bottom quartile and decile of the backward cosine similarity go up by 11 to 20%. At the same

[25] 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.

[26] Relative to similarity metrics based on Latent Semantic Analysis (LSA) (Iaria et al., 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.

time, men reduce the number of papers that are very similar to earlier work in the same field in response to explicit performance incentives: they produce fewer papers in the top two quartiles of the backward cosine similarity by 6 to 13%. But the additional papers produced in response to career-concern incentives only have medium impact (there is an increase in 3rd quartile forward cosine similarity papers of 16%) and there is no significant decrease in papers with little to no impact.

In contrast, women produce more papers that are quite similar to papers that were published in the same field, but which do garner a substantial amount of follow-on research. In response to explicit performance incentives, women produce more papers in all but the bottom decile and quartile bin of backward cosine similarity, by 40 to 63%, and more papers in the 2nd, 3rd and 4th quartile bin of forward cosine similarity, by 49 to 67% (see Figure 7-B).

<Figure 7 about here>

4.2 Implications for Female Representation

So far, the results show that both male and female academics respond to performance incentives by producing more research - albeit different kinds of incentives. Moreover, women produce not only more output, they produce more output that is highly cited. Men do not produce more of the most highly cited research, and the average quality of research declines. Finally, though it may not be the most novel, women produce more work that garners a lot of follow-on research. Men, on the other hand, increase output that is novel but has a more modest impact. This begs the question what implications performance incentives have for female representation. Does performance pay increase the number of women - particularly high-performing women - in academia?

To study this, I analyze the selection effect of performance pay. If selection is positive, more productive academics are more likely to select into performance pay. Selection into academia should follow the same pattern as selection into performance pay, since the latter drives the former (see Proposition 4 in Ytsma (2022)). Hence, if more productive women are more likely to select into performance pay, this may bode well for female representation in academia, as we would expect to see the same selection pattern into academia.

<Figure 8 about here>

4.2.1 Non-Parametric Analysis

As a first test, I analyze the hazard rates of switches to performance pay of men and women 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 to switch pay scheme. They do so by changing affiliation or position, or renegotiating their contract after 2004. Accordingly, I label any such event after implementation of the pay reform as a switch to performance pay (the "failure" event). In what follows, I will refer to contract renegotiations, affiliation and position changes collectively as "contract changes" for brevity. As described in section 3.1, 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, so that I observe their full tenured affiliation history, and hence the moment they become "at risk" of switching.[27] 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 8 shows the Epanechnikov kernel density estimates of the hazard function for switches from age-related to performance pay for male or female academics whose average productivity falls in the top quartile or bottom three quartiles of the average productivity distribution. I base productivity quartiles on the average of the citation-weighted number of publications in the three pre-implementation years (2002-04), calculating quartiles separately by academic field and tenure cohort.[28] Figure 8-A shows that the hazard rate for switching to the performance pay scheme is greater for top quartile productive men throughout. A log-rank test rejects equality of the survival functions at 0.1% significance level. More productive men are thus more likely to sort into performance pay. In contrast, Figure 8-B shows that the more productive women are not more likely to switch to performance pay (log-rank test Chi-squared statistic 0.37 (p-value 0.5437)). Results are similar with alternative productivity quantiles, for instance based on the average impact-factor rated number of publications in 2002-2004 (Figure 16).

These hazard rates are the result of two-sided matching; the result of who receives offers and who accepts offers. Hence, the switching rates for men could be because more productive men are simply more likely to receive (outside) offers. The pattern for women could be because women do not self-select into or accept performance pay, or it could

[27] 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. To avoid this, I restrict attention to academics tenured after 1998. As in the effort effect analysis, I also restrict the sample to academics who 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.

[28] Using pre-implementation years ensures quartiles are based on the most recent productivity data before switching decisions can be made, precludes assignment to productivity quartiles from being affected by switching decisions, and prevents simultaneity bias. Results are robust to using quartiles based on the averages of the impact factor-rated number of publications published in the pre-announcement years 1999-2001.

be that more productive women were not more likely to get (better) job offers. In order to distinguish between these two forces, I estimate the direction of self-selection into performance pay in another difference-in-differences framework.

<Figure 9 about here>

4.2.2 Difference-in-Differences Estimation

To identify the (self-)selection effect of performance pay, I use the framework in (Ytsma, 2022) and exploit variation along two dimensions: tenure cohort and age. 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 after 2004. Hence academics who start their first tenured position before 2005 are the treated cohort here, while academics who make tenure after 2004 form the control cohort. 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 (self-)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.

Selection incentives are weaker for older academics due to a single-crossing property of the basic wage schemes for age-related and performance pay (cf. Figure 9). The performance-based scheme pays a basic wage on top of which academics can earn substantial bonuses. The age-related pay scheme pays a basic wage that increases every two years until the age of 49. Hence for older academics, the basic wage is lower than the wage they would receive under the age-related pay scheme. Consequently, older academics 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.

To test this, I estimate the following Weibull proportional hazard model:

$$\begin{aligned}\lambda_{i,t} = \rho * exp[\beta_0 + \beta_1 Treat_i + \beta_1 Age_{i,t} + \beta_2 AvgProd_i + \beta_3 Age_{i,t} * Treat_i + \beta_4 AvgProd_i * Treat_i \\ + \beta_1 AvgProd_i * Age_{i,t} + \beta_2 AvgProd_i * Age_{i,t} * Treat_i + X_i + u_{i,t}] * t^{\rho-1} \quad (3)\end{aligned}$$

Here $\lambda_{i,t}$ is the hazard (risk) of changing contract for academic i at time t and $Treat$ is a treatment group indicator.[29] $AvgProd$ is an academic's average productivity calculated as three-year pre-implementation averages (2002-2004) of the citation-weighted number of publications[30]. 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[31]. 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 Panels A and C of Table 5 report estimation results of specifications without and with age-at-tenure as additional control, respectively. It follows from the discussion above that $AvgProd * Treat$ should be positive and $AvgProd_i * Age_{i,t} * Treat_i$ negative if there is positive (self-) selection into performance pay.

The estimation results for men, in 5 Panel A, show that more productive men who are treated are more likely to switch to performance pay ($AvgProd * Treat$ is positive and significant), but more productive men who are treated and who are older are only marginally less likely to select into performance pay ($AvgProd_i * Age_{i,t} * Treat_i$ is negative, but small and either not significant (column 2a) or significant at 10% only (column 2b)). That is, the selection effect of performance pay, net of general differential sorting patterns by age and productivity types, appears to be only weakly positive for men, in line with the theoretical prediction of Ytsma (2022) that positive selection is less likely in response to career-concern incentives.

In contrast, 5 Panel C shows that there is positive and significant self-selection into performance pay for women. More productive women who are treated are more likely to select into performance pay, but less so if they are older (columns 2a and 2b). Hence treated women need to be of even higher productivity to want to switch into performance pay when they're older. This too aligns with the theoretical prediction of Ytsma (2022) that positive selection is more likely in response to explicit performance incentives.

Reassuringly, I do not find any such significant effects in equivalent placebo estimations (5 Panels B and C). 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

[29] 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.

[30] Using pre-implementation years ensures quartiles are based on the most recent productivity data before switching decisions can be made, precludes assignment to productivity quartiles from being affected by switching decisions, and prevents simultaneity bias.

[31] 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. To avoid this, I restrict attention to academics tenured after 1998.

selection incentives. The estimation results of the placebo estimations for both men and women show that 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.

<Tables 5 and 6 about here>

4.2.3 Implications

Comparing the positive self-selection findings from the difference-in-differences estimation above with the switching patterns from the hazard analysis in section 4.2.1 suggests that more productive women may not be more likely to receive (better) outside offers. The hazard function in section 4.2.1 shows that more productive women are not more likely to end up in the performance-based scheme. The D-i-D estimation on the other hand shows that there is positive self-selection into performance pay for women. Hence, the lack of higher switching rates of more productive women has to be because departments are not more likely to extend them a job offer. This would limit female representation - despite women's apparent willingness to self-select into performance pay (and, by extension, academia), and despite women's strong positive response to performance incentives in terms of output quantity and quality.

It can also explain why women do not respond to market-based career-concern incentives on average. I turn to this next.

5 Mechanisms

5.1 Rational Response to Absence of Job Offers

Market-based career concern incentives require job offers and competition for candidates. Absent (outside) offers, workers are unlikely to be able to negotiate higher pay. In particular, we would expect more productive workers to be in higher demand and receive more offers. This would strengthen their bargaining position and should yield higher pay. Conversely, if more productive workers are not more likely to receive outside offers, they are less likely to be able to negotiate higher pay, and they have less incentive to exert effort to appear more productive. The evidence in section 4.2 shows this to be the case for women. Even though more productive women are more likely to self-select into performance pay, they are not more likely to end up in performance pay. Hence more productive women appear not to receive job offers at a higher rate. But if more productive women are not more likely to receive outside offers, women have no incentive to exert effort to appear more productive. This means that they effectively do not face market-based career-concern incentives, and hence the absence of a response to career-concern incentives is rational.

This does beg the question why - or when - more productive women are not more likely to receive outside offers. In what follows, I explore three possible reasons: signal noise, evaluation bias, and lack of mobility.

5.2 Signal Noise

If the output signal that the market receives is a noisier signal of the ability of women than men (higher signal-to-noise ratio), the market does not update their beliefs about women's ability as much based on observed output. This gives women weaker career-concern incentives to exert effort and produce more or higher quality output, because more/better output will translate into smaller increases in expected pay (Holmstrom1982; 1999).

To test whether a higher signal-to-noise ratio can explain the lack of career-concern incentive responses in women, I compare the effort response of women in fields with large or small research teams. Work produced in larger teams may be a noisier signal of individual ability, particularly for women, who have been shown to receive less credit for research performed in large teams (Sarsons et al., 2021; Bircan et al., 2024)). If this is the case, women should have a positive effort response to career-concern incentives in fields with small research teams, but a weak/no response to career-concern incentives in fields with large research teams. Instead, in fields with large research teams, women should have a positive effort response to explicit performance incentives, which make up for the lack of career-concern incentives.

I use two definitions of fields with large/small research teams: above/below median average team size, and below/above median proportion of single-authored papers. To generate these metrics, I use a finer categorization of academic fields than used in FuL, to allow for better separation of lab and non-lab fields, in line with Tripodi et al. (2025).[32] For each field, I calculate the average number of authors on a paper published in a publication of the field (journal, proceeding, etc.), for all publications in the Web of Science database published in 1993-2012 with at least one author with a German affiliation. I then run separate baseline effort effect regressions (1) for female academics in large and small team size fields[33]. The results are depicted in Figure (10).

Panel A in Figure (10) shows that there is no significant effort response to career-concern incentives in women in fields with above median average team size: none of the blue bars indicating the $Post'02{*}Treatment$ coefficients are statistically different from zero. In contrast, in fields with below median average team size, women do have a significant

[32] My classification is a concatenation of the fields in Tripodi et al. (2025) and FuL.

[33] For this, I assign academics to the field they have published in most during my sample period.

effort response to career-concern incentives, particularly in output quantity (Panel B). We see a similar pattern for below vs. above median single-author fraction fields: women have no significant effort response to career-concern incentives in fields with a below median fraction of single authors (panel C), but they do in fields with an above median fraction of single-author teams (panel D). This aligns with women receiving less credit for work performed in large teams, and hence the signal of their ability being noisier when most of their output is produced as part of large teams.

Furthermore, Figure (10) shows that in fields with large teams (panels A and C), women do have a significant effort response to explicit performance incentives (orange bars). They do not in fields with small teams, where women have a significant effort response to career-concern incentives (panels B and D, blue bars). This aligns with the prediction that explicit performance incentives are stronger when career-concern incentives are weaker (Gibbons and Murphy, 1992).

For men, there is no such difference in effort response to the two types of incentives across fields with large or small research teams. Figure (17) shows that men have a significant effort response across large and small research team fields.

Taken together, this suggests that lack of credit for joint work, and thus noisiness of the signal of ability that the market perceives for women in many academic fields, is a driver of the lack of an average effort response to career-concern incentives in women in the baseline regression findings in Figure (1). In the absence of strong market-based career-concern incentives, women respond to explicit performance incentives instead. If this is the case indeed, we should also see that more productive women are not more likely to receive outside offers in fields with large teams, so that market-based career concern incentives are moot, while there should be differential offer rates for women in small team fields.

To test this, I revisit the switching rate analysis from section (4.2.1), and estimate switching rates separately for high/low productivity women in large team fields and low team fields. The results are show in Figure (11). Panel a shows that more productive women are not more likely to switch to performance pay in fields with below median single-author fraction papers. In contrast, in fields with above median single-author fraction publications, more productive women are more likely to switch to performance pay. Combined with the earlier positive self-selection results for women, this aligns with more productive women having higher offer rates in fields with small research teams. In turn, this explains the significant effort response to career-concern incentives in those fields. In small teams women receive (more) credit for their work, hence signal noise is limited, more productive women are more likely to receive offers, and women - rationally - exert more effort since they do face market-based career-concern incentives.

5.3 Bias in Performance Evaluation

The second mechanism I explore, is potential bias in performance evaluation. Career concerns are implicit incentives and therefore potentially more subjective. If there is bias in the assessment of the performance of men and women, with women's performance undervalued, this could weaken career-concern incentives for women. To examine this, I compare the female effort response in fields with a higher or lower fraction of female tenured professors. If the performance (output) of women is evaluated more favorably if more women are involved in the assessment, career-concern incentives should be stronger for women in those fields.

To test this, I compare the effort responses of women in fields with above/below median female representation (measured as fraction of female tenured professors during the sample period). Figure (12) shows that women do not have a significant effort response to career-concern incentives (blue bars) in fields with low female representation, but neither do they in fields with higher female representation. I therefore find no evidence to suggest that differences in performance evaluation bias drive the absence of an effort response to career concern incentives in women on average.

5.4 Low mobility

In order to look at the potential effect of low mobility, I compare the effort response of married women to women who are not married. Married women may be less mobile. This would make it less likely that they accept an outside offer. This could reduce their bargaining power in the market and hence make career-concern incentives weaker.

In order to test this, I use double last names as a proxy for being married. In Germany, women would traditionally take their husband's last name upon marriage, and they can add their maiden name as a second last name, often hyphenated to his (Luehrs, ????). Women with a double last name are almost surely a subset of married women. This would lead me to underestimate any differences in the response to incentives between women who are married and those who are not.

Here, as before, I estimate the baseline effort effect regression separately for women with a double last-name and those without. Figure 13 shows that there is no significant effort response to career-concern incentives in women without a hyphenated last name. As in the baseline effort regression, they only have a significant effort response to explicit incentives. For women with a hyphenated last name, the response to explicit performance incentives in terms of (impact factor-weighted) number of publications and total citations is at least as large as that of women who do not appear to be married. In fact, the average quality of research of married women also shows a positive effort response to career-concern incentives. This is the opposite of what would be expected if married women were less mobile and in a weaker bargaining position, and suggests that (lack of mobility) cannot explain the lack of response to career concern incentives in women on average.

6 CONCLUSION

This paper provides causal evidence of differential effort and selection effects of performance incentives in male and female academics. I find that men and women respond to different kinds of performance incentives. Male academics have a significant effort response to career-concern incentives only, of 16% to 19%. Female academics have a significant effort response to explicit performance incentives only, of 36% to 40%. I find no evidence that this is due to differences in negotiating power due to women's more limited mobility, or more biased assessments of women's performance.

The dichotomous response to incentives has implications for performance and female representation. In response to performance pay, women increase work that ends up being most highly cited, while men do not. This is because more productive female academics increase work of the highest quality. In contrast, the most productive male academics do not increase work of the highest quality, and medium productive men even decrease it. Men do increase work that is more novel, but which has only medium impact. Women increase work that is less novel, but which garners more follow-on research. Finally, I find that there is positive self-selection of both men and women into performance pay. This would suggest academia can attract more productive men and women if it features career-concern and explicit performance incentives. Yet I also find suggestive evidence that more productive women may not be more likely to receive (better) job offers. This could limit female representation in academia.

The finding that more productive women may not be more likely to receive academic offers is puzzling, because they do display a strongly positive, high(er) quality effort response. It begs the question why this is the case. One possible reason may lie in the finding that men produce more novel work in response to performance incentives, while women produce more work that is more similar to existing work. If novelty is easier to measure than impact or highly valued and an important determinant in hiring decisions, the differential novelty response could explain the relative lack of job offers for (highly productive) women. This could explain the absence of a significant effort response to career-concern incentives in women and has implications for female representation. Moreover, such hiring decisions could be detrimental for an institution's research impact, since the work that men produce in response to performance incentives garners only modest follow-on research, while women's work garners more follow-on research. It therefore seems important to examine to what extent novelty is a deciding factor in hiring decisions, and how this affects both female representation and research impact. I examine this in ongoing work.

When it comes to the effects of performance incentives in academia in men and women, it is an open question whether the benefits outweigh the costs. The paper shows that performance incentives significantly increase effort towards research quantity in both men and women. Women also increase research quality by producing more of the most highly cited work. 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 of women is not the most novel and the additional output of men is not the most impactful. It could be that some of the time dedicated to producing research was previously enjoyed as leisure or spent providing consulting services. Or perhaps the additional research effort comes at the cost of teaching or mentoring effort. In the latter case, whether teaching quality declines or improves will depend on whether research and teaching effort are substitutes 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 the effort and selection of men and women 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 gender differences in the response to incentives and the implications this may have for innovation and female representation 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 career-concern incentives may thus not be as strong. It would be interesting to see to what extent there are still gender differences in the response to incentives in such settings. It would therefore be valuable to study how the publicness of output and the mobility of agents affects the effort and selection responses to performance incentives in men and women. This is left for future research as well.

References

AGARWAL, R. AND P. GAULE (2020): "Invisible Geniuses: Could the Knowledge Frontier Advance Faster?" *American Economic Review: Insights*, 2, 409–24.

AGHION, P., M. DEWATRIPONT, C. HOXBY, A. MAS-COLELL, AND A. SAPIR (2010): "The Governance and Performance of Universities: Evidence from Europe and the Us," *Economic Policy*, 25, 7–59.

Aghion, P., J. Van Reenen, and L. Zingales (2013): "Innovation and Institutional Ownership," *The American Economic Review*, 103, 277–304.

Antecol, H., K. Bedard, and J. Stearns (2018): "Equal but inequitable: Who benefits from gender-neutral tenure clock stopping policies?" *American Economic Review*, 108, 2420–2441.

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

Auriol, E., G. Friebel, A. Weinberger, and S. Wilhelm (2022): "Underrepresentation of women in the economics profession more pronounced in the United States compared to heterogeneous Europe," *Proceedings of the National Academy of Sciences*, 119, e2118853119.

Azoulay, P., W. Ding, and T. Stuart (2009): "The Impact of Academic Patenting on the Rate, Quality and Direction of (public) Research Output," *The Journal of Industrial Economics*, 57, 637–676.

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

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

Azoulay, P., I. Ganguli, and J. G. Zivin (2017): "The mobility of elite life scientists: Professional and personal determinants," *Research policy*, 46, 573–590.

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

Babcock, L. and S. Laschever (2009): "Women don't ask: Negotiation and the gender divide," in *Women Don't Ask*, Princeton University Press.

Bagues, M., M. Makany, G. Vattuone, and N. Zinovyeva (2024): "Women in Top Academic Positions: Is There a Trickle-down Effect?" .

Bagues, M., M. Sylos-Labini, and N. Zinovyeva (2017): "Does the gender composition of scientific committees matter?" *American Economic Review*, 107, 1207–1238.

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

Bandiera, O., G. Fischer, A. Prat, and E. Ytsma (2021): "Do women respond less to performance pay? Building evidence from multiple experiments," *American Economic Review: Insights*, 3, 435–454.

Bertrand, M. (2018): "Coase lecture–the glass ceiling," *Economica*, 85, 205–231.

Biasi, B. and S. Ma (2022): "The Education-innovation Gap," Tech. rep., working paper, Yale.

Biasi, B. and H. Sarsons (2022): "Flexible wages, bargaining, and the gender gap," *The Quarterly Journal of Economics*, 137, 215–266.

Björk, B.-C. and D. Solomon (2013): "The Publishing Delay in Scholarly Peer-reviewed Journals," *Journal of Informetrics*, 7, 914–923.

BMBF (2002): "Gesetz Zur Reform Der Professorenbesoldung," .

BMI (2007): "Bericht Zum Besoldungsrechtlichen Vergaberahmen Bei Der Professorenbesoldung Nach p. 35 Abs. 5 Bundesbesoldungsgesetz," Tech. rep., Bundesministerium des Innern.

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

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

Borusyak, K., X. Jaravel, and J. Spiess (2021): "Revisiting Event Study Designs: Robust and Efficient Estimation," *Arxiv Preprint Arxiv:2108.12419*.

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

Boudreau, K. J., K. R. Lakhani, and M. 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, 140–165.

Caldwell, S. and O. Danieli (2024): "Outside Options in the Labour Market," *Review of Economic Studies*, rdae006.

Callaway, B. and P. H. SantAnna (2021): "Difference-in-differences with Multiple Time Periods," *Journal of Econometrics*, 225, 200–230.

Camargo, B., F. Lange, and E. Pastorino (2021): "On the Role of Learning, Human Capital, and Performance Incentives for Wages," Tech. rep., Working Paper.

Card, D., S. DellaVigna, P. Funk, and N. Iriberri (2020): "Are referees and editors in economics gender neutral?" *The Quarterly Journal of Economics*, 135, 269–327.

Chan, H. F., B. S. Frey, J. Gallus, and B. Torgler (2014): "Academic Honors and Performance," *Labour Economics*, 31, 188–204.

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

Clarivate Analytics (1993-2012): "Isi Web of Science," .

——— (2000-2012): "Journal Citation Report," .

Cortés, P. and J. Pan (2023): "Children and the remaining gender gaps in the labor market," *Journal of Economic Literature*, 61, 1359–1409.

Coupé, T., V. Smeets, F. Warzynski, et al. (2006): "Incentives, Sorting and Productivity along the Career: Evidence from a Sample of Top Economists," *The*

Journal of Law, Economics, and Organization, 22, 137–167.

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

De Gruyter (2006): "Kuerschners Deutscher Gelehrten Kalender," cd-rom.

——— (2008): "Kuerschners Deutscher Gelehrten Kalender," cd-rom.

——— (2010): "Kuerschners Deutscher Gelehrten Kalender," cd-rom.

——— (2013): "Kuerschners Deutscher Gelehrten Kalender Online," .

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

Detmer, H. and U. Preissler (2004): "Abenteuer W, Strategien, Risiken Und Chancen," 308–311.

——— (2005): "Die Neue Professorenbesoldung, Ein Ueberblick," 256–258.

——— (2006): "Die W-besoldung Und Ihre Anwendung in Den Bundeslaendern," *Beitraege Zur Hochschulforschung*, 28, 50–65.

DFG (2016): "Excellence Initiative 2005-2017," .

DHV (1999-2013): "Forschung Und Lehre," .

Dohmen, T. and A. Falk (2011): "Performance Pay and Multidimensional Sorting: Productivity, Preferences, and Gender," *The American Economic Review*, 556–590.

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

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

Expertenkommission (2000): "Bericht Der Expertenkommission - Reform Des Hochschuldienstrechts," Tech. rep.

Ferrer, R. (2016): "The Effect of Lawyers' Career Concerns on Litigation," Tech. rep., Working Paper.

Flämig, C., V. Grellert, O. Kimminich, E.-J. Meusel, H.-H. Rupp, D. Scheven, H. J. Schuster, and F. Stenbock-Fermor (2013): *Handbuch des Wissenschaftsrechts*, Springer-Verlag.

Franzoni, C., G. Scellato, and P. Stephan (2011): "Changing incentives to publish," *Science*, 333, 702–703.

Gentzkow, M., B. Kelly, and M. Taddy (2019): "Text As Data," *Journal of Economic Literature*, 57, 535–74.

Ghirelli, C., E. Havari, E. C. Meroni, and S. Verzillo (2023): "The long-term causal effects of winning an ERC grant," Tech. rep., IZA Discussion Papers.

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

505.

Gibbs, M., S. Neckermann, and C. Siemroth (2017): "A Field Experiment in Motivating Employee Ideas," *Review of Economics and Statistics*, 99, 577–590.

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

Ginther, D. K. and S. Kahn (2004): "Women in economics: moving up or falling off the academic career ladder?" *Journal of Economic perspectives*, 18, 193–214.

——— (2021): "Women in academic economics: Have we made progress?" in *AEA Papers and Proceedings*, American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203, vol. 111, 138–142.

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

Graff Zivin, J. and E. Lyons (2021): "The Effects of Prize Structures on Innovative Performance," *AEA Papers and Proceedings*, 111, 577–81.

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

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

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

Handel, K. C. (2005): *Die Umsetzung Der Professorenbesoldungsreform in Den Bundesländern*, CHE.

Hengel, E. (2022): "Publishing while female: Are women held to higher standards? Evidence from peer review," *The Economic Journal*, 132, 2951–2991.

Holmström, B. (1982): "Managerial Incentives Schemes-a Dynamic Perspective," *Essays in Economics and Management in Honor of Lars Wahlbeck*.

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

Hong, H., J. D. Kubik, and A. Solomon (2000): "Security Analysts' Career Concerns and Herding of Earnings Forecasts," *The Rand Journal of Economics*, 121–144.

HRK (2014): "Hochschulkompass," .

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

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

Jochheim, L. (2015): "Leistungsförderung durch Leistungszulagen? Effekte der W-

Besoldung untersucht am Beispiel deutscher Universitätsprofessoren," .

Jones, B. F. (2009): "The Burden of Knowledge and the Death of the Renaissance Man: Is Innovation Getting Harder?" *The Review of Economic Studies*, 76, 283–317.

Kelly, B., D. Papanikolaou, A. Seru, and M. Taddy (2021): "Measuring Technological Innovation Over the Long Run," *American Economic Review: Insights*, 3, 303–20.

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

——— (2008): "Incentives and Invention in Universities," *The Rand Journal of Economics*, 39, 403–433.

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

Lazear, E. P. and P. Oyer (2012): "Personnel Economics," *The Handbook of Organizational Economics*, 479.

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

Leuven, E., H. Oosterbeek, J. Sonnemans, and B. Van Der Klaauw (2011): "Incentives Versus Sorting in Tournaments: Evidence from a Field Experiment," *Journal of Labor Economics*, 29, 637–658.

Luehrs, J. (????): "Mueller-Schmidt, Schmidt oder Mueller," .

Lundberg, S. and J. Stearns (2019): "Women in economics: Stalled progress," *Journal of Economic Perspectives*, 33, 3–22.

Lünstroth, P. (2011): "Leistungslohn Und Kooptation-eine ökonomische Analyse Der Reform Der Professorenbesoldung," Ph.D. thesis, Universitaet Trier.

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

Macleod, B. (2022): *Beyond Price Theory*, MIT Press.

MacLeod, W. B. and D. Parent (2000): "Job Characteristics and the Form of Compensation Form of Compensation," in *Research in Labor Economics*, Emerald Group Publishing Limited.

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

McCormack, J., C. Propper, and S. Smith (2014): "Herding Cats? Management and University Performance," *The Economic Journal*, 124, F534–F564.

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

Nagin, D. S., J. B. Rebitzer, S. Sanders, and L. J. Taylor (2002): "Monitoring, motivation, and management: The determinants of opportunistic behavior in a field experiment," *American Economic Review*, 92, 850–873.

Oeffentlicher Dienst (2004): "Gesetz Ueber Die Erhoehung Von Dienst- Und Versorgungsbezuegen in Bund Und Laendern 2003/2004," Accessed 11.08.2015.

Oyer, P. and S. 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.

Sarsons, H., K. Gërxhani, E. Reuben, and A. Schram (2021): "Gender differences in recognition for group work," *Journal of Political economy*, 129, 101–147.

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

Sherman, M. G. and H. E. Tookes (2022): "Female representation in the academic finance profession," *The Journal of Finance*, 77, 317–365.

Silva, J. S. and S. Tenreyro (2006): "The Log of Gravity," *The Review of Economics and Statistics*, 88, 641–658.

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

Universitaet Regensburg, U. (2016): "Grundsaetze Der Universitaet Regensburg Zur Vergabe Von Leistungsbezuege," .

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

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

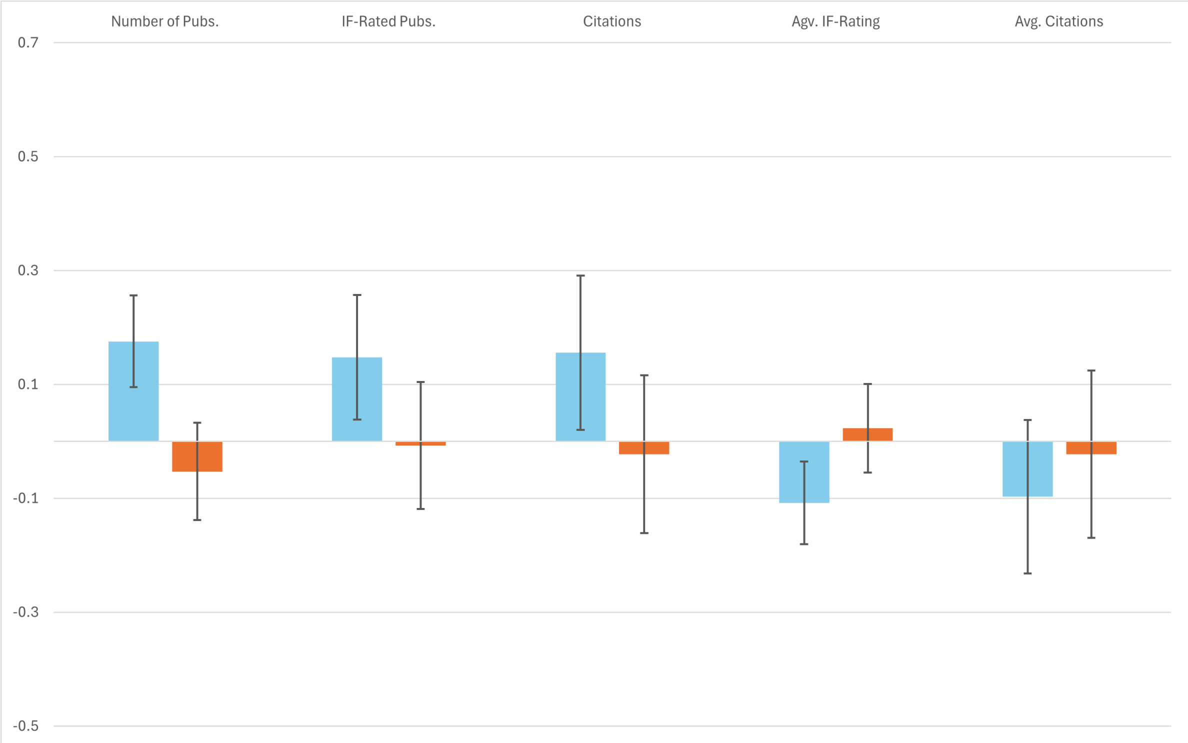
Wissenschaftsrat (2005): "Empfehlungen zur Ausgestaltung von Berufungsverfahren," .

Wu, Y. and F. Zhu (2022): "Competition, Contracts, and Creativity: Evidence from Novel Writing in a Platform Market," *Management Science*.

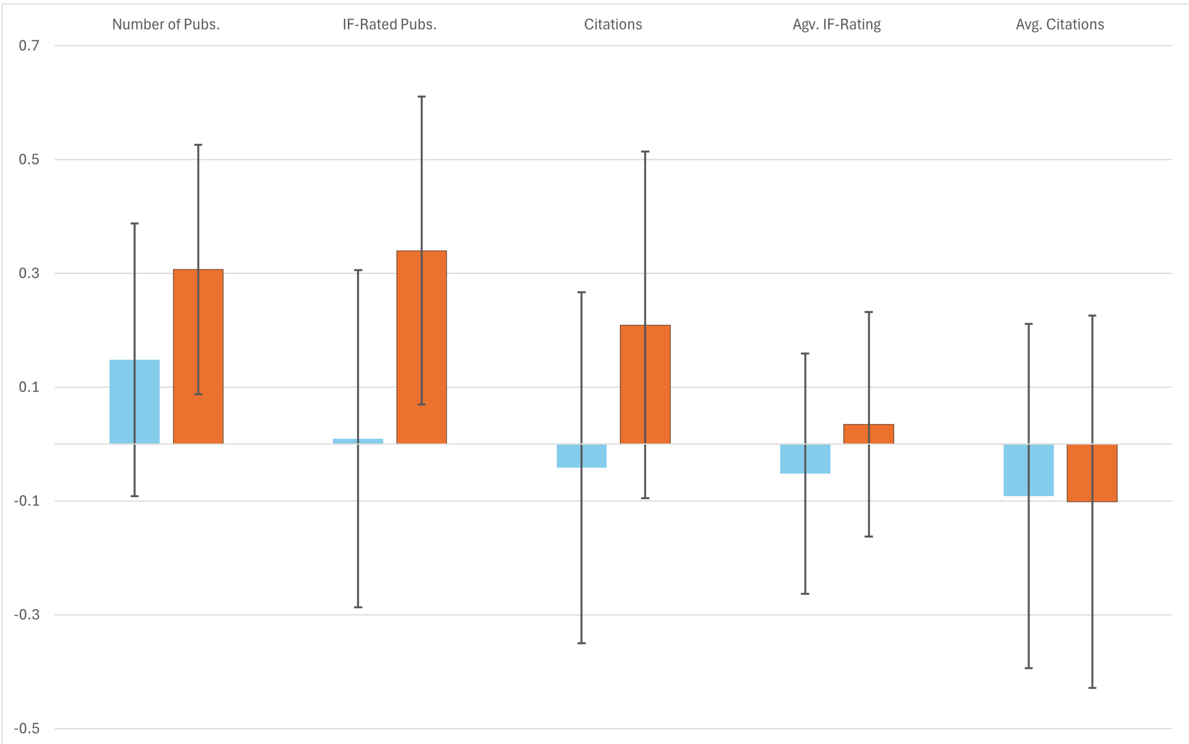
Xu, L., T. Nian, and L. Cabral (2020): "What Makes Geeks Tick? A Study of Stack Overflow Careers," *Management Science*, 66, 587–604.

Ytsma, E. (2022): "Effort and Selection Effects of Performance Pay in Knowledge Creation," *CESifo working paper*.

Figure 1: Effort Response



(a) Men

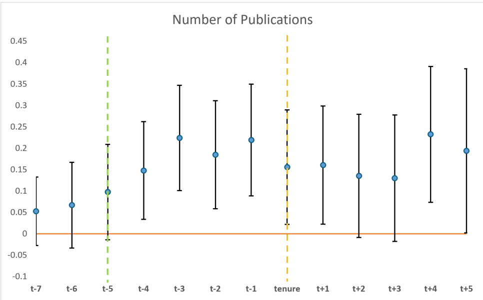


(b) Women

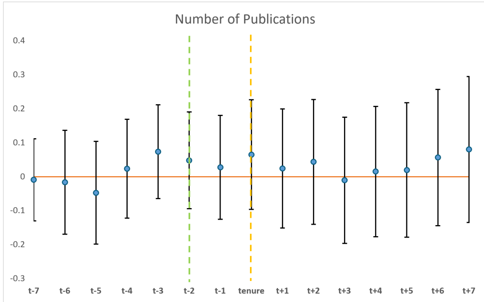
Notes: The plots depict the coefficient estimates and corresponding 95% confidence intervals of the $post'02 * Treatment$ (blue bars) and $Tenure * Treatment$ (orange bars) interactions from separate regressions with the following dependent variables: number of publications, impact factor-rated number of publications, total number of citations, average impact factor rating, and average number of citations. In sub-figure A, the sample is restricted to men and in sub-figure B to women. See section 3.2.1 for further details.

Treatment vs. Control

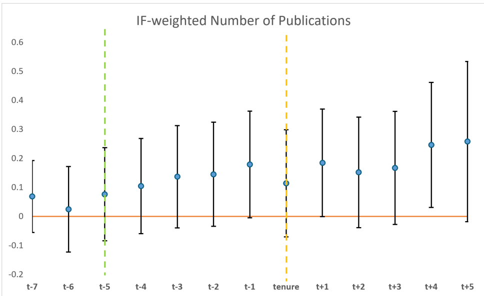
Placebo



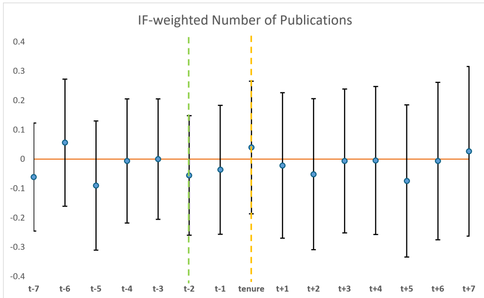
(c) Number of Publications



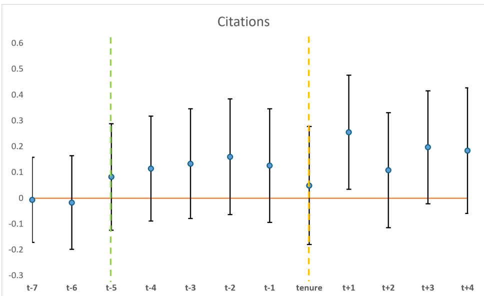
(d) Number of Publications



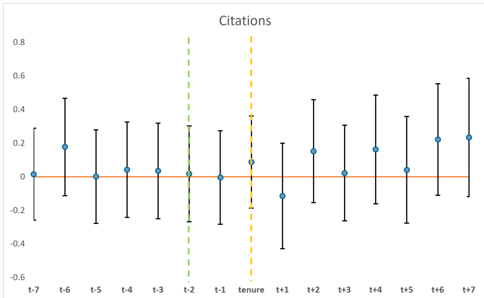
(e) Impact Factor-Rated Number of Publications



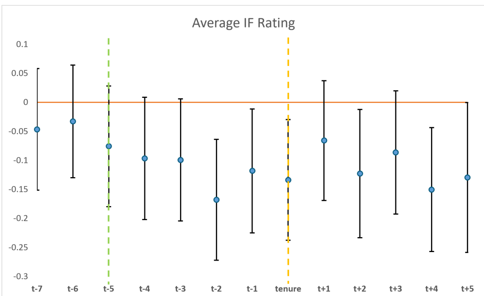
(f) Impact Factor-Rated Number of Publications



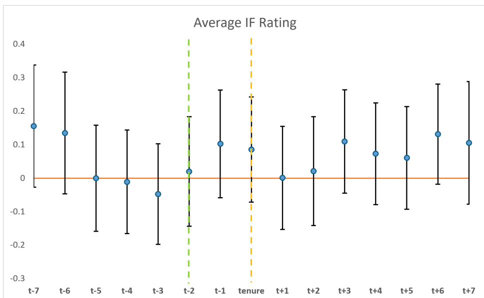
(g) Total Number of Citations



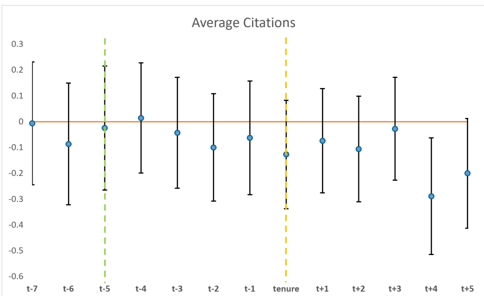
(h) Total Number of Citations



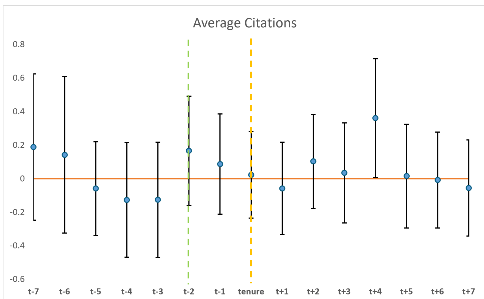
(i) Average Impact Factor Rating



(j) Average Impact Factor Rating



(k) Average Number of Citations



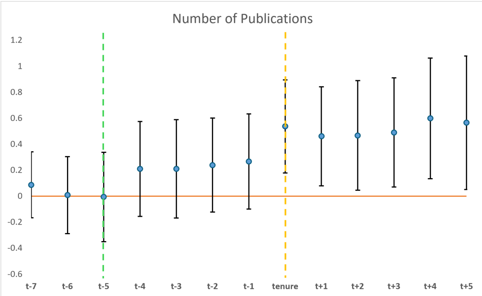
(l) Average Number of Citations

Figure 2: Pre-trends and Effect Dynamics - Men

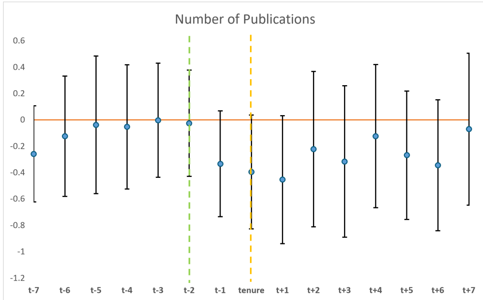
Notes: The plots depict the coefficient estimates and corresponding 95% confidence intervals of the interactions of a treatment dummy and time-to-tenure fixed effects in regressions with the following dependent variables: number of publications, impact factor-rated number of publications, total number of citations, average impact factor rating, and average number of

Treatment vs. Control

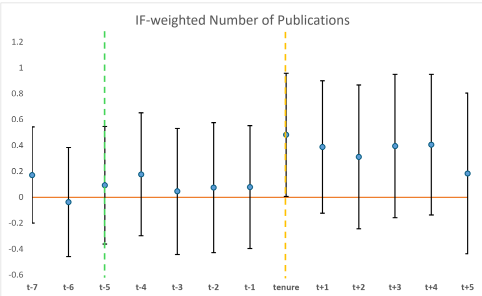
Placebo



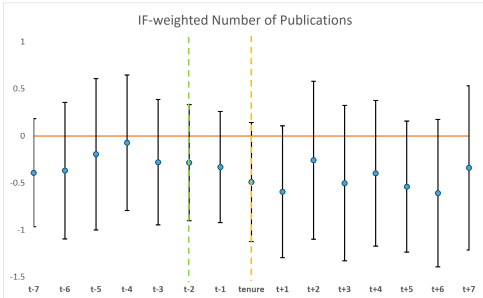
(a) Number of Publications



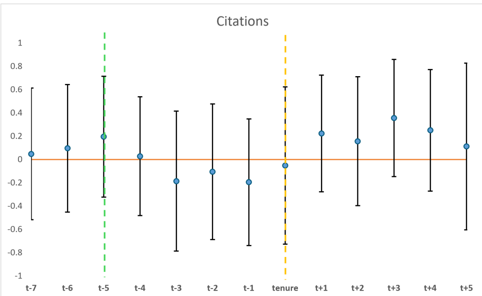
(b) Number of Publications



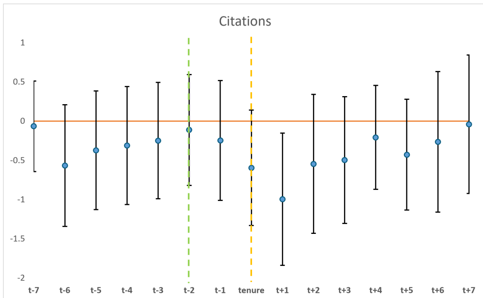
(c) Impact Factor-Rated Number of Publications



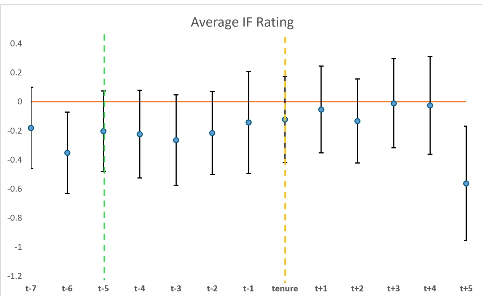
(d) Impact Factor-Rated Number of Publications



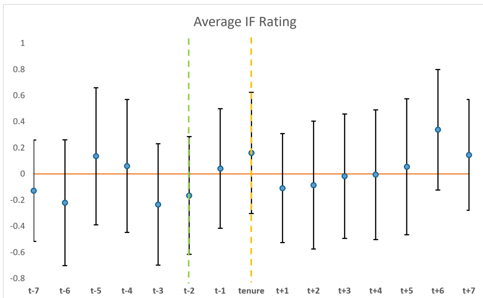
(e) Total Number of Citations



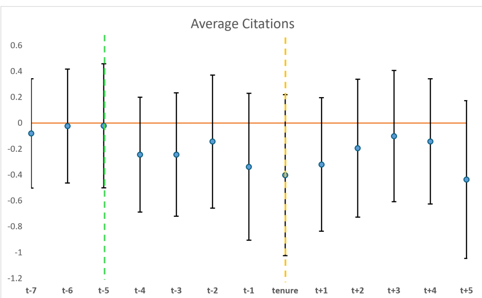
(f) Total Number of Citations



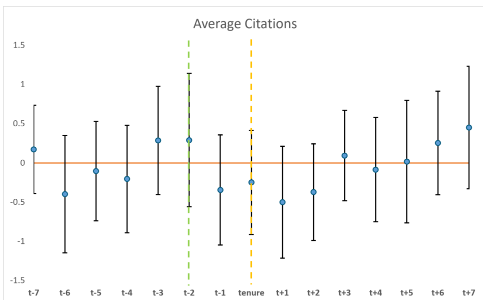
(g) Average Impact Factor Rating



(h) Average Impact Factor Rating



(i) Average Number of Citations

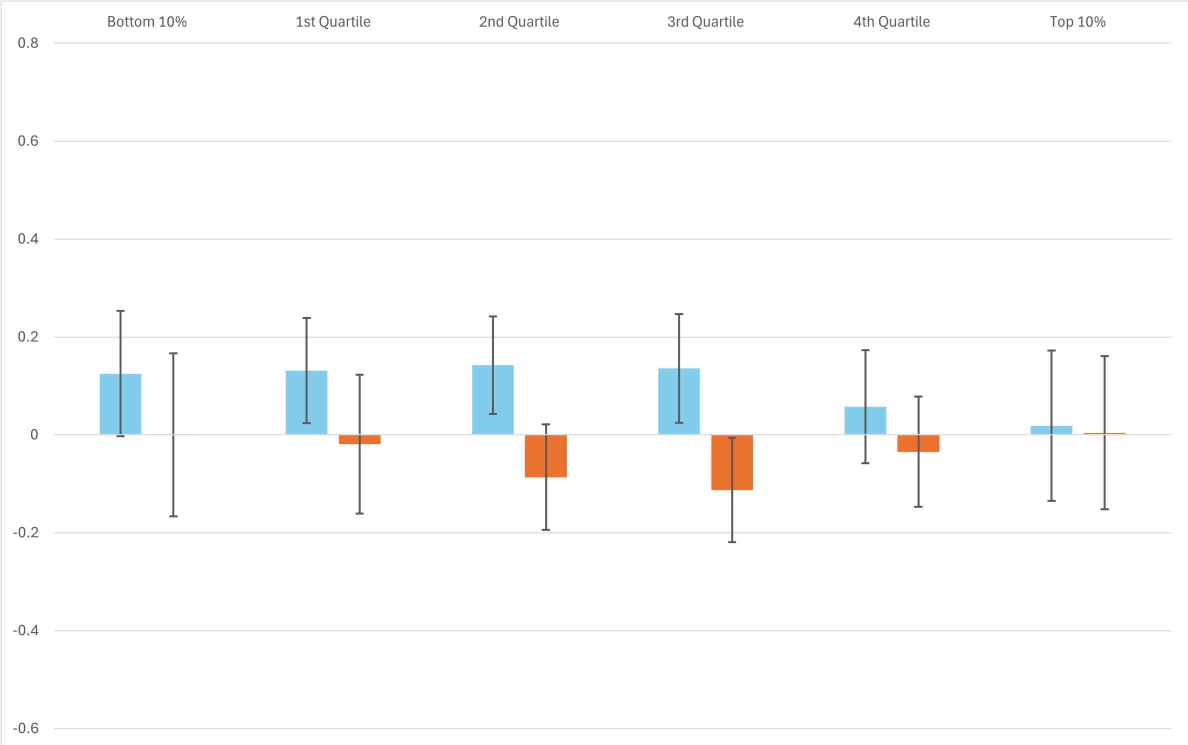


(j) Average Number of Citations

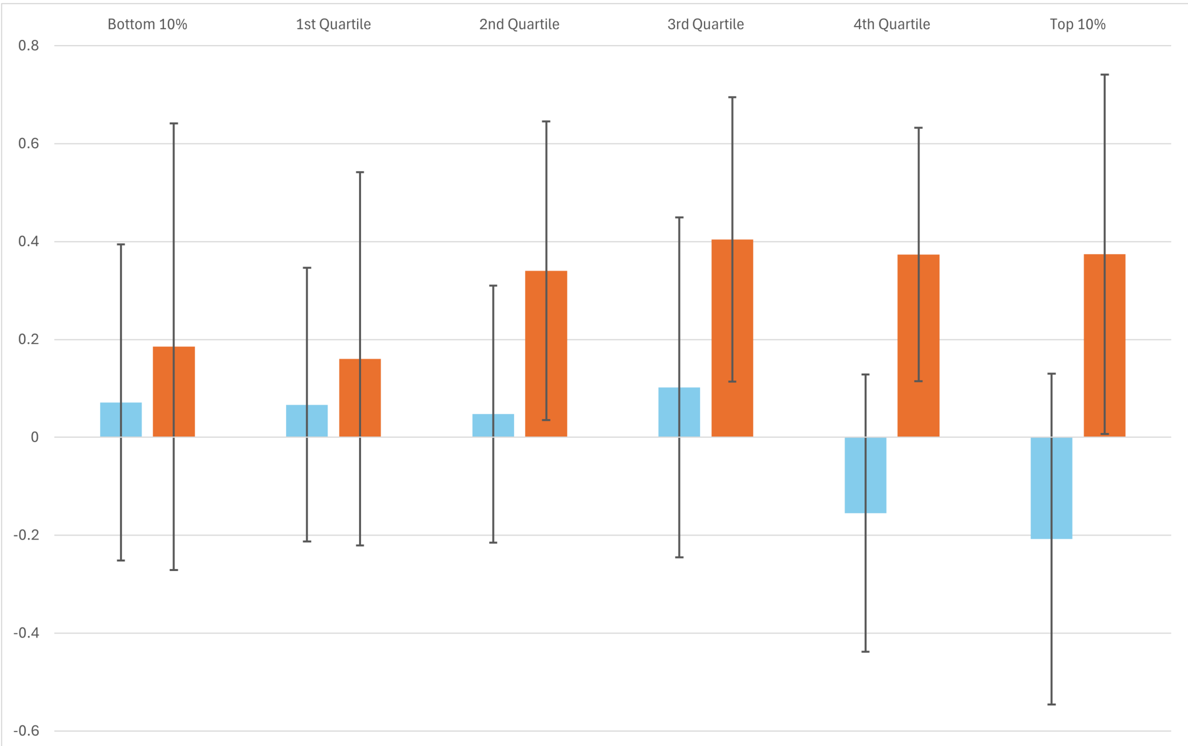
Figure 3: Pre-trends and Effect Dynamics - Women

Notes: Same specification as for Fig. 2. Sample restricted to women. See sections 3.3.1 and 3.3.2 for further details.

Figure 4: Citation Distributions - Men vs. Women



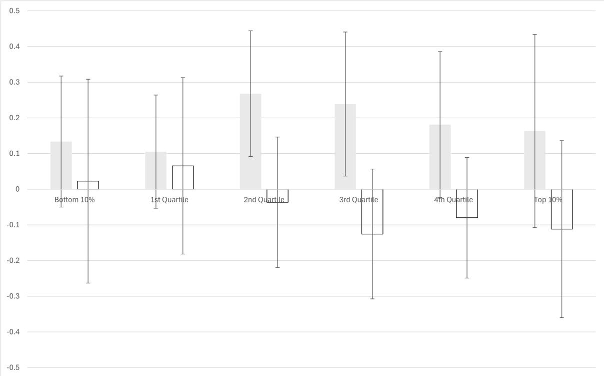
(a) Men



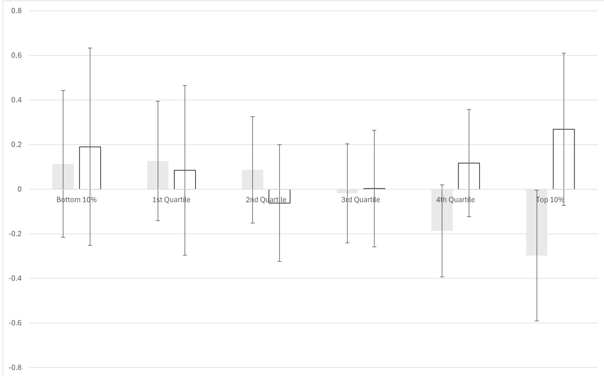
(b) Women

Notes: The plots depict the coefficient estimates and corresponding 95% confidence intervals of the $post'02 * Treatment$ (blue bars) and $Tenure * Treatment$ (orange bars) interactions from separate regressions with citation quantile frequency variables as dependent variables. To generate these dependent variables, percentiles of citations 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. In sub-figure A, the sample is restricted to men and in sub-figure B to women. See section 4.1.1 for further details.

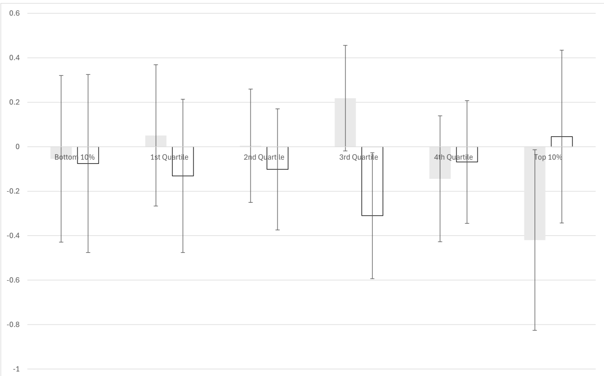
Figure 5: Citation distributions by productivity quantiles - Men



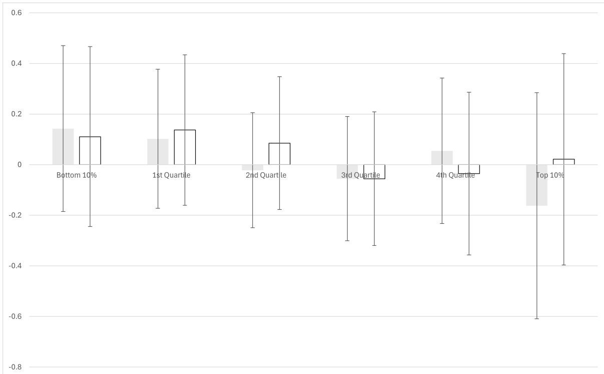
(a) Top decile



(b) 9th decile



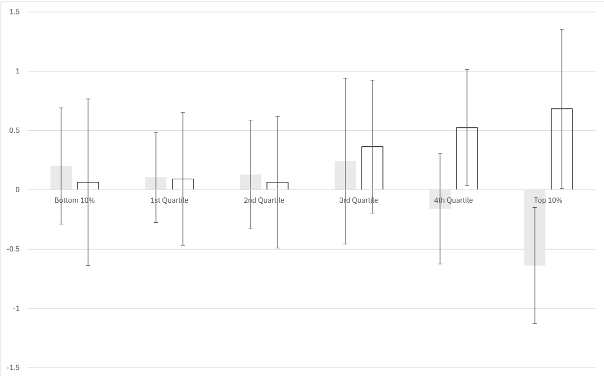
(c) 8th decile



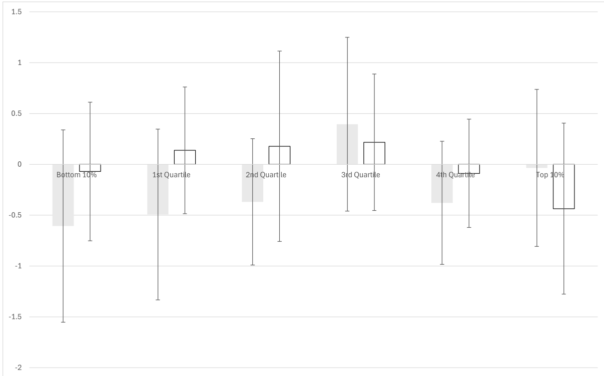
(d) Below Median

Notes: The plots 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 (for the bottom 10%, 1st/2nd/3rd/4th quartile and top 10% citation quantiles). Samples are restricted to the top three productivity deciles (sub-figures A-C) and below median productive men (sub-figure D) respectively. 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 4.1.2 for further details.

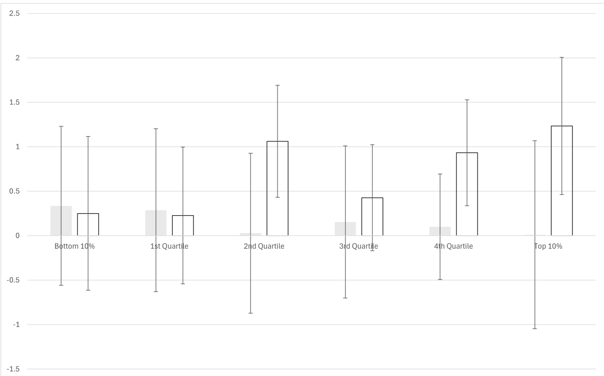
Figure 6: Citation distributions by productivity quantiles - Women



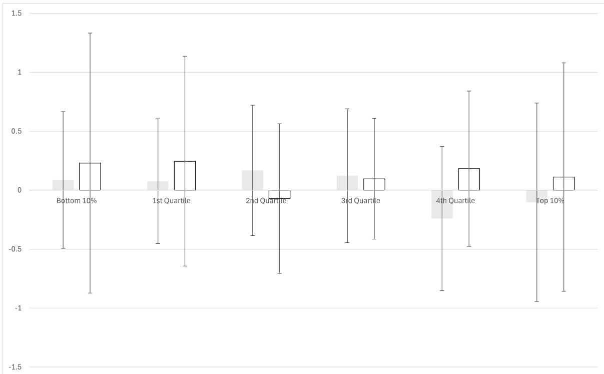
(a) Top decile



(b) 9th decile



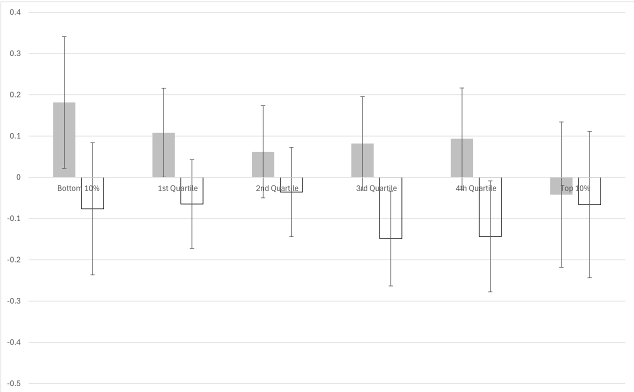
(c) 8th decile



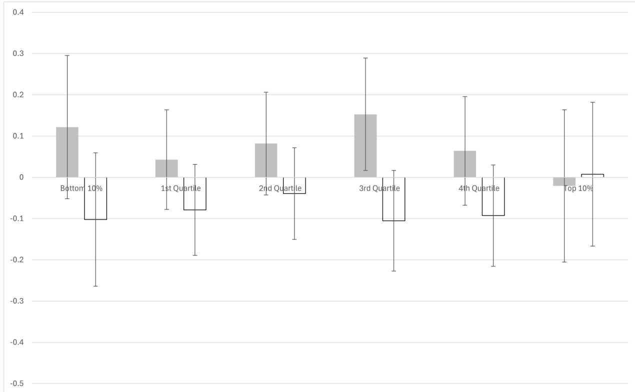
(d) Below Median

Notes: Same specification as for Fig. 5. Sample restricted to women. See section 4.1.2 for further details.

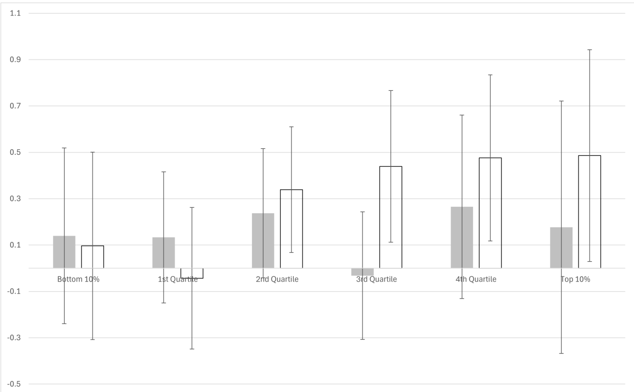
Figure 7: Cosine Similarity Bins



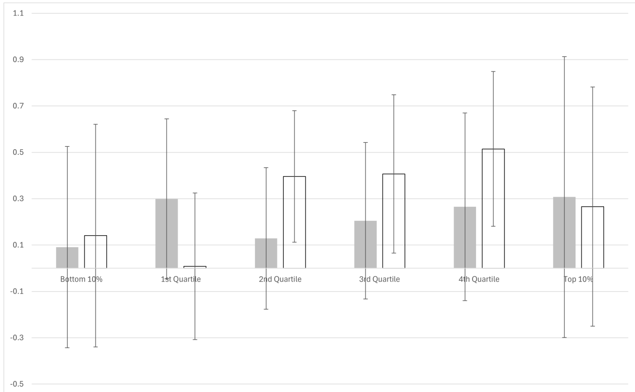
(a) Backward Cosine Similarity - Men



(b) Forward Cosine Similarity - Men



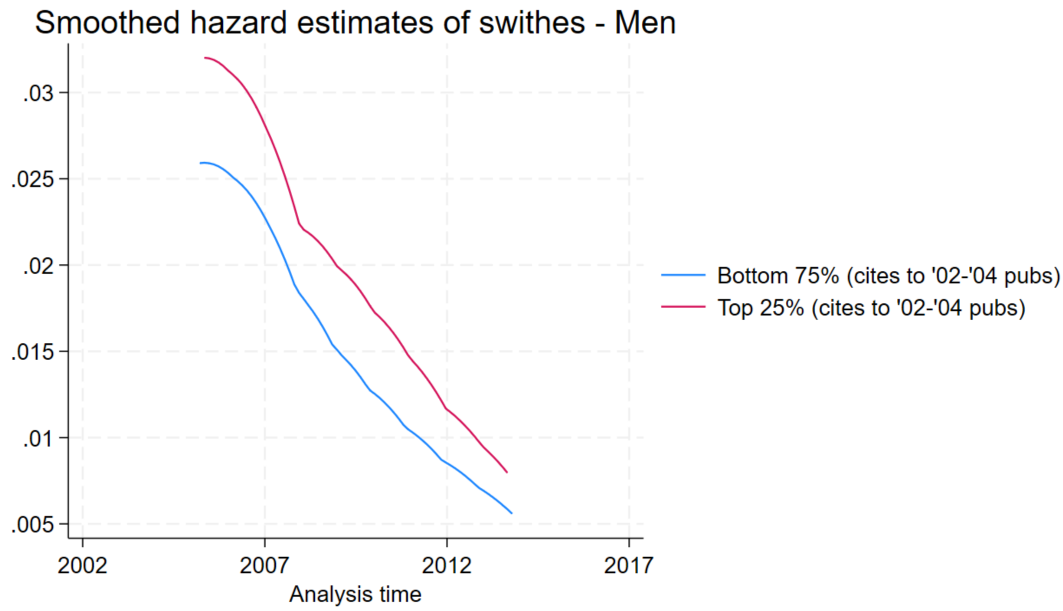
(c) Backward Cosine Similarity - Women



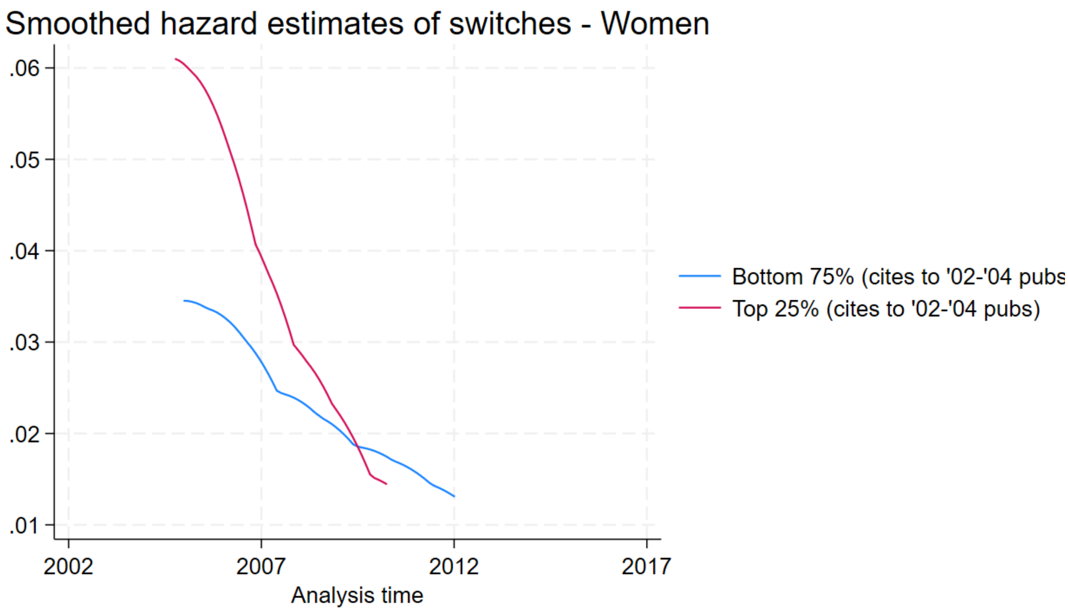
(d) Forward Cosine Similarity - Women

Notes: The plots 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-figures A and C) and forward cosine similarity quantile frequency variables (sub-figures B and D) as dependent variable. The sample is restricted to men in sub-figures A and B, and to women in sub-figures C and D. All other specifications as in the baseline regression of the effort effect. See section 7 for further details.

Figure 8: Switching rates to performance pay of top 25% and bottom 75% productive academics



(a) Men



(b) Women

Notes: The figure 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 2004. Quartiles are determined on the basis of the averages of the citation-weighted 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 4.2.1 for further details.

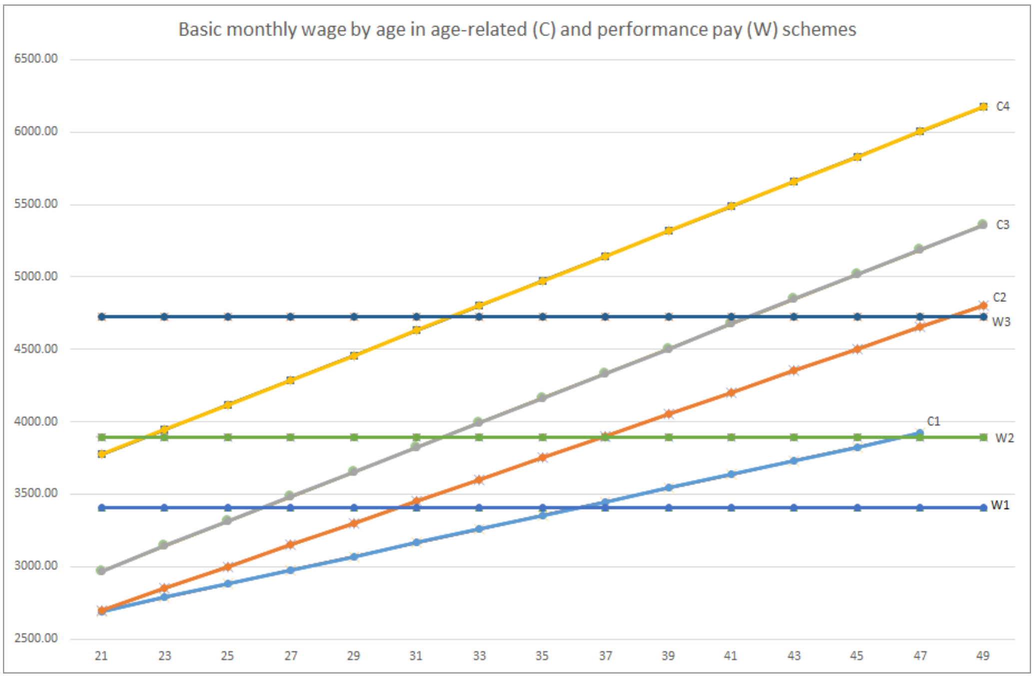
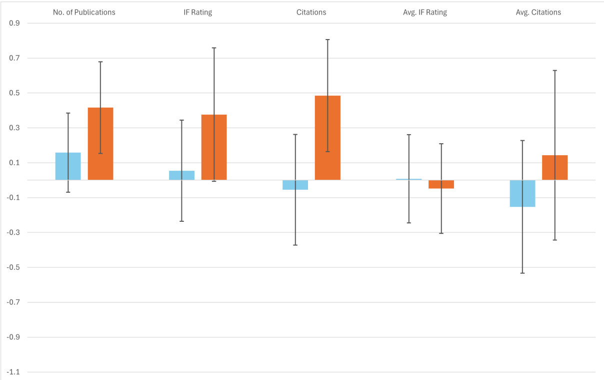


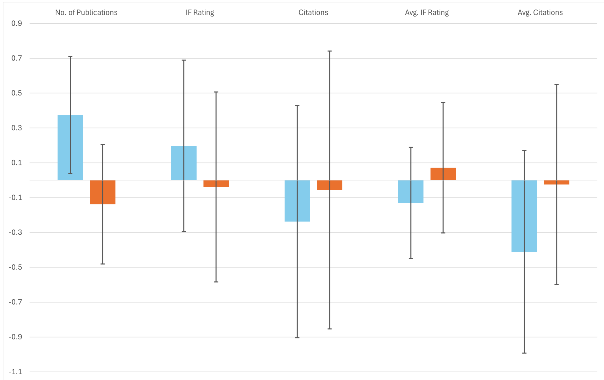
Figure 9: Basic monthly wages by age

Notes: The figure 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).

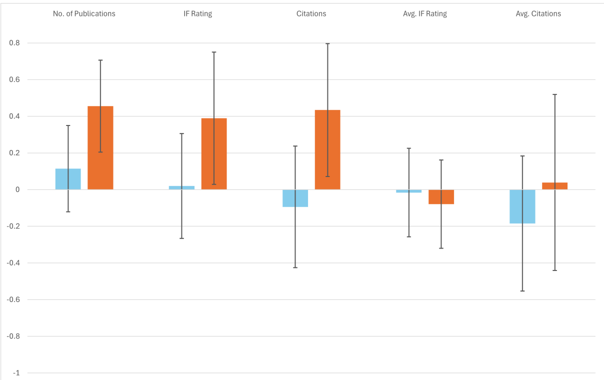
Figure 10: Effort Response in Above/Below Median Average Team Size Fields - Women



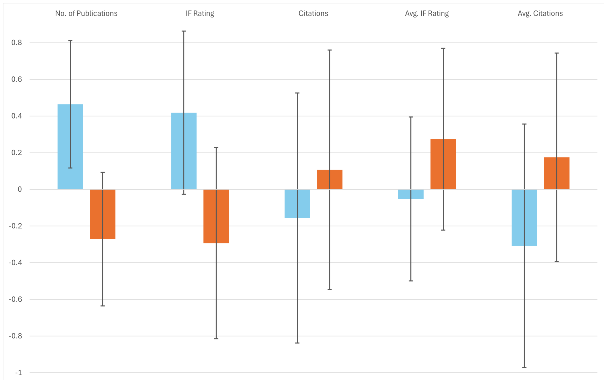
(a) Above Median Average Team Size - Women



(b) Below Median Average Team Size - Women



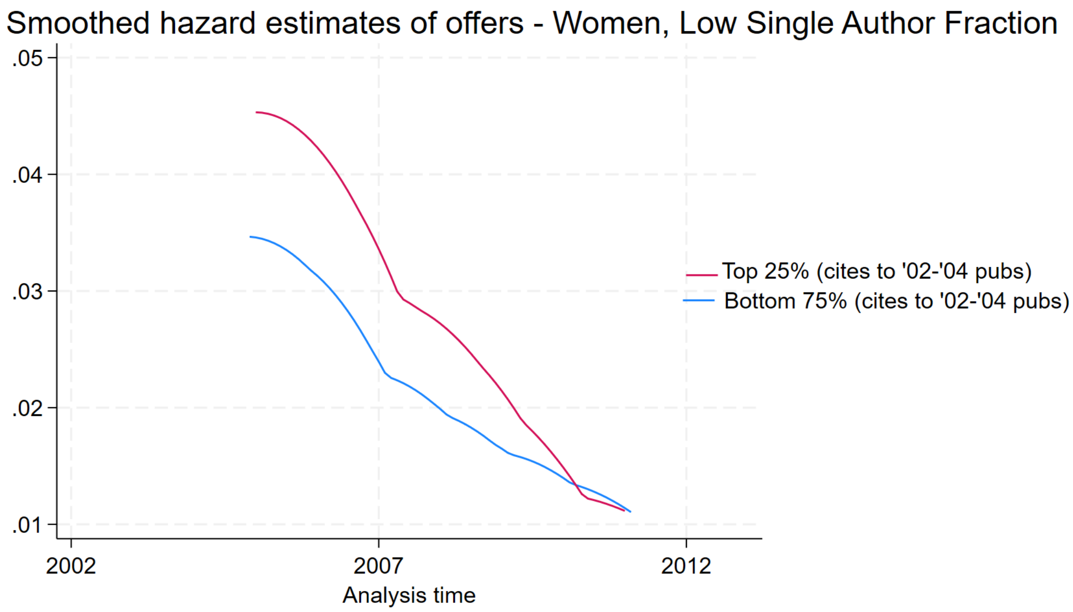
(c) Below Median Single-Author Fraction - Women



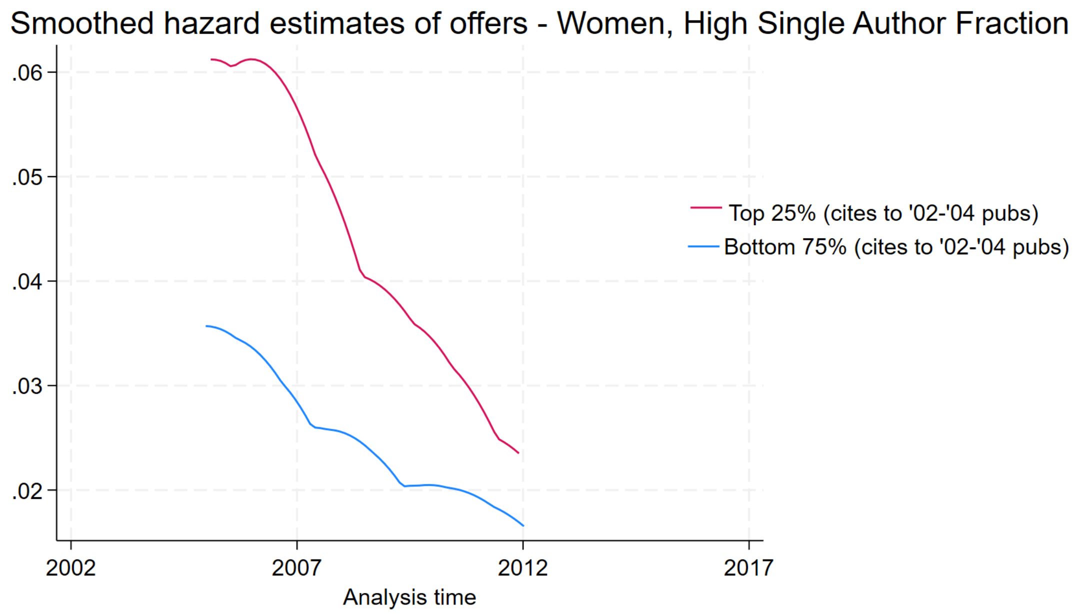
(d) Above Median Single-Author Fraction - Women

Notes: The plots depict the coefficient estimates and corresponding 95% confidence intervals of the $post'02 * Treatment$ (blue bars) and $Tenure * Treatment$ (orange 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 all sub-figures, the sample is restricted to women. The sample is further restricted to female academics in above median (A) vs. below median (B) average team size fields, and below median (C) vs. above median (D) single-author fraction fields. All other specifications as in the baseline regression of the effort effect. See section 5 for further details.

Figure 11: Switching rates to performance pay of top 25% and bottom 75% productive academics - Below/Above median single-author fraction fields



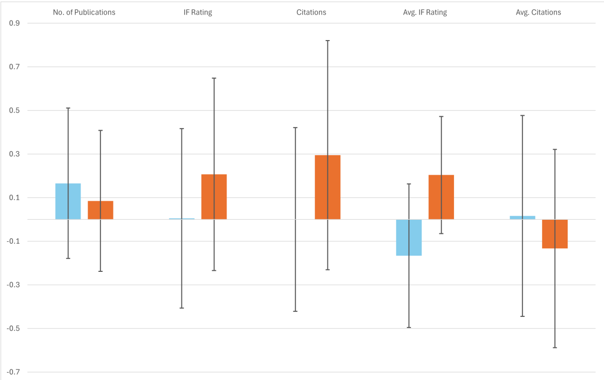
(a) Switching rates of women in below median single-author fraction fields



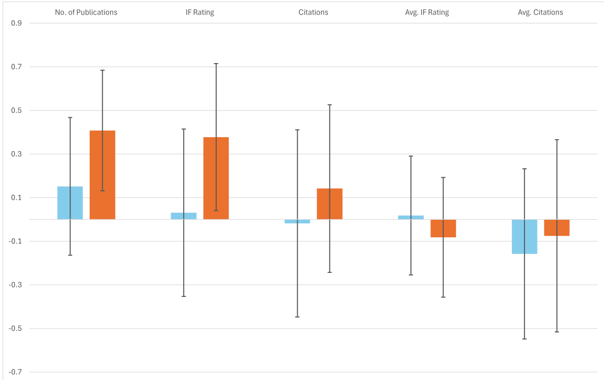
(b) Switching rates of women in above median single-author fraction fields

Notes: The figure 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 2004. Quartiles are determined on the basis of the averages of the citation-weighted 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 female academics in fields with below median (panel a) and above median (panel b) single-author fraction. The sample is restricted to academics who held a tenured affiliation at a public university before 2005 throughout. See section 4.2.1 for further details.

Figure 12: Effort Response in Above/Below Median Female Representation Fields



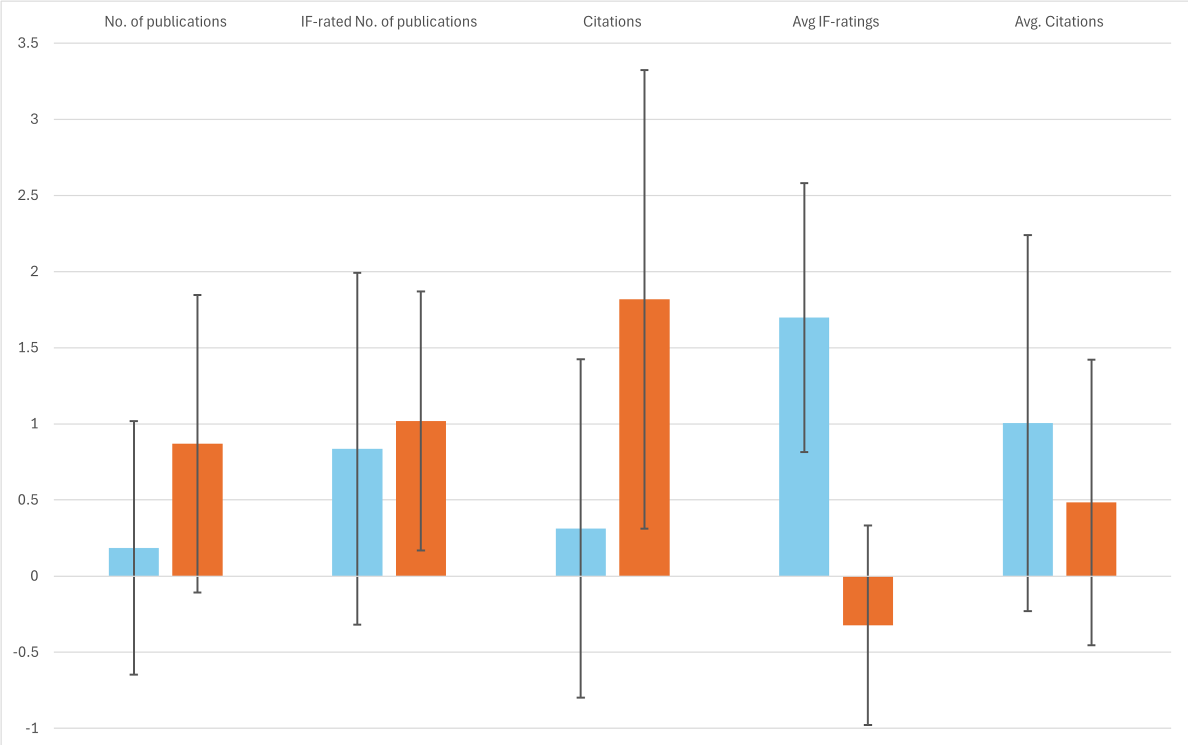
(a) Above Median Female Representation



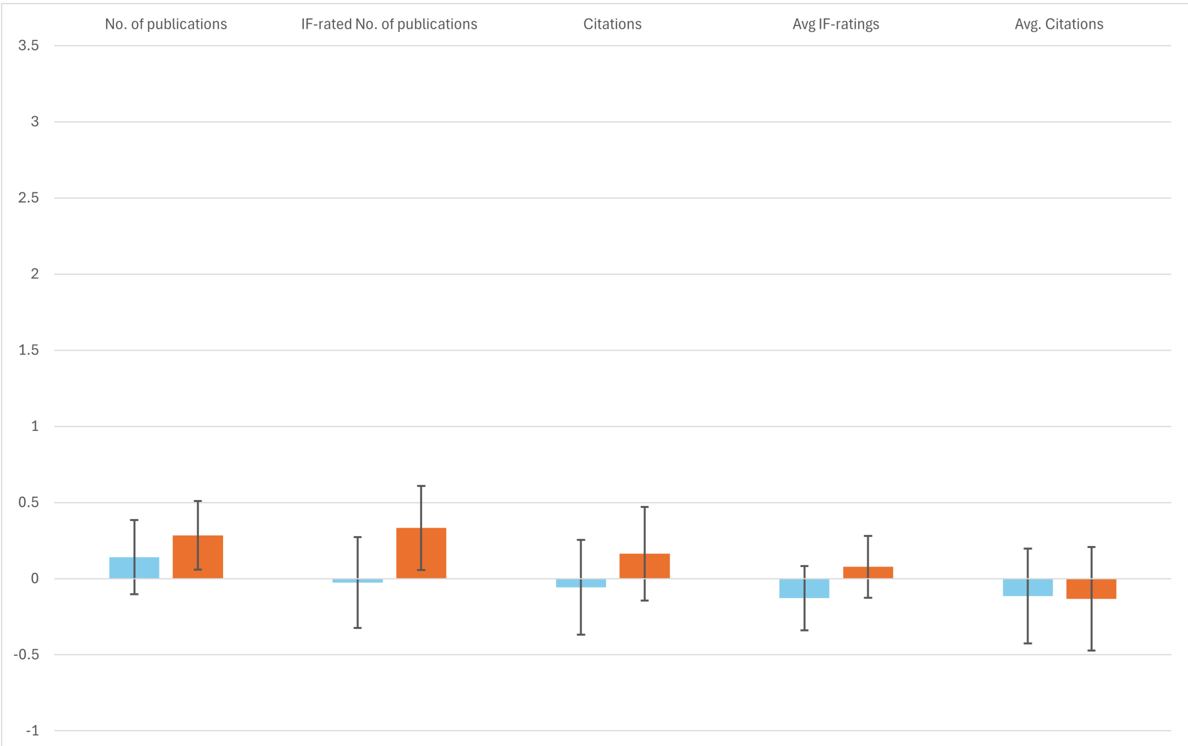
(b) Below Median Female Representation

Notes: The plots depict the coefficient estimates and corresponding 95% confidence intervals of the $post'02 * Treatment$ (blue bars) and $Tenure * Treatment$ (orange 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 all sub-figures, the sample is restricted to women. The sample is further restricted to women in fields with above median proportion of female professors during the sample period in the plot on the left (A), and to fields with below median proportion of female professors in the plot on the right (B). All other specifications as in the baseline regression of the effort effect. See section 5 for further details.

Figure 13: Effort Response of Women Who are Married and Not Married



(a) Married



(b) Not Married

Notes: The plots depict the coefficient estimates and corresponding 95% confidence intervals of the $post'02 * Treatment$ (blue bars) and $Tenure * Treatment$ (orange 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 women with a hyphenated last-name (as a proxy for being married) and in sub-figure B to women who do not have a hyphenated last name (as a proxy for not being married). See section 5 for further details.

Table 1: Summary Statistics

Research output variables	N	Mean	SD	P10	P25	Median	P75	P90	P99
Panel A: Women									
Number of publications	20276	1.59	5	0	0	0	1	4	24
IF-rated publications	20276	5.21	21.8	0	0	0	0.57	11.75	91.36
Citations	20276	53.84	222.68	0	0	0	3	116	1010
Average IF-rating	6204	2.58	2.95	0	0.64	1.85	3.58	5.6	13.33
Average citations	6204	31.41	54.81	0	5	17.46	37	69.33	250
Panel B: Men									
Number of publications	82741	3.33	7.66	0	0	0	3	9	36
IF-rated publications	82741	11.31	33.77	0	0	0	6.85	31.84	158.02
Citations	82741	116.89	358.99	0	0	0	71	323	1624
Average IF-rating	40775	2.7	2.8	0.19	0.81	2.09	3.74	5.66	12.6
Average citations	40775	32.63	62.34	1	6.67	19.25	39	70	225.9

Notes: The table shows summary statistics for the samples used for the baseline effort effect estimation. The unit of observation is academic i. The samples are 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 A shows summary statistics for women, Panel B for men.

Table 2: Descriptive Statistics - Productivity Quantiles

Research output variables	Below Median	8th	9th	Top
Panel A: Women				
Number of publications	0.27	2.63	3.3	8.97
IF-rated publications	0.48	9.55	11.87	33.25
Citations	6.48	111.51	125.79	400.96
Bottom 10% citations	0.51	0.92	1	2.52
Bottom 25% citations	0.6	1.05	1.29	2.91
Top 25% citations	0.37	0.92	1.18	3.08
Top 10% citations	0.12	0.48	0.47	1.18
Panel B: Men				
Number of publications	0.55	3.94	5.78	13.16
IF-rated publications	0.84	12.22	20.12	48.01
Citations	12.86	175.76	246.15	532.78
Bottom 10% citations	0.48	0.87	1.27	2.92
Bottom 25% citations	0.65	1.09	1.69	3.75
Top 25% citations	0.41	1.39	1.7	3.7
Top 10% citations	0.12	0.58	0.71	1.51

Notes: The table shows the mean of research output variables in the three pre-implementation years (1999-2001) for four productivity quantiles (the top 3 deciles and below median productivity academics) of women (panel A) and men (panel B). Productivity quantiles are based on the averages of the impact factor-rated number of publications in 1999-2001. The output variables in the last four rows are citation quantile frequency variables (for the bottom 10%, 1st/4th quartile and top 10%). See sections 4.1.1 and 4.1.2 for details.

Table 3: Validity Checks - Men

	# Publications	IF-rated publications	Citations	Average IF-rating	Average citati
Panel A: Placebo					
Post'02 * Treatment	0.002	-0.043	-0.087	0.030	-0.073
	(0.058)	(0.081)	(0.087)	(0.051)	(0.108)
Tenure * Treatment	-0.001	0.028	0.174*	0.042	0.160
	(0.058)	(0.074)	(0.099)	(0.048)	(0.098)
Number of Observations	43440	39061	40780	25415	25743
Number of Individuals	2419	2175	2270	2057	2145
Log Likelihood	-74451.328	-180176.372	-2453351.490	-36800.197	-389783.433
Chi-squared	1456.177	1451.976	577.705	463.123	310.285
Panel B: Without Switchers					
Post'02 * Treatment	0.210***	0.192***	0.175**	-0.112***	-0.112
	(0.043)	(0.060)	(0.075)	(0.043)	(0.074)
Tenure * Treatment	-0.047	-0.026	-0.027	0.025	0.004
	(0.048)	(0.065)	(0.082)	(0.046)	(0.084)
Number of Observations	58629	52869	55429	34708	35201
Number of Individuals	3264	2943	3085	2802	2921
Log Likelihood	-100234.545	-251033.163	-3379274.804	-52350.628	-551548.818
Chi-squared	2174.505	2103.629	798.506	479.538	379.364
Panel C: Publication Variables Weighted by Number of Authors					
Post'02 * Treatment	0.175***	0.115*	0.133*	-0.120***	-0.088
	(0.039)	(0.059)	(0.080)	(0.040)	(0.076)
Tenure * Treatment	-0.018	0.093	0.082	0.036	-0.001
	(0.046)	(0.069)	(0.087)	(0.044)	(0.086)
Number of Observations	67135	60457	63449	39681	40251
Number of Individuals	3737	3365	3531	3207	3350
Log Likelihood	-54734.528	-93327.239	-915211.303	-58899.421	-582403.164
Chi-squared	1636.601	1514.298	631.794	384.436	546.716

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 shows the result of the baseline regression with a placebo treatment and control group, where the placebo treatment group is comprised of academics appointed to their first tenured affiliation in 2003-2004, and the placebo control in 2001-2002. . 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. Sample restricted to men. See sections 3.3.2-3.3.4 for further details.

Table 4: Validity Checks - Women

Panel A: Placebo					
	# Publications	IF-rated publications	Citations	Average IF-rating	Average citatio
Post'02 * Treatment	-0.084	-0.038	-0.318	0.131	-0.349*
	(0.150)	(0.207)	(0.237)	(0.138)	(0.203)
Tenure * Treatment	-0.104	-0.264	0.081	-0.023	0.413
	(0.158)	(0.216)	(0.267)	(0.134)	(0.299)
Number of Observations	7748	6363	6903	3467	3579
Number of Individuals	431	354	384	326	357
Log Likelihood	-9763.515	-22515.364	-285086.670	-4972.619	-48903.950
Chi-squared	356.859	426.793	243.619	80.482	112.189
Panel B: Without Switchers					
Post'02 * Treatment	0.223	0.070	0.072	-0.018	0.026
	(0.136)	(0.167)	(0.169)	(0.120)	(0.166)
Tenure * Treatment	0.269**	0.291*	0.110	-0.006	-0.237
	(0.120)	(0.150)	(0.172)	(0.110)	(0.190)
Number of Observations	11637	9369	10161	5037	5154
Number of Individuals	647	521	565	487	520
Log Likelihood	-14351.983	-34084.702	-443519.974	-7595.228	-79188.709
Chi-squared	453.146	493.648	243.028	56.274	72.377
Panel C: Publication Variables Weighted by Number of Authors					
Post'02 * Treatment	0.222**	0.056	-0.084	-0.026	-0.195
	(0.099)	(0.134)	(0.152)	(0.113)	(0.167)
Tenure * Treatment	0.206*	0.317*	0.346**	-0.026	0.032
	(0.112)	(0.170)	(0.167)	(0.112)	(0.200)
Number of Observations	13544	10971	11835	5854	6000
Number of Individuals	753	610	658	566	609
Log Likelihood	-8307.124	-12696.128	-120700.641	-8617.739	-81214.400
Chi-squared	354.746	433.594	297.660	53.979	132.005

Notes: Same specifications as in Table 3. Sample restricted to women. See sections 3.3.2-3.3.4 for further details.

Table 5: Selection Analysis -Men

	Panel A: Treatment versus Control				Panel B: Placebo	
	1a	1b	2a	2b	3a	3b
Treatment	0.406	0.049	0.352	-0.293	-1.366	-2.546***
	(0.601)	(0.637)	(0.659)	(0.695)	(0.882)	(0.958)
Age	-0.117***	-0.269***	-0.121***	-0.283***	-0.155***	-0.483***
	(0.011)	(0.023)	(0.012)	(0.024)	(0.015)	(0.059)
Avg Productivity	-0.000	-0.000*	-0.001	-0.003**	0.000	-0.000
	(0.000)	(0.000)	(0.001)	(0.001)	(0.001)	(0.001)
Age * Treatment	-0.023*	-0.008	-0.021	-0.001	0.024	0.052***
	(0.013)	(0.014)	(0.014)	(0.015)	(0.018)	(0.020)
Avg Productivity * Treatment	0.000***	0.001***	0.001	0.003**	-0.000	-0.000
	(0.000)	(0.000)	(0.001)	(0.001)	(0.001)	(0.001)
Avg Productivity * Age			0.000	0.000**	0.000	0.000
			(0.000)	(0.000)	(0.000)	(0.000)
Avg Productivity * Age * Treatment			-0.00002	-0.00005*	0.000	0.000
			(0.000)	(0.000)	(0.000)	(0.000)
Age at Tenure		0.176***		0.181***		0.336***
		(0.020)		(0.021)		(0.054)
Constant	2.167***	1.216**	2.342***	1.649***	3.169***	2.210***
	(0.569)	(0.592)	(0.606)	(0.623)	(0.721)	(0.755)
Number of Observations	40011	40011	40011	40011	26583	26583
Number of Subjects	7237	7237	7237	7237	3595	3595
Number of Switches	1163	1163	1163	1163	579	579
Log Likelihood	-3591.743	-3529.561	-3591.219	-3527.124	-1703.172	-1666.889
Chi-squared	623.592	587.870	620.301	583.239	287.917	251.425
Rho	1.353	1.632	1.354	1.644	1.425	2.702

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 appointed to their first tenured position before 2005 and 0 for those appointed to their first tenured position afterwards. Panel B reports the results for a placebo experiment where the placebo-treatment group comprises academics appointed to their first tenured position before 2002, while academics appointed to their first tenured affiliation in 2002-2005 act as placebo-control group. "Avg Productivity" is calculated as three year pre-implementation averages (2002-2004) of the citation-weighted 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 also control for synthetic age at tenure. Standard errors are robust and clustered by individual academic. Sample restricted to men. See section 4.2.2 for further details.

Table 6: Selection Analysis -Women

	Panel A: Treatment versus Control				Panel B: Placebo	
	1a	1b	2a	2b	3a	3b
Treatment	3.176**	1.941	1.904	0.804	-3.574	-4.992*
	(1.354)	(1.487)	(1.412)	(1.564)	(2.370)	(2.626)
Age	-0.064***	-0.328***	-0.086***	-0.341***	-0.149***	-0.500***
	(0.024)	(0.053)	(0.024)	(0.053)	(0.027)	(0.116)
Avg Productivity	-0.000	-0.000	-0.009***	-0.007***	-0.001	-0.001
	(0.000)	(0.000)	(0.003)	(0.003)	(0.004)	(0.004)
Age * Treatment	-0.080***	-0.043	-0.052*	-0.018	0.065	0.100*
	(0.029)	(0.032)	(0.031)	(0.034)	(0.049)	(0.055)
Avg Productivity * Treatment	-0.000	-0.000	0.011***	0.010***	0.007	0.007
	(0.000)	(0.000)	(0.004)	(0.003)	(0.006)	(0.006)
Avg Productivity * Age			0.000***	0.000***	0.000	0.000
			(0.000)	(0.000)	(0.000)	(0.000)
Avg Productivity * Age * Treatment			-0.0003***	-0.0002***	-0.000	-0.000
			(0.000)	(0.000)	(0.000)	(0.000)
Age at Tenure		0.288***		0.282***		0.367***
		(0.050)		(0.050)		(0.115)
Constant	0.282	-0.490	1.307	0.407	3.875***	2.928**
	(1.233)	(1.282)	(1.236)	(1.299)	(1.377)	(1.402)
Number of Observations	6362	6362	6362	6362	3568	3568
Number of Subjects	1284	1284	1284	1284	484	484
Number of Switches	239	239	239	239	100	100
Log Likelihood	-711.737	-688.909	-708.711	-686.554	-270.967	-263.531
Chi-squared	893.119	1228.848	895.309	1045.166	773.176	868.503
Rho	1.313	1.825	1.312	1.814	1.371	2.601

Notes: Same specifications as in Table 5. Sample restricted to women. See section 4.2.2 for further details.

7 Appendix

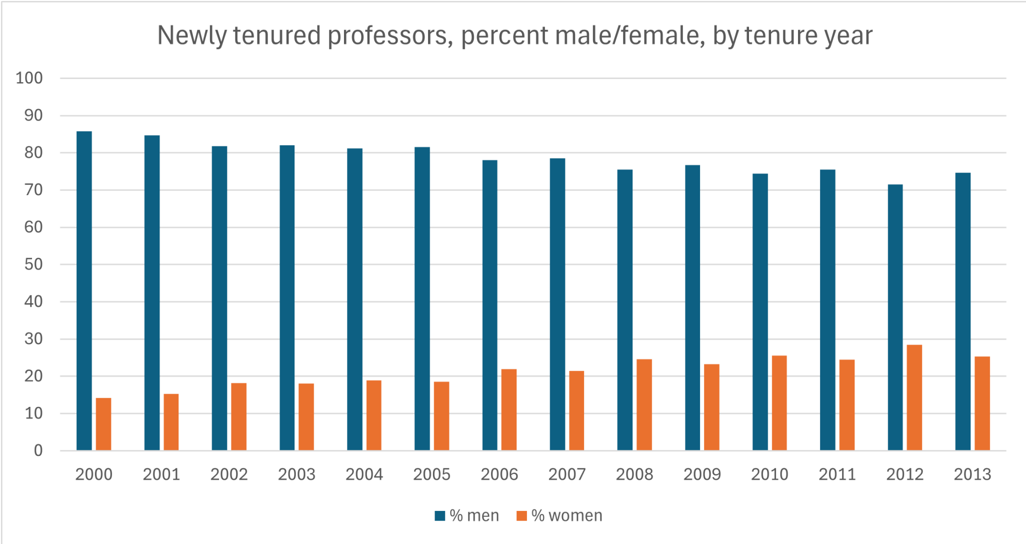


Figure 14: Percent female and male professors appointed to first tenured position

Notes: The figure shows the percentage of male and female professors appointed to their first tenured position by year. Only first tenured affiliations at a public university are taken into account, and academics who hold a foreign affiliation before their first tenured affiliation in Germany are excluded.

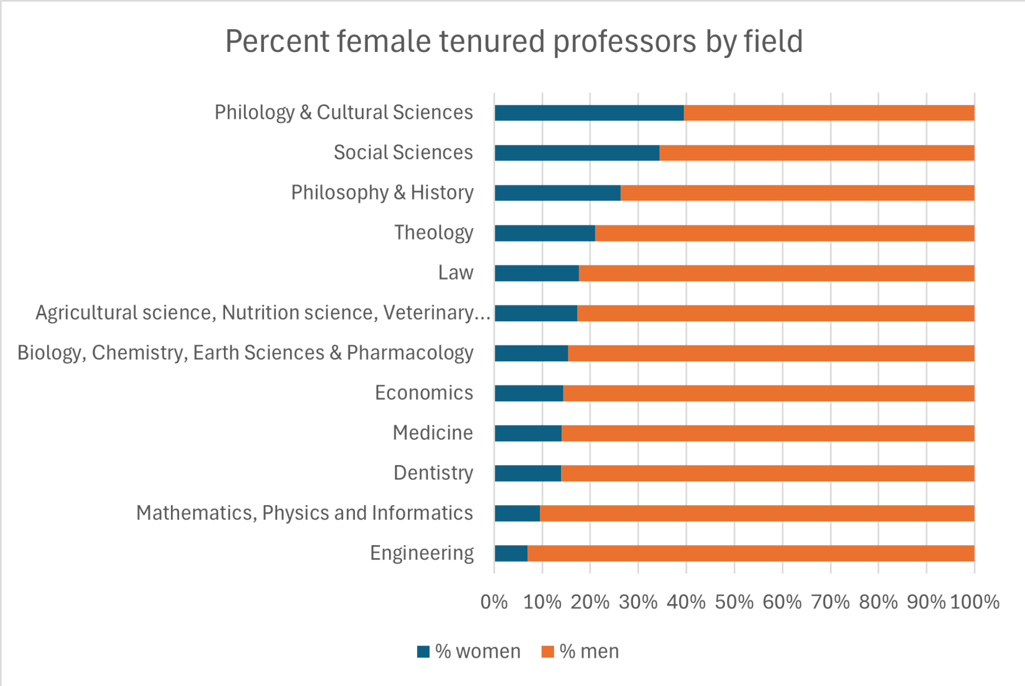
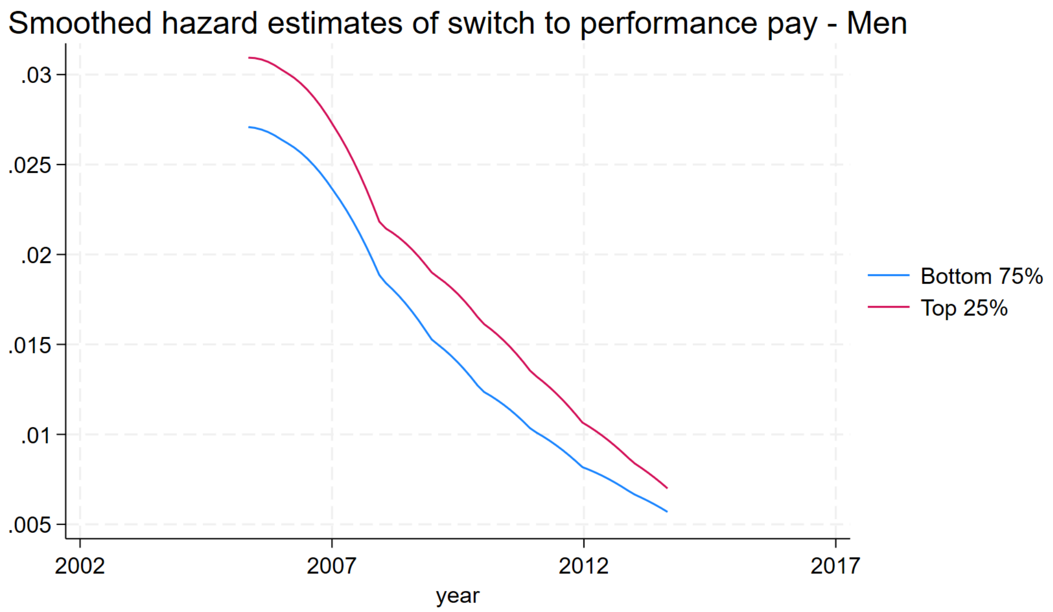


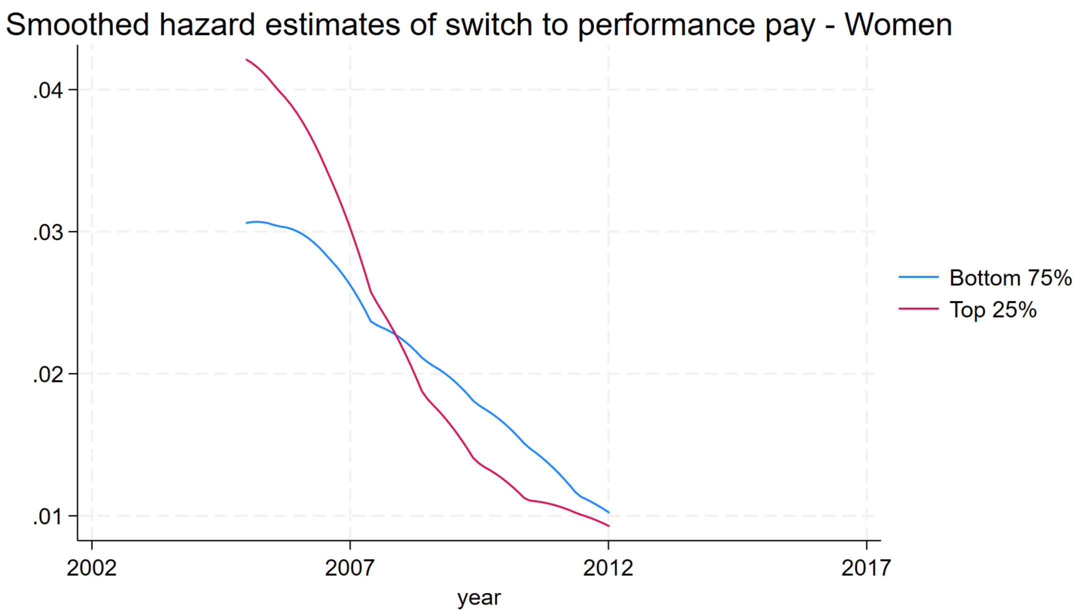
Figure 15: Percent female and male professors across fields

Notes: The figure shows the percentage of male and female professors at German public universities in 2012, separately by 12 broad academic fields.

Figure 16: Switching rates to performance pay of top 25% and bottom 75% productive academics



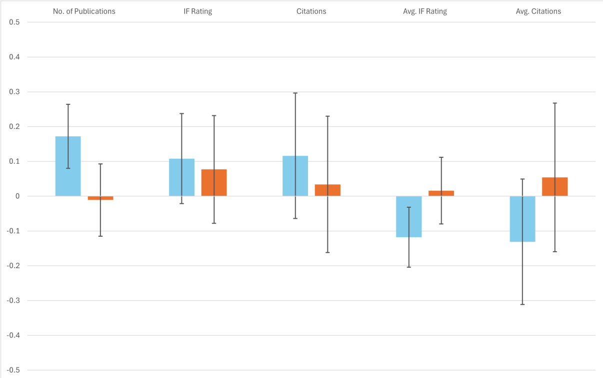
(a) Men



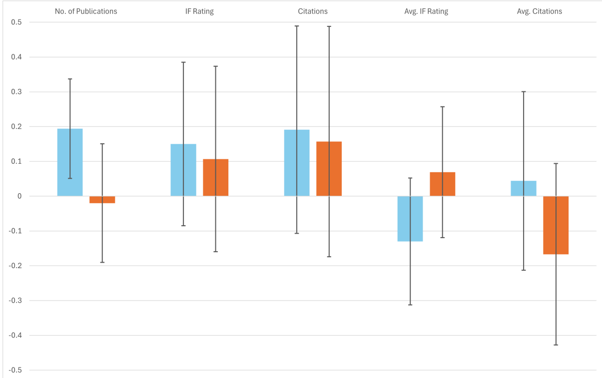
(b) Women

Notes: The figure 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 2004. 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 4.2.1 for further details.

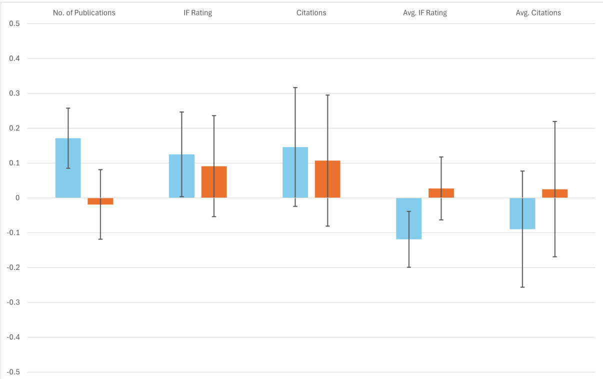
Figure 17: Effort Response in Above/Below Median Average Team Size Fields - Men



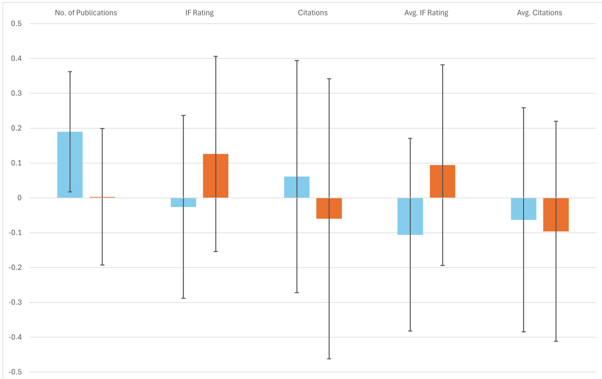
(a) Above Median Average Team Size - Men



(b) Below Median Average Team Size - Men



(c) Below Median Single-Author Fraction - Men



(d) Above Median Single-Author Fraction - Men

Notes: The plots depict the coefficient estimates and corresponding 95% confidence intervals of the $post'02 * Treatment$ (blue bars) and $Tenure * Treatment$ (orange 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 all sub-figures, the sample is restricted to men. The sample is further restricted to female academics in above median (A) vs. below median (B) average team size fields, and below median (C) vs. above median (D) single-author fraction fields. All other specifications as in the baseline regression of the effort effect. See section 5 for further details.