Truly Legendary Freedom: Funding, Incentives, and the Productivity of Scientists

Matthias Wilhelm*

July 4, 2019

How do elite scientists react to an increase in grant size and duration? We compare recipients of the most important German research prize before and after an increase (i) in its total grant size of €1m and (ii) in its grant duration by two years using a difference-in-differences design. Scientists react to these changes by reducing their overall number of publications while increasing the number of papers in top ranked journals, but do not change the novelty of their research. Additional analysis suggests that it is the interaction of larger grant size and longer grant duration that drives these productivity effects.

Keywords: Science Funding; Scientific Productivity
JEL-Codes: O38; O32

*Department of Economics, Ludwig-Maximilians-University Munich, Akademiestr. 1, 80799 Munich, matthias.wilhelm@econ.lmu.de. For helpful comments and suggestions I thank Jeff Furman, Anna Gumpert, Benjamin Häusinger, Cathrin Mohr, Frank Müller-Langer, Markus Nagler, Monika Schnitzer, Andreas Steinmayr, Scott Stern, Reinhilde Veugelers, and Martin Watzinger as well as seminar and conference participants at LMU Munich, the Max-Planck-Institute for Innovation and Competition, the 8th ZEW/MaCCI Conference on the Economics of Innovation and Patenting, and the 13th Workshop on the Organisation, Economics and Policy of Scientific Research. For sharing her experiences with the Leibniz Prize, I thank Antje Boetius. The title is adapted from the book "Von märchenhafter Freiheit" by Marco Finetti. I gratefully acknowledge financial support from the Elite Network of Bavaria through the doctoral program Evidence-Based Economics and from the German Academic Exchange Service. Parts of this paper were written while visiting MIT Sloan. I thank the department and especially Antoinette Schoar for their hospitality.

1. Introduction

Scientific breakthroughs have spurred productivity growth for the past decades and many great inventions have roots in the labs of universities and research institutes (Mokyr, 2016; Bush, 1945; The Economist, 2011). Most, if not all, of these breakthroughs would not have been possible without funding to pay for the research (Stephan, 2010). Funding comes in a variety of shapes and sizes. It ranges from grants for individual projects to grants for entire research institutes and from a few thousand to several million or even billions of dollars (Stephan, 2012).

The design of funding is a key lever for governments to incentivize scientific research. Potential design choices include the size of individual grants, the duration over which grants can be spent, how a grant can be spent, and how the allocation of a grant is decided. These design choices are the topic of an extensive debate, as critics contend that the current system of academic funding discourages risky research, limits academic freedom, and pushes scientists to publish ever larger numbers of papers of little scientific value (Stephan 2012, Nicholson and Ioannidis 2012, Sarewitz 2016).[1]

Unfortunately, we know little about how different funding amounts and structures affect scientific output. There is little variation within funding programs and comparisons across programs are fraught with selection issues. Scientists self-select and are selected into different programs based on past and expected future productivity. Hence, any comparison across schemes usually suffers from endogeneity issues.

This paper studies a reform of Germany's most important research prize to assess how elite scientists react to an increase in the total grant amount and duration of a grant that comes without any strings attached. It compares recipients of the *same* prize with *different* grant amounts and durations in a difference-in-differences framework, circumventing selection issues. We find that scientists after the reform reduce their overall number of publications, but increase their number of publications in top ranked journals. Additional analysis suggests that this effect is due to the combination of both a larger grant amount and a longer grant duration.

[1]Examples in the popular press are *Dr. No Money: The Broken Science Funding System* (Scientific American, 01 May 2011) or *Grant System Leads Cancer Researchers to Play it Safe* (The New York Times, 27 June 2009).

The reform of the Gottfried Wilhelm Leibniz Prize is the ideal testing ground to assess how the amount and structure of funding impacts scientific productivity. It is Germany's most prestigious research prize and recipients cannot apply, but must be nominated. It bestows both honor for past achievement and comes with a research grant for the following years. In the words of the German Research Foundation's (DFG) former president Hubert Markl, the DFG wants to provide the recipients with *truly legendary freedom* to conduct their research (Finetti, 2010). Hence, the grant can be spent at the full discretion of the recipient.[2] In 2007, there was an increase in this *truly legendary freedom*. The total grant amount was increased from €1.55m to €2.50m, and the period over which these funds could be spent was increased from five to seven years. Since the selection criteria and selection process of Leibniz Prize recipients remained the same, we can use this natural experiment to study how very similar researchers behave under different funding schemes.

We use a difference-in-differences identification strategy and study how the productivity of prize recipients differs before and after the reform, comparing the change in publication output before and after receiving the Leibniz Prize. We measure academic productivity in several ways. First, we count the number of publications per year, irrespective of the outlet (journals, conference proceedings, books). Second, we differentiate journal publications by the rank of the journal. We rank journals by their impact factor, i.e. by the average citations to the papers published in the journal. Third, we complement the analysis by studying how scientists change their research direction over time by studying the text similarity of the abstracts of their research papers. We measure how similar abstracts are to each other within a given year and how similar subsequent research is to the research of a winner conducted five years prior to the prize. Lastly, we build on the approach of Uzzi et al. (2013) and study how novel or conventional their research is, based on which journal combinations are referenced together in a paper.

We find that the post reform prize recipients reduce their overall number of publications at least 65 percent relative to the mean, or by 5.91 publications per year. This

[2] Whitley (2014, p.370) uses the term *protected space* to describe "the period of time in which scientists have discretion over needed resources, including their own efforts, (..:) before having to produce publishable and collectively valued results." The reform of the Leibniz Prize can be interpreted as an increase in protected space.

reduction in the overall quantity of publications is accompanied by an increase in the number of publications in top ranked journals, where the number of publications in the top 1 percent and top 2 percent of journals more than doubles. We find some evidence for a change in the research direction relative to the early stock of publications, but no change in the diversity of research within a given year. Similarly, there is no significant effect on the average novelty or conventionality of the prize winners' research. These results are similar across all scientific fields (engineering, life sciences, natural sciences, social sciences) under study.

A key concern would be time-varying shocks that differentially affect the treatment cohorts relative to the control cohorts. Examples include the introduction of ERC grants.[3] We extensively discuss this issue and use a matched control group of other eminent scientists who did not receive the Leibniz Prize to argue that these concurrent shocks cannot plausibly explain our observed patterns.

We can rule out many prominent alternative explanations for the productivity effects through our within prize comparison. A form of "Matthew effect" (Merton, 1968) where Leibniz Prize recipients find it easier to publish their research in top ranked journals due to their increased prestige should affect Leibniz Prize recipients before and after the reform in the same fashion. The same holds for an increase in the personal threshold of what is deemed "publishable" research by the scientist. Moreover, the Leibniz Prize is just as prestigious before and after the reform.[4]

Additionally, we shed light on the question whether it is the increase in the grant amount, the grant duration, or the combination of the two that matters. We use the fact that the funding of the Leibniz Prize stayed constant in *nominal* terms from 1986 to 2006, whereas money lost 45 percent of its value in *real* terms. Hence, the earliest Leibniz prize recipients prior to 1992 had almost the same grant amount in real terms as the post reform cohorts after 2007.

Three comparisons aim at disentangling the effects of grant amount and grant dura-

[3] The European Research Council (ERC) is a public funding body for scientific research within the EU. It has a budget of €13 billion and funds early and peak career researchers with grants of €1.5m to €3.5m, similar in magnitude to the Leibniz Prize. These grants were introduced in 2007.

[4] As suggestive evidence that there is no discontinuity in the interest of the general public into the Leibniz Prize, Figure B.3 plots Google Search trends related to the Leibniz Prize around the reform. There are no visible differences.

tion. First, to isolate the effect of the grant amount, we conduct a comparison within the control group. We compare the 1986 to 1992 prize recipients to the 2000 to 2006 cohorts. Both groups had five years to conduct their research, but the former received €600,000 more funding in real terms than the latter. Second, to isolate the effect of the grant duration, we compare the 1986 to 1992 recipients to the 2007 to 2010 recipients. Both groups received comparable amounts of funding in real terms, but the latter had two more years to conduct their research than the former. Lastly, we compare the 2007 to 2010 recipients to the 2000 to 2006 prize winners, where there was an increase in the grant duration and the increase in grant amount was strongest.

In the last comparison, the pattern is very similar to the baseline estimate whereas in the other two comparisons the estimated effects are much smaller and often insignificant. This indicates that increasing the grant amount or the grant duration alone would have had little impact on scientific productivity. Hence, it seems that the combination of the two gave Leibniz prize recipients *truly legendary freedom* to conduct their research.

This study speaks to two strands of the literature, which have focused on two separate, but related questions.[5] First, it adds to the literature studying how scientists react to the *amount* of funding. Most of these studies compare recipients of a competitive grant to non-recipients.[6] Jacob and Lefgren (2011) use an instrumental variables approach to compare recipients of National Institutes of Health (NIH) grants to equally qualified researchers who were barely rejected for the same type of grant. They find an increase in the subsequent number of publications, but this effect is small in magnitude as the rejected researchers simply shift to other grants. Benavente et al. (2012) find a sizable increase in the number of publications for recipients of a Chilean research fund using a regression discontinuity design. Similar increases are found in the case of New Zealand (Gush et al., 2018) and Denmark and Norway (Langfeldt et al., 2015). An exception to the finding of positive effects of funding on output is Lerchenmueller (2018). He finds negative effects of NIH funding on the subsequent

[5]Although we study the most important German research prize, our within prize comparison does not directly speak to the incentive effects of receiving a prestigious research prize on subsequent productivity (Borjas and Doran 2015, Chan et al. 2014).

[6]We only survey the literature on the effects of grants on individual researchers. Whalley and Hicks (2014) study how the research output of universities reacts to increases in funding and find positive effects.

number of publications. Theoretically, it is unclear whether funding should have positive effects, zero effects, or negative effects. For example, increases in funding may be used to increase wages of scientists (Goolsbee, 1998) or scientists may shift their research strategy in such a way that their output declines, e.g. by crowding out intrinsic motivation. Our study contributes to this literature by studying an increase in funding within the same program, holding many other factors constant, such as the prestige of a grant or the strings attached to the grant. In addition, we complement the analysis of large scale programs such as NIH grants by focusing on a set of elite scientists.

Second, this paper also speaks to the literature on how elite scientists react to the *structure* of funding, such as the duration of a grant or the amount of discretion in spending funds. Azoulay et al. (2011) look at the effects of becoming a Howard-Hughes Medical Institute Investigator (HHMI) on subsequent research productivity.[7] They compare recipients of the HHMI to matched-on-observable recipients of prestigious early-career prizes who are funded by the NIH. The HHMI program has, among other features, longer funding periods, higher funding amounts, and more discretion for researchers in spending their funds than grants by the NIH. The paper finds an increase in both the overall number and the number of highly cited publications of HHMI researchers relative to non-HHMI researchers. Our different finding of a reduction in the overall number of publications is most likely explained by the different structure of the HHMI and the Leibniz Prize program. Whereas both offer scientists a lot of freedom and funding to conduct their research, HHMI researchers are subject to evaluations and renewal rounds. The Leibniz Prize, in contrast, is non-renewable and hence does not have any reward (or punishment) for long-term success (or failure). This paper contributes to this literature by showing how elite scientists react to an increase in the amount and grant duration of discretionary funding across multiple scientific disciplines. Furthermore, our approach of comparing different cohorts of Leibniz Prize recipients mitigates remaining concerns about selection on unobservables that cannot be ruled out by matching only on observables.[8]

[7] An additional example is Wang et al. (2018) who survey scientists in Japan and compare the correlation between the novelty of research and whether it was based on competitive or block funding.

[8] This paper also complements recent analyses of incentive structures in German academia, but focuses on a set of elite scientists instead of the universe of management researchers (Bian et al., 2016) or the universe of university scientists (Ytsma, 2017).

The remainder of this paper is structured as follows. Section 2 describes the Leibniz Prize and the reform of 2007 in detail. Section 3 explains measurement, data and identification. Section 4 presents the results and Section 5 discusses the mechanism. Section 6 concludes.

2. The Gottfried Wilhelm Leibniz Prize

The Gottfried Wilhelm Leibniz Prize is the most important research prize in Germany. Since 1986, the DFG has awarded it annually to around ten recipients. The prize is both recognition of past achievement and *aims to improve the working conditions of outstanding researchers, expand their research opportunities, relieve them of administrative tasks, and help them employ particularly qualified early career researchers*.[9] It is awarded to peak career researchers (recipients are on average 45 years old when receiving the prize) across all scientific disciplines. Apart from prestige and accolades, prize winners receive a non-renewable seven figure research grant for several years. This research funding is not attached to specific projects, institutions, or other strings and can be spent by prize winners with full discretion, as long as it is for research purposes. Anecdotally, funds have been spent on hiring junior scientists, purchasing equipment and books, undertaking travel and expeditions, and hosting guests (Finetti, 2010).

In this paper, we exploit a change in the amount of funding and in the funding period in 2007. Prior to 2007, each prize was endowed with €1.55m in total and the funds could be spent over five years.[10] From 2007 onwards, the funding for each prize was increased to €2.50m and the funds could now be spent over a period of seven years, increasing both grant size and duration by at least 40 percent. The DFG undertook this reform to adjust the funding amount for inflation (funding had stayed constant in nominal terms from 1986 to 2006), to signal the status of the Leibniz Prize as the premier research prize in Germany, and in response to feedback of past recipients that the time period was too short for some projects (Finetti, 2010). This

[9] See `http://www.dfg.de/en/funded_projects/prizewinners/leibniz_prize/index.html`, last accessed on 01 August 2018.

[10] In less than 10 percent of cases, one *prize* is split among several (usually two, in one case four) *prize winners*. In these cases, the award sum is split equitably. Furthermore, prior to 2002, the DFG differentiated between theoretical and more capital intensive research. Theoretical researchers received half as much funding as researchers more reliant on physical equipment.

reform was announced publicly on 30 May 2006 by the DFG, prior to the communication of decisions for the 2007 prize in December 2006.[11] Hence, any anticipation effect of scientists should be minimal. In addition, apart from the change in funding amount and period, no other feature of the prize changed. Importantly, the funding amount and time frame is the same across disciplines, only researchers affiliated with a German research institution are eligible, and the nomination process also remained unchanged.

The amount of funding from the Leibniz Prize and the increase in 2007 constitute a sizable shock to recipients' research budgets. For example, in 2010 the average amount of third-party funding per university professor in Germany was €261,000 (Statistisches Bundesamt, 2014b, p.70). In medicine and engineering, which have the highest third party funding of all scientific fields, average funds per professor were around €550,000 (Statistisches Bundesamt, 2014b, p.70). Hence, the reform of the Leibniz Prize with its increase in funding of €1m constitutes at least two years of third party funding for the average university professor. Some Leibniz Prize recipients head research institutes outside of universities, such as Max-Planck Institutes or institutes of the Helmholtz Society. Unfortunately, there is no systematic data on the average third-party funding per professor for these institutes. A back-of-the-envelope calculation for one selected institute suggests average third-party funding of around €1m per professor.[12] Here, the increase in Leibniz Prize funding constitutes a smaller, but still relevant shock.

Researchers cannot apply themselves for the Leibniz Prize, but must be nominated by a third party. These third parties are mostly universities (represented by their presidents) who put forward a slate of nominations to the DFG. Each year, around 120 to 150 researchers are nominated. In a next step, the selection committee for the Leibniz Prize of the DFG comes up with a recommendation and the final selection is then made by the joint committee of the DFG, its main funding decision body. The selection committee for the Leibniz Prize consists of former prize winners and other

[11] Prize winners themselves are informed shortly before the general public each year in early December. The prize itself (and the funding) is then awarded in March of the following year.

[12] Specifically, we look at the Alfred-Wegener-Institute for Polar and Marine Research. In 2010, this institute had €21m in third-party funding and 21 professors according to its annual report (Alfred-Wegener-Institut für Polar- und Meeresforschung in der Helmholtz-Gemeinschaft, 2012).

eminent scientists.[13] The joint committee consists of scientists as well as representatives of the German federal government and the state governments. Due to this highly regulated multistage process, any strategic selection of scientists into earlier or later prize years seems unlikely.

In line with the aim of the Leibniz Prize to fund outstanding researchers, recipients of the Leibniz Prize have gone on to receive other distinctions as well. Seven recipients have received a Nobel Prize and two received the Fields Medal in mathematics.[14] Most prize winners are tenured professors when they receive the prize and usually continue in academia, often taking on prestigious positions such as heading research institutes of the Max Planck Society, the Fraunhofer Society, and the Helmholtz Society.[15]

3. Empirical Framework

In this section, we first describe how we measure scientific productivity and which data sources are used. Additional details on the data construction can be found in Appendix A. Second, we present the identification strategy and evidence for the validity of the identifying assumption.

3.1. Data and Measurement

The sample encompasses all Leibniz Prize recipients from 1986 to 2010, except those from the humanities and law. These fields are not covered in our publication data. In total, we study 257 Leibniz Prize winners, of whom 36 received the prize after 2007. We follow the literature and use bibliometric measures to proxy for scientific output and productivity (e.g. Azoulay et al., 2011; Borjas and Doran, 2015; Lee et al., 2015;

[13]The composition of the selection committee did also not change substantively in 2007. Several old members left and new members joined the committee, but not out of line with turnover in previous year (Finetti, 2010). To our knowledge, there are no term limits for the committee.

[14]Nobel laureates are Christiane Nüsslein-Volhard (Leibniz Prize: 1986/ Nobel Prize: 1995), Erwin Neher (1987/1991), and Bert Sakmann (1987/1991) in physiology, Hartmut Michel (1986/1988), Gerhard Ertl (1991/2007), and Stefan W. Hell (2008/2014) in chemistry, and Theodor W. Hänsch (1989/2005) in physics. Fields medalists are Gerd Faltings (Leibniz Prize: 1996/ Fields Medal: 1986) and Peter Scholze (2016/2018).

[15]Two notable exceptions are Wolfgang A. Herrmann (Leibniz Prize 1987), who has served as president of Technical University Munich since 1995, and Joachim Milberg (Leibniz Prize 1989), who was the CEO of BMW from 1999 to 2002 and chairman of the board from 2004 to 2015.

Wang et al., 2017a). All of our measures are based on data from Microsoft Academic, which contains information on the title, authors, outlet, document type, references, and abstracts of scientific publications including journals, some books, and conference proceedings (Tang et al., 2008, Sinha et al., 2015). In our primary set of results, we focus on *publication counts*, irrespective of the type of publication (journal article, conference proceeding, book, book chapter). We weight all publications equally. In additional regressions, we split publication counts in academic journals by the quality of the outlet. Journals are ranked in year t according to the average number of citations received by the papers in this journal in years $t-3$ to $t-1$.[16] We then construct percentiles of this ranking in each year and count all publications in the top 1%, the top 2%, the top 3% and so on. This cumulative sum mitigates the issue that scientists do not publish in all percentiles and hence looking at individual percentiles would be too noisy.

To get a more nuanced measure of whether and how the output of Leibniz Prize winners changes, we look at the *text similarity* of abstracts.[17] We focus on two measures that are defined for each scientist. The first measure compares publications within a given year to each other to identify how broad the research portfolio of a winner is. The second measure compares the publications in a given year to the early stock of publications. This measures how much a prize winner branches out over time. We use standard text analysis methods and first pre-process all abstracts by removing very common words (stop words) and by stemming words (i.e. `innovate` and `innovation` are stemmed to `innov`). We then treat each abstract as a document and construct a document-term matrix of all abstracts and words appearing in the corpus of abstracts. In addition, we use term frequency–inverse document frequency (tf-idf) weighting. Each abstract is a row in the document-term matrix and each word a column. For the first measure, we then calculate the similarity of all abstracts within a given year to each other by calculating the cosine-similarity between all pairs of ab-

[16]Specifically, we sum all forward citations of all papers published in a given journal j in years $t-3$ to $t-1$ and divide this by the total number of papers published in journal j in years $t-3$ to $t-1$. This is akin to the journal impact factor. Furthermore, we drop the bottom 25 percent of journals in terms of total number of publications over the publishing history of the journal and the bottom 25 percent of journals in terms of the total number of citations over the publishing history of the journal.

[17]This limits attention to publications for which Microsoft Academic includes an abstract. This is the case for two thirds of the publications in our data.

stracts and taking the average.[18] For the second measure, we treat all early abstracts (from 10 years prior to receiving the prize to six years prior) as a single document and calculate the average cosine similarity of the abstracts in a given year to this early stock.

The measures of abstract similarity are only defined within a scientist. We additionally build on the approach of Uzzi et al. (2013) to see how a prize winner's research changes relative to science as a whole. The underlying idea is that new scientific ideas are often recombinations of old ideas (Weitzman, 1998). Hence, the combination of prior literature referenced together in a publication is indicative for how novel or conventional an idea is. Uzzi et al. (2013) look at the pairwise combinations of journals referenced together in a paper and define a measure of *novelty* and *conventionality* for each publication.[19] They find that papers that score high on novelty and conventionality are more likely to be impactful, i.e., to land in the top five percent of the citation distribution. We use the approach of Lee et al. (2015) that is based on the same idea as Uzzi et al. (2013), but computationally easier to implement.

The basis for measuring *novelty* and *conventionality* on the paper level is to look at how common a combination of two referenced journals is. Simply counting how often two journals are referenced together would give disproportionate weight to journals that are heavily cited. Hence, we standardize the number of actually observed co-appearances with the expected number of co-appearances. For the expected number of co-appearances, we assume that journals are cited independent of each other and calculate how many co-appearances one would expect based on how often each individual journal is cited.[20] We follow Lee et al. (2015) and sort all journal combinations on the publication level by their commonness. We then measure the *novelty* of a paper

[18] The cosine-similarity between two abstracts A_1 and A_2 is given by the cosine-similarity between the two corresponding row vectors in the document-term matrix, $similarity = \frac{\sum_{i=1}^{n} A_{1,i} A_{2,i}}{\sqrt{\sum_{i_1}^{n} A_{1,i}^2}\sqrt{\sum_{i_1}^{n} A_{2,i}^2}}$.

[19] Specifically, they compare how the observed frequencies of observing a given journal pair compare to those expected by chance. Comparing these two distributions, one can generate a z score for each journal combination. Uzzi et al. (2013) then use two summary statistics on the publication level to characterize how novel (a paper's 10[th] percentile z score) and how conventional (a paper's median z score) a given paper is.

[20] Letting t denote a given year, the commonness between two journals j_1 and j_2 is defined as $\frac{N_{j_1,j_2,t}}{\frac{N_{j_1,t}}{N_t} \cdot \frac{N_{j_2,t}}{N_t} \cdot N_t}$ where $N_{j_1,j_2,t}$ is the number of times journal j_1 and journal j_2 are referenced together in year t. N_t, $N_{j_1,t}$, and $N_{j_2,t}$ are the number of all journal pairs, the number of journal pairs containing j_1 and the number of pairs containing j_2 in year t, respectively.

using the 10^{th} percentile of this distribution. As in Lee et al. (2015), we transform this value using the negative of the logarithm to facilitate the interpretation such that increases in this score are increases in novelty. In addition, we use the (logarithm) of the 50^{th} percentile as a measure of the *conventionality*.

All of our measures are ex-ante measures, i.e., they measure the novelty, conventionality, and quality at the point of publication. We cannot condition on ex-post measures as evidenced by citations as the reform of the Leibniz Prize was fairly recently and the publications of the treatment cohorts have not had enough time to garner citations. This issue is especially relevant as Wang et al. (2017b) have shown that papers with high novelty take longer to accrue citations.[21]

3.2. Identification

To estimate the impact of the increased funding amount and duration, we use a difference-in-differences model. We compare the change in productivity of prize winners receiving the increased grant size and duration to that of prize winners under the old scheme. This means we compare the productivity of the prize winners from 2007 to 2010 to the productivity of the prize winners from 1986 to 2006, both before and after receiving the Leibniz Prize. As pre-period we study the ten years prior to receiving the Leibniz Prize and as post period we use the seven years after prize reception. We focus on this window to capture the full seven years of Leibniz Prize funding after the reform of 2007. The ten years prior to receiving the Leibniz Prize are chosen to limit attention to the period where researchers have usually received tenure and head their own labs or research groups.[22] We cannot study any prize winners who received their prize after 2010 as they have not yet completed their Leibniz Prize funding period.

We estimate the following difference-in-differences specification for the prize cohorts

[21] Appendix B presents results weighting all publication counts with three year forward citations, excluding the most recent years. The results are qualitatively similar.

[22] On average, professors in Germany receive tenure at age 41 (Statistisches Bundesamt, 2014a, p.189). In addition, Leibniz Prize recipients receive the prize on average 24 years after their first publication.

from 1986 to 2010:

$$\begin{aligned} y_{i,t} = \beta_1 \cdot \text{Post Prize}_t + \beta_2 \cdot \text{Post 2007}_i \cdot \text{Post Prize}_t \\ + \text{Winner FE} + \text{Year FE} + \epsilon_{i,t} \end{aligned} \tag{1}$$

where i indexes prize winners and t indexes time. As dependent variable, $y_{i,t}$, we use the various measures of scientific productivity described above. *Post Prize* is an indicator equal to one for the years after receiving the Leibniz Prize. *Post 2007* is an indicator equal to one if the prize recipient received her prize between 2007 and 2010. We include both winner and (calendar) year fixed effects as controls.[23] Standard errors allow for clustering on the level of the prize cohort to account for serial correlation across years and potential correlation within prize cohorts. Since this yields relatively few clusters (25), we also use the wild cluster bootstrap proposed by Cameron et al. (2008) as a robustness check in Appendix B. We use weights to adjust for the different number of control and treatment winners to arrive at the average treatment effect on the treated (Iacus et al., 2012).[24]

The coefficient of interest, β_2, measures the average yearly change in the dependent variable in the seven years after receiving the Leibniz Prize post 2007 relative to the period before receiving the prize and relative to the prize winners receiving the Leibniz Prize prior to 2007. To be able to interpret β_2 causally, Leibniz Prize recipients prior to 2007 must be a good counterfactual for Leibniz Prize recipients after 2007. Although this assumption is fundamentally untestable, we present several pieces of evidence for its plausibility. First, the selection mechanism and criteria of the Leibniz Prize did not change with the reform and hence the selection of prize recipients should be the same.

Second, on observables the two groups look similar. Summary statistics for all main variables prior to receiving the Leibniz Prize can be found in Table 1. Leibniz Prize recipients are on average 28 when they receive their PhD and 45 when they receive the Leibniz Prize. The natural sciences account for the largest share of prize recipients and the social sciences for the smallest. All of these covariates are balanced across

[23]Note that the baseline effect of *Post 2007* is taken up by the winner fixed effects.

[24]Specifically, we calculate the weights for each broad field by type of institution at prize reception. There are eight strata in total.

the two groups, as expected given that the selection process remained unchanged. However, there are significant differences in the share of female recipients, the share of recipients at research institutes such as Max Planck Institutes, and the number of authors per publication. This is likely due to a general time trend in academia, an issue we revisit below.[25] In terms of our outcome variables, there are differences in levels for the overall publication count and the text similarity of abstracts relative to the early stock of publications. Figure 1 shows the kernel density for the distribution of publications by percentile of the journal impact factor ranking prior to the prize. Both groups publish the largest share of their papers in top ranked journals and generally exhibit a similar distribution.

Lastly, we estimate yearly treatment effects by interacting the treatment dummy with an indicator for each year before and after the prize.[26] For brevity, we once again focus on the overall number of publications as our main dependent variable.[27] As can be seen in Figure 2, the coefficients prior to receiving the prize are all insignificant and centered around 0. This is further evidence for the plausibility of the identifying assumption that Leibniz Prize recipients from 1986 to 2006 are a valid counterfactual to Leibniz Prize recipients from 2007 to 2010.

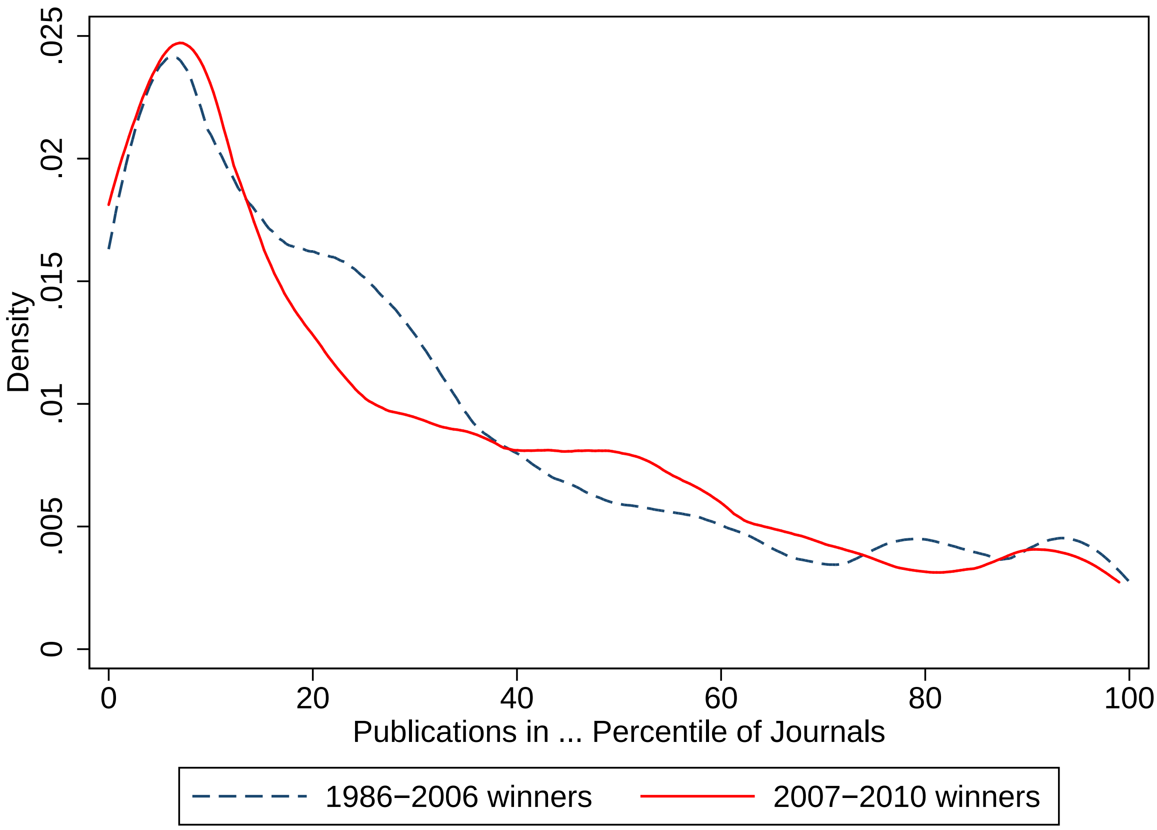
A separate concern would be broader time trends affecting prolific researchers at the same time as the reform of the Leibniz Prize. An example would be the introduction of ERC grants in 2007, which may have changed how researchers react to receiving a Leibniz Prize independently of the reform of the Leibniz Prize itself. We tackle this issue in Section 4.2.

4. Results

This section presents the main findings. Leibniz Prize recipients react to the increased funding amount and increased funding period in two ways. They reduce their overall number of publications and increase the number of articles in top ranked journals. This finding is not driven by concurrent shocks such as the introduction of ERC grants.

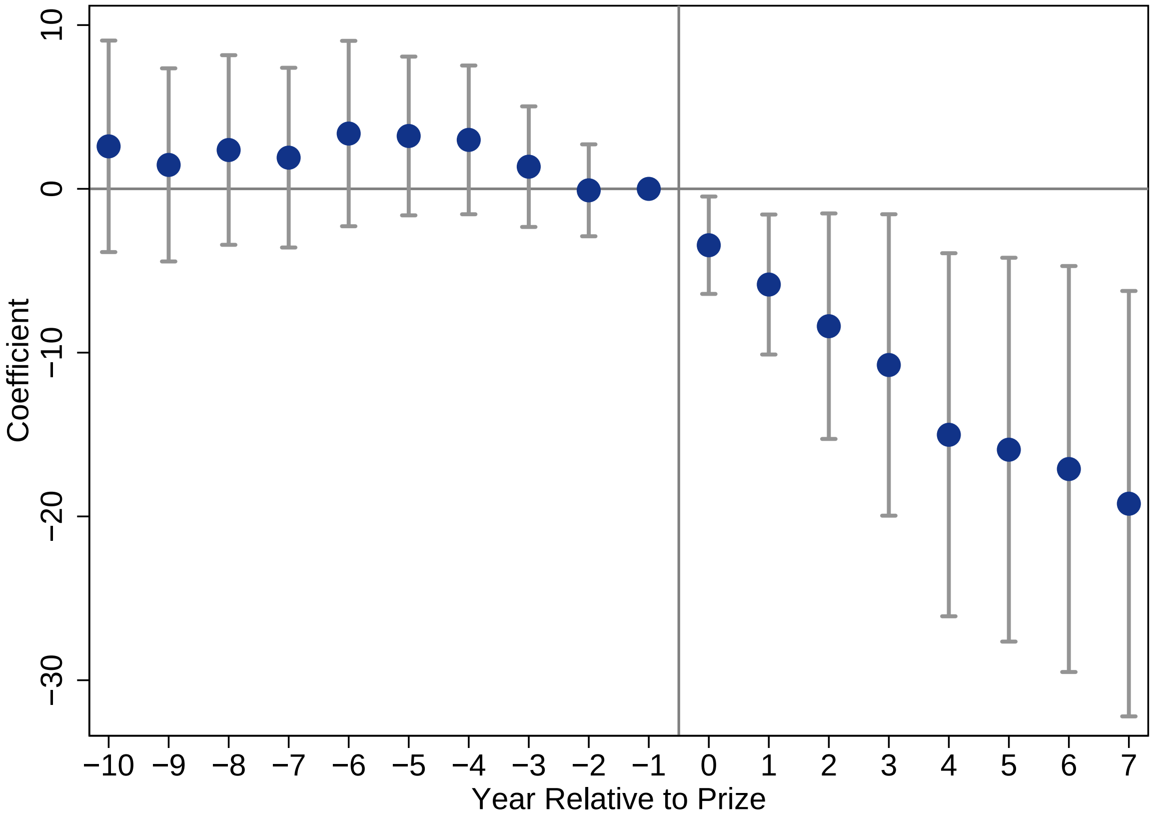
[25]See e.g. Wuchty et al. (2007) for evidence on increasing team sizes.
[26]The estimation equation is: $y_{i,t} = \sum_{\tau=-5}^{7} \beta_\tau \cdot \text{Post 2007}_i \cdot \mathbb{1}\{t = \tau\} + \text{Winner FE} + \text{Year FE} + \epsilon_{i,t}$.
[27]Figure B.2 in Appendix B present the results for the other dependent variables.

Figure 1: Density of Publications by Journal Rank Prior to Prize



Note: This figure shows the kernel density of the number of publications by percentiles of the journal quality ranking, separately for the treatment and control group for all publications published in the decade prior to receiving the Leibniz Prize. The ranking of journals is based on the average citations to the publications in the journal in the three years prior.

Figure 2: Time-Varying Treatment Effect on the Number of Publications



Note: This figure shows the yearly average treatment effects on the treated of receiving the Leibniz Prize in 2007 or later on the average number of publications (all types) per year relative to the average number of publications of researchers who received the Leibniz Prize in 2006 or prior. In all regressions, we use the weights suggested by Iacus et al. (2012) to identify the average treatment effect on the treated, calculated within each institution type by appointment and broad scientific field stratum. 95 percent confidence intervals are based on standard errors clustered on the year of prize reception.

Table 1: Summary Statistics Prior to Prize

	1986-2006	2007-2010	Difference	
Age at Prize	45.19	45.03	0.16	(0.83)
Age at PhD	27.74	27.92	−0.18	(0.60)
Social Sciences	0.06	0.08	−0.02	(0.62)
Engineering	0.17	0.19	−0.03	(0.71)
Life Sciences	0.31	0.31	0.01	(0.94)
Natural Sciences	0.46	0.42	0.04	(0.62)
Female	0.07	0.19	−0.13*	(0.07)
University	0.78	0.56	0.23**	(0.01)
Number of authors per pub	3.15	3.88	−0.73***	(0.00)
Number of publications per year	6.43	10.89	−4.46***	(0.00)
Avg. novelty of referenced journals	−0.89	−0.66	−0.22	(0.35)
Avg. conventionality of referenced journals	1.93	1.99	−0.06	(0.82)
Avg. abstract similarity	0.09	0.09	0.00	(0.47)
Avg. similarity to early abstracts	0.16	0.19	−0.02***	(0.01)
Observations	221	36	257	

Note: This table shows summary statistics for the treatment and control group prior to receiving the Leibniz Prize. Values are averaged over the five years prior to receiving the Leibniz Prize. *, ** and *** denote significance on the 10 percent, 5 percent and 1 percent level, respectively.

4.1. Main Findings

Table 2 presents the main results using the baseline specification from equation (1). In all columns, the dependent variable is the count of all types of publications per year. The different columns introduce different type of time fixed effects, e.g. in column (2) we control for a separate calendar year fixed effect for each broad scientific field. Independent of the type of time fixed effects, the coefficients are negative, large, and statistically significant on standard levels. In terms of relative magnitude, the estimates imply a decrease of at least 64 percent.

The timing of the effect can be seen in Figure 2. The number of publication decreases fairly quickly and continues to fall over time. In addition, all coefficients after receiving the Leibniz Prize are statistically significant. Given that turnaround times in the natural and life sciences are much quicker than in economics, this is plausible.[28]

Next, we investigate whether the reform of the Leibniz Prize had an affect on the type

[28]Moreover, another immediate response might be to stop existing projects that were not submitted for publication yet after receiving the Leibniz Prize to focus on other projects.

Table 2: Effect of the Leibniz Prize Reform on the Number of Publications (Diff-in-diff Estimates)

	Number of Publications (all)				
	(1)	(2)	(3)	(4)	(5)
Post Prize	1.51**	1.33**	1.46**	1.38*	1.44**
	(0.71)	(0.53)	(0.67)	(0.68)	(0.63)
Post Prize ×	−7.92*	−6.08**	−8.25*	−7.09*	−5.91**
Post 2007	(3.87)	(2.24)	(4.37)	(3.61)	(2.63)
Fixed Effects	Year	Year × Field	Year × Univ.	Year × Gender	Year × Subject
Mean Dep.	9.22	9.22	9.22	9.22	9.22
R^2	0.22	0.39	0.27	0.23	0.50
Winners	257	257	257	257	257
Observations	4626	4626	4626	4626	4626

Note: This table shows the results from a difference-in-differences estimation with ten years before receiving the Leibniz Prize as pre-period and seven years after as post-period. The estimation equation is as in equation (1). The dependent variable in all columns is the count of all types of publication per year. In the first column, we include a dummy for each calendar year. In the second column, we include separate calendar year dummies for four broad scientific fields. In the third column, we include separate calendar year dummies for the type of institution at prize reception (university or research institute). In the fourth column, we include separate calendar year dummies for male and female prize recipients, respectively. In the fifth column, we include a separate calendar year dummy for thirteen scientific subjects. In all regressions, we use the weights suggested by Iacus et al. (2012) to identify the average treatment effect on the treated, calculated within each institution type by appointment and broad scientific field stratum. Standard errors in parentheses are clustered on the level of the year of the prize. *, ** and *** denote significance on the 10 percent, 5 percent and 1 percent level, respectively.

of publication. First, we differentiate publications according to quality by using the (impact factor) rank of the outlet as a proxy for quality. The results of this can be seen in Figure 3. Each point is the point estimate of the baseline model in equation (1) using as dependent variable the number of publications in the top z percentiles of the journal ranking, where z ranges from 1 to 100, where 100 implies the number of publications in all journals. The red diamond is the baseline coefficient using publications of all types.[29] One can see an increase at the very top of the distribution, with point estimates of 0.32 (p-value = 0.09), 0.92 (p-value = 0.003), and 0.60 (p-value = 0.15) for the number of publications in the top 1%, top 2%, and top 3% of journals, respectively. Going down the quality distribution, the point estimate turns negative starting with publications in the top 16%. In terms of relative magnitude, the effect for publications in the top 1% and top 2% corresponds to a more than 150% increase relative to the baseline mean and to a roughly 75% increase for the number of publications in the top 3% of journals. This indicates that the later Leibniz Prize recipients may have substituted their decrease in the quantity of publications with an increase in publications in highly ranked journals.

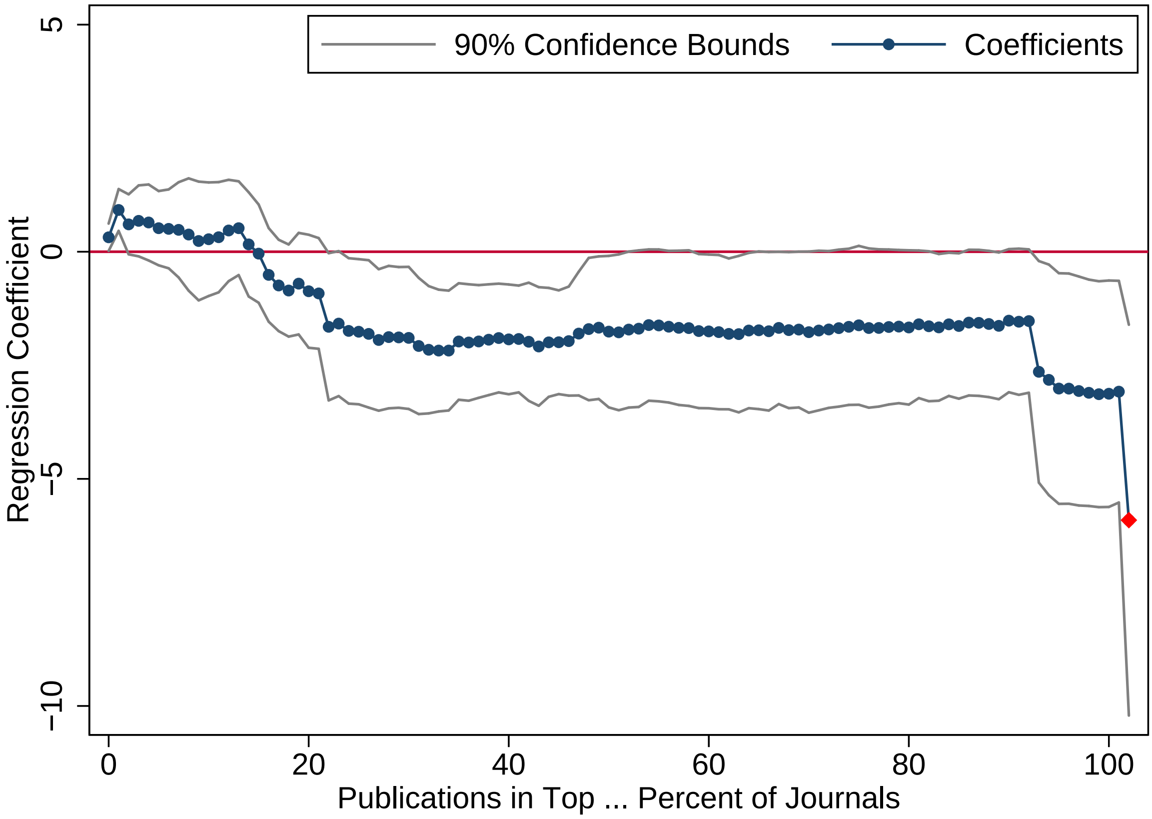
Second, we look at the text similarity of abstracts and the novelty and conventionality of reference journal combinations to assess whether the direction or nature of research changes. Table 3 presents results for these dependent variables. There is only a significant effect for the similarity of abstracts relative to the early stock of publications. The decrease implies that the treatment group moves further away from their initial research portfolio than the control group after receiving the Leibniz Prize. There is no significant effect on the similarity of abstracts to each other within a given year nor the average novelty or conventionality as measured by the combinations of journals referenced together. In addition, these point estimates are all very small.

One contribution of this study is to extend the analysis from one discipline to multiple scientific fields (we only exclude the humanities and law). In Figure 4 , we explore heterogeneity across fields by interacting the treatment indicator with an indicator for each field.[30] Figure 4 shows the results for the number of publications differentiated

[29] This count encompasses publications in all ranked journals, publications in non-ranked journals and publications outside of journals, e.g. book chapters.

[30] The field assignment of the DFG at the time of the prize is used. Since fields are defined broadly (social sciences, engineering, life sciences, natural sciences), movement of prizewinners across fields

Figure 3: Effects over the Journal Quality Distribution



Note: This figure shows the estimated coefficients of repeatedly estimating equation (1) for the number of publications in the top z percent of journals. Journals are ranked according to the average number of citations per paper in the three years prior. The red diamond depicts the baseline coefficient for the number of publications of all types, so encompassing also publications outside of journals. 90 percent confidence intervals are based on standard errors clustered on the year of prize reception. In all regressions, we use the weights suggested by Iacus et al. (2012) to identify the average treatment effect on the treated, calculated within each institution type by appointment and broad scientific field stratum.

Table 3: Effect of the Leibniz Prize Reform on Text Similarity and Novelty (Diff-in-diff Estimates)

	Text Sim. I	Text Sim. II	Novelty	Conventionality
Post Prize	-0.00	-0.01^{*}	0.10^{**}	-0.01
	(0.00)	(0.00)	(0.05)	(0.03)
Post Prize × Post 2007	-0.00	-0.02^{***}	-0.08	0.05
	(0.01)	(0.01)	(0.09)	(0.09)
Fixed Effects	Year	Year	Year	Year
Mean Dep.	0.08	0.13	-0.70	1.87
R^2	0.06	0.39	0.03	0.02
Winners	252	248	256	256
Observations	3536	2536	3974	3974

Note: This table shows the results from a difference-in-differences estimation with ten years before receiving the Leibniz Prize as pre-period and seven years after as post-period. The estimation equation is as in equation (1). The dependent variable in column (1) is the average similarity of abstracts to each other within a given year. In column (2), the pre-period is limited to five years prior to prize reception and the dependent variable is the average similarity of abstracts to the abstracts of papers published between ten and six years prior to the prize. In columns (3) and (4), the dependent variables are novelty and conventionality, respectively, as defined by Lee et al. (2015). In all regressions, we use the weights suggested by Iacus et al. (2012) to identify the average treatment effect on the treated, calculated within each institution type by appointment and broad scientific field stratum. Standard errors in parentheses are clustered on the level of the year of the prize. *, ** and *** denote significance on the 10 percent, 5 percent and 1 percent level, respectively.

by journal quality. The qualitative pattern of the coefficients is very similar across fields, though quantitatively the effects are most pronounced in the life sciences. In terms of significance, the decrease for the number of publications of all types is statistically significant for all fields except the social sciences, though the point estimate is still negative (p-value = 0.19).[31] Overall, there is little evidence of heterogeneity across fields. The effect in social sciences differs somewhat from the effect in all other fields, but by and large is qualitatively similar. One potential reason may be that knowledge production and the publication process work differently in social sciences than in other fields. However, the small number of social scientists in the sample does not allow us to draw firm conclusions.

Taken together, increasing funding by €1m and extending the time frame by two years had the following effect: Leibniz Prize recipients reduce their overall number of publications. However, they increase their number of publications in top ranked journals. The effects are also large in economic terms, with a decrease in the overall number of publications of more than 60 percent and an even larger increase in the number of publications in top ranked journals. There is little evidence that there is an effect on the direction or content of research. This pattern is similar across fields.

4.2. Robustness: Is This a Cohort Effect?

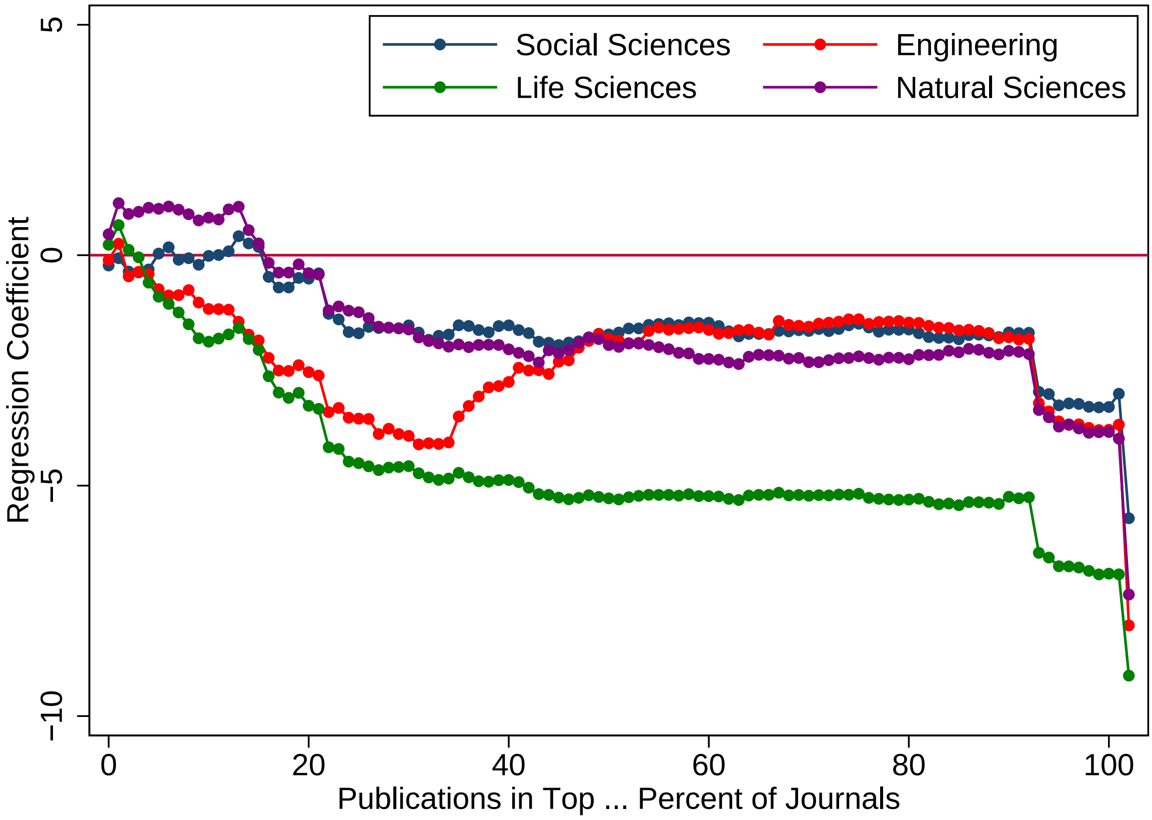
One key concern are cohort effects, i.e. other time-varying shocks that differentially affect younger and older researchers. For example, the introduction of ERC grants in 2007 may have affected the treatment group differently than the control group, as the former was younger than the latter in 2007. Additional examples are the so-called "Exzellenzinitiative" for German universities starting in 2005/06 or the decision of the DFG to limit the length of publication lists on grant applications in 2010. The introduction of these programs may have changed the incentives regarding which type of research to focus on. Younger researchers might respond more strongly to this change, as career concerns loom larger for them.

We argue that these concerns cannot explain the main results using three pieces of ev-

does not play a role.

[31] Table B.1 in the Appendix presents the results for text similarity and the novelty/conventionality of reference journal combinations.

Figure 4: Heterogeneity by Field – Publication Counts



Note: This figure shows the estimated coefficients of repeatedly estimating $y_{i,t} = \beta_1 \cdot \text{Post Prize}_t + \beta_{\text{Social Sciences}} \cdot \text{Post 2007}_i \cdot \text{Post Prize}_t \cdot \mathbb{1}\{\text{Field} = \text{Social Sciences}\} + \beta_{\text{Engineering}} \cdot \text{Post 2007}_i \cdot \text{Post Prize}_t \cdot \mathbb{1}\{\text{Field} = \text{Engineering}\} + \beta_{\text{Life Sciences}} \cdot \text{Post 2007}_i \cdot \text{Post Prize}_t \cdot \mathbb{1}\{\text{Field} = \text{Life Sciences}\} + +\beta_{\text{Natural Sciences}} \cdot \text{Post 2007}_i \cdot \text{Post Prize}_t \cdot \mathbb{1}\{\text{Field} = \text{Natural Sciences}\} + \text{Winner FE} + \text{Year FE} + \epsilon_{i,t}$ for the number of publications in the top z percent of journals. Journals are ranked according to the average number of citations per paper in the three years prior. The last coefficient on the right depicts the baseline coefficient for the number of publications of all types, so encompassing also publications outside of journals. In all regressions, we use the weights suggested by Iacus et al. (2012) to identify the average treatment effect on the treated, calculated within each institution type by appointment and broad scientific field stratum.

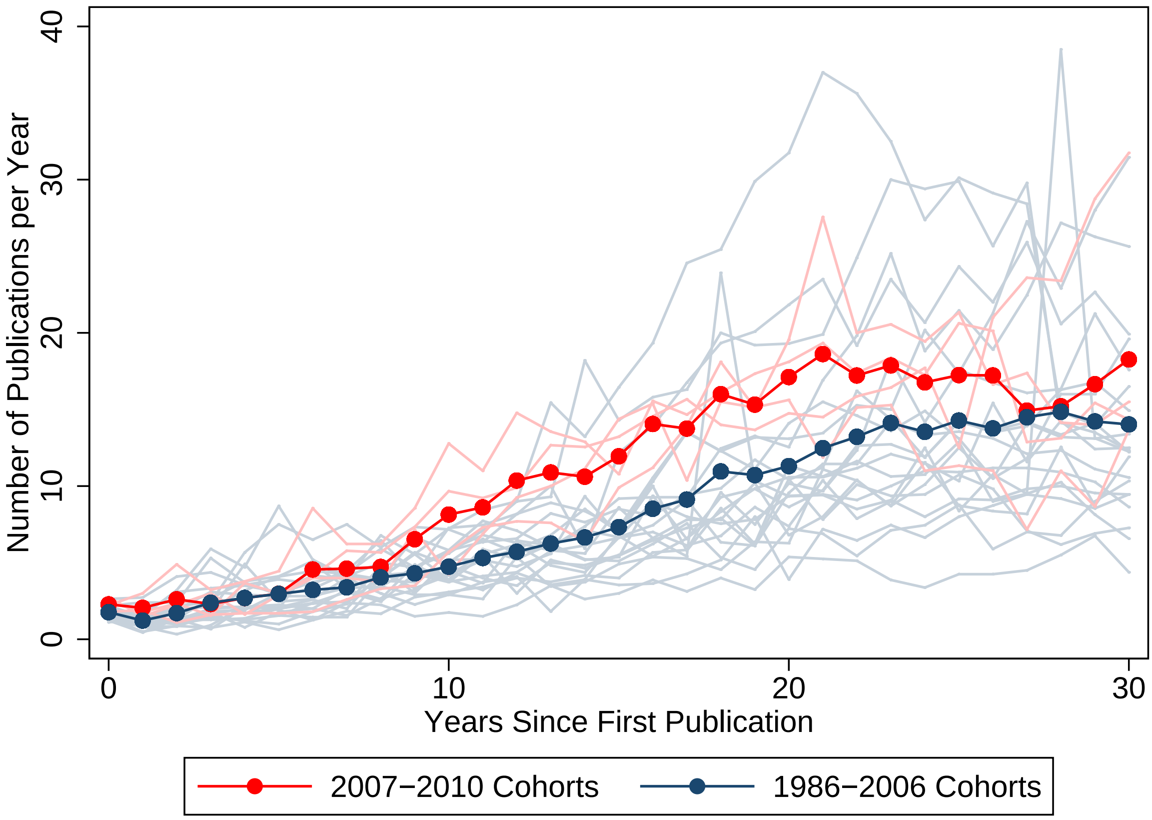
idence. First, we show that the life-cycle profiles, i.e. the publication output over the career of a researcher relative to the start of her career, does not differ greatly between the treatment and control group. Second, we construct a matched control group of eminent scientists who do not receive a Leibniz Prize and run a placebo regression within this sample. Lastly, we conduct a difference-in-differences-in-differences specification using this matched control group as an additional control group.

To investigate whether our treatment effect can be explained by changes in the life-cycle profiles of academics that happen to coincide with the Leibniz Prize, we graphically examine these life-cycle profiles. We count all publications per year from the first year a scientist published up to 30 years after. Years without a publication are assigned zero. Figure 5 shows the results for Leibniz Prize recipients. The light blue and red lines are the average number of publications per year for each prize cohort, where blue lines indicate prize cohorts prior to 2007 and red lines indicate the Leibniz Prize recipients after 2007. The connected solid lines are the averages for the treatment and control group, respectively. One sees that in general the life-cycle profiles are upward sloping and the treatment group does not seem to be on a fundamentally different trend. There is a stronger increase in the number of publications for the treatment group around nine to 12 years after their first publication, but then the two groups develop in a parallel fashion.[32] The comparison of life-cycle profiles suggests that it is not a mere change in life-cycle productivity that happens to coincide with receiving the Leibniz Prize that drives our results.

Although any differences in life-cycle profiles at first glance do not seem to be able to explain our treatment effect, they do not fully mitigate the concern that cohort effects are driving the results. As a further check, we construct a matched control group for Leibniz Prize recipients of researchers who are similar to Leibniz Prize recipients and hence exposed to the same shocks such as the introduction of ERC grants, but do not receive a Leibniz Prize. Using this group, we run a placebo regression where we compare their scientific output before and after they *should* have received the Leibniz Prize, comparing researchers who should have received the prize before 2007 and after 2007. We approximate the year when a researcher should have received a Leibniz Prize based on her age.

[32]On average, recipients receive the Leibniz Prize 24 years after their first publication.

Figure 5: Number of Publications over the Life Cycle



Note: This figure plots the number of publications per year relative to the first year of publication for all Leibniz Prize recipients. Each shaded line represents the average for one prize cohort. Blue lines indicate control group prize cohorts and red lines the treament group cohorts. The connected lines depict the respective averages for the treatment and control group.

We construct the matched control group using the "categories" feature in Wikipedia to coarsened exact match: For each Leibniz Prize recipient, we obtain all individuals who share the same year of birth category (e.g. "born in 1972"), the same broad field category (e.g. "mathematician"), and the same gender. In addition, we require control group recipients to be (or have been) academics at a German higher education institution. This results in 1,819 individuals matched to 219 Leibniz Prize recipients.[33] The number of controls per Leibniz Prize recipient ranges from one to 37.[34] We then retrieve the same type of publication data as for the Leibniz Prize recipients from Microsoft Academic.[35]

Whilst the individuals in the matched control group are matched on eminence (they all have a Wikipedia entry), age, gender, field, and being (or having been) active in German academia, we do not match on any publication counts. Hence, we can check whether the matching works and the matched control group from Wikipedia appears to be a reasonable control group for the Leibniz Prize recipients. Figure 6 shows the number of publications (all types) of these placebo prize recipients, in addition to the actual Leibniz prize recipients, from ten years before the prize to seven years after. Whilst there are level differences between the recipients before and after 2007, the trends are similar and even more importantly, the trends of the placebo prize recipients and actual Leibniz prize recipients are very similar. This indicates that the matching does indeed work.

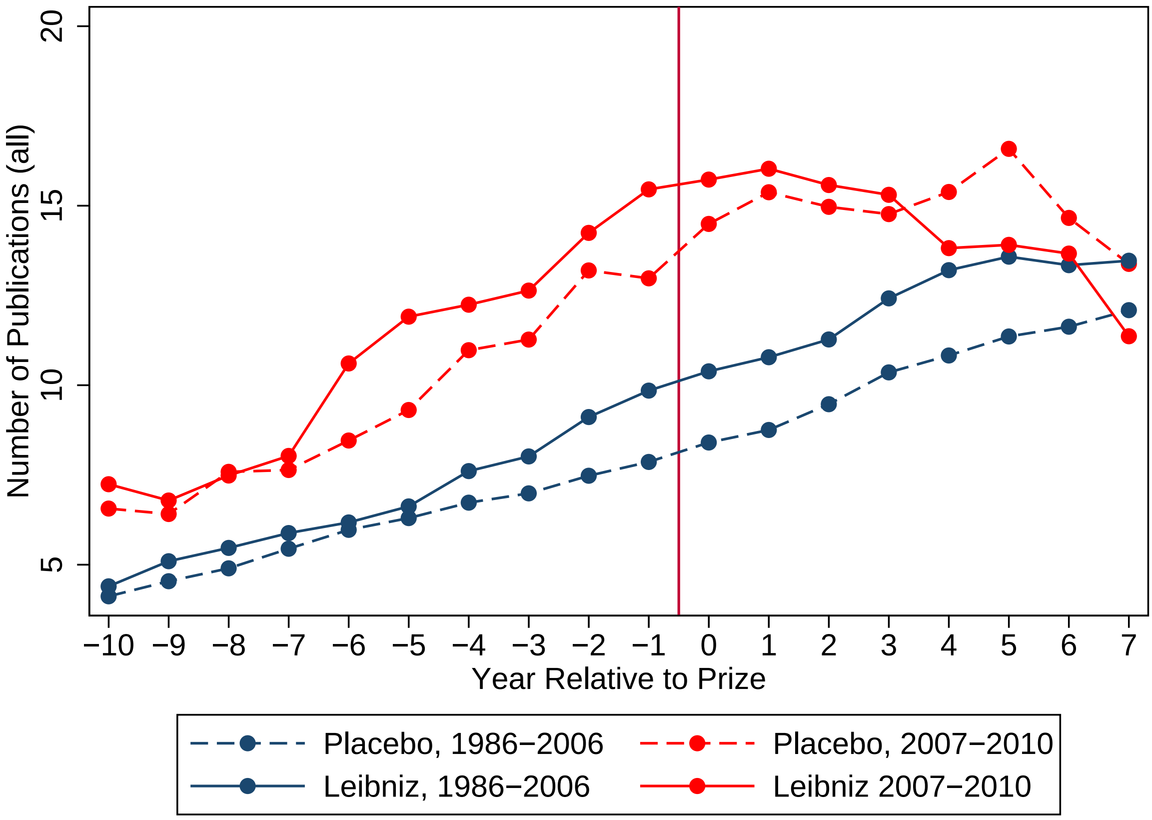
Figure 7 and Table B.2 repeat the main analysis within the matched control group. If the effects we observed earlier were due to concurrent changes such as the introduction of ERC grants or the "Exzellenzinitiative", we would expect to observe a similar pattern as amongst the Leibniz Prize recipients. However, in Figure 7, the coefficients are much smaller in absolute magnitude and tend to be positive. In addition, the baseline coefficient on the number of all publications (red diamond), is very small

[33] We are unable to match all Leibniz Prize recipients for two reasons. First, there is not a Wikipedia article for all Leibniz Prize recipients. Second, for some Leibniz Prize recipients who are featured in Wikipedia, there is no corresponding match.

[34] Note that the individuals in the matched control group may be matched to multiple Leibniz Prize recipients.

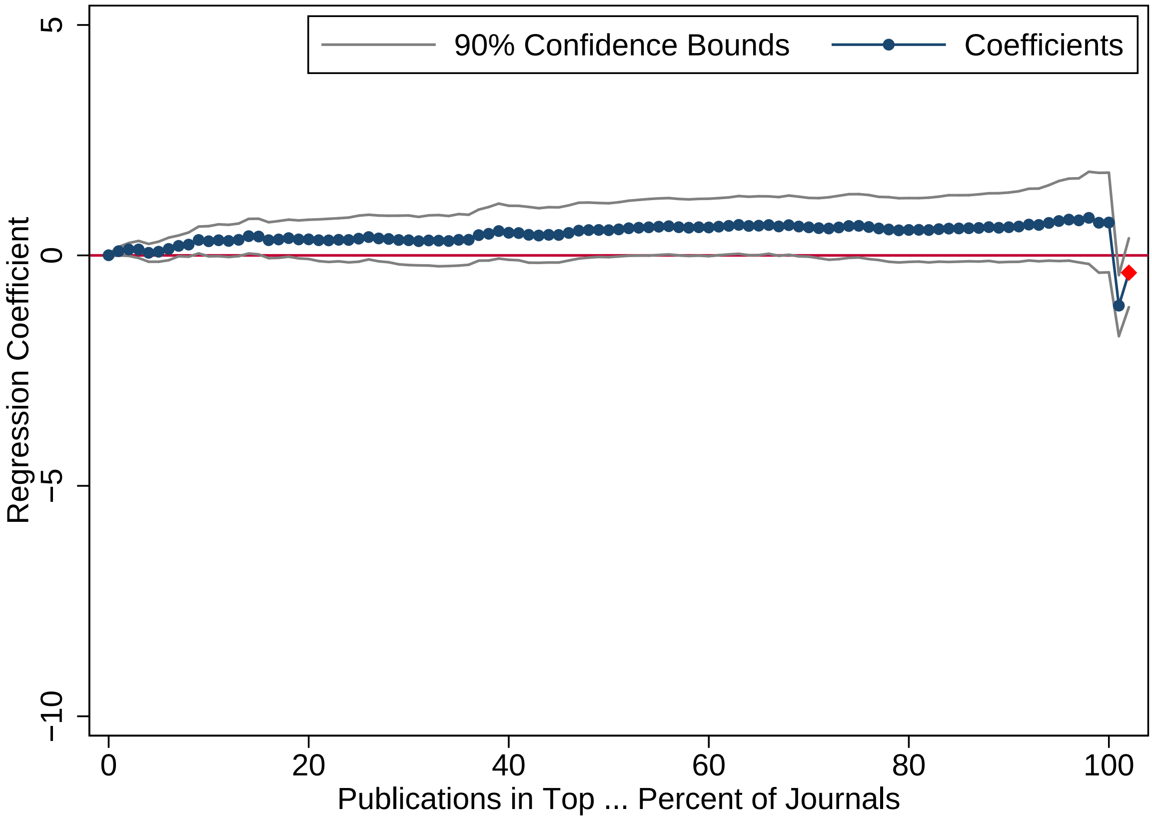
[35] Since there are so many more individuals in the matched control group than Leibniz Prize recipients, it is not feasible to engage in the same type of extensive manual cleaning of the data. Hence, we drop outliers and drop the bottom 5% and top 1% in terms of lifetime publications in the matched control group. This corresponds to a total number of lifetime publications of less than 4 and more than 1291, respectively.

Figure 6: Number of Publications Relative to Leibniz Prize: Actual and Placebo Recipients



Note: This figure shows the yearly average number of publications (all types) for actual and placebo Leibniz Prize recipients. Averages are weighted with the weights suggested by Iacus et al. (2012) within each year of birth, field, and gender stratum. Year of prize reception for placebo recipients is approximated by their age.

Figure 7: Effects over the Journal Quality Distribution – Placebo Exercise



Note: This figure shows the estimated coefficients of repeatedly estimating equation (1) for the number of publications in the top z percent of journals for the sample of placebo Leibniz prize recipients. Journals are ranked according to the average number of citations per paper in the three years prior. The red diamond depicts the baseline coefficient for the number of publications of all types, so encompassing also publications outside of journals. 90 percent confidence intervals are based on standard errors clustered on the year of prize reception. In all regressions, we use the weights suggested by Iacus et al. (2012) to identify the average treatment effect on the treated, calculated within each broad scientific field stratum.

and statistically insignificant, indicating that concurrent shocks to not drive the observed pattern.[36]

Apart from the placebo exercise, we can also use the matched control group in a triple

[36]Table B.2 in Appendix B shows the results for the other dependent variables.

differences approach. The full triple differences specification is

$$\begin{aligned} y_{i,t} = {} & \beta_1 \cdot \text{Post Prize}_t + \beta_2 \cdot \text{Post Prize}_t \cdot \text{Leibniz}_i \\ & + \beta_3 \cdot \text{Post Prize}_t \cdot \text{Post 2007 Cohort}_i \\ & + \beta_4 \cdot \text{Post Prize}_t \cdot \text{Leibniz}_i \cdot \text{Post 2007 Cohort}_i \\ & + \text{Winner FE} + \text{Year FE} + \epsilon_{i,t} \end{aligned} \tag{2}$$

where Leibniz_i is an indicator equal to one if the researcher received a Leibniz Prize and all other variables are defined as in equation (1).[37] The coefficient of interest is now β_4. It measures the change in the outcome variable (e.g. the number of publications) of receiving the Leibniz Prize after 2007 (relative to before 2007), relative to the differential change between Leibniz Prize and placebo prize winners before and after receiving the Leibniz Prize.

The results of this specification can be found in Figure 8.[38] The pattern for the publication counts is very similar to the baseline estimate, with significant increases for publications in the top 20% of journals. The coefficient for the overall number of publications is -2.91, a 35% decrease relative to the mean. Broadly, the results of this triple-difference specification are in line with our baseline results, indicating that other contemporaneous shocks do not explain the results.

4.3. Additional Robustness

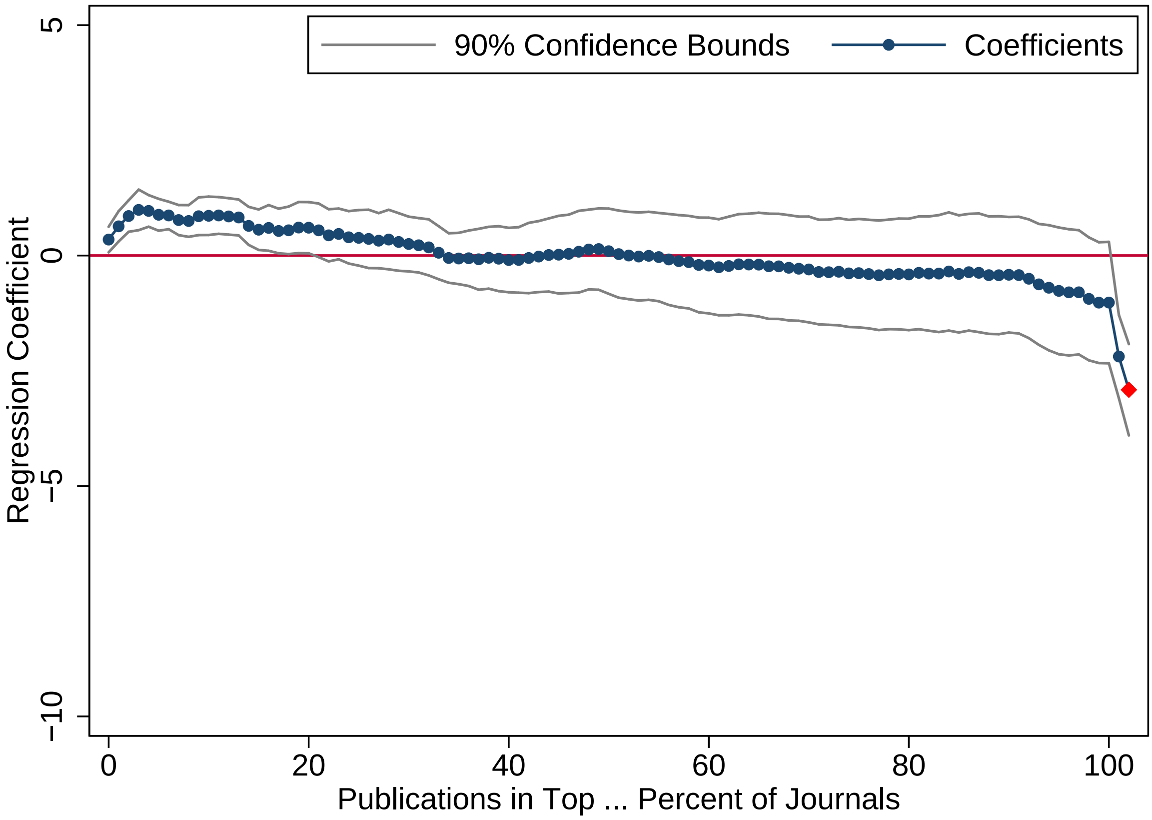
Appendix B presents additional robustness checks in detail, which we briefly summarize here. One possible concern pertains to inference. Standard errors allow for clustering on the level of the prize year. This yields comparatively few clusters (25). For most coefficients, inference is unchanged if we use the wild cluster bootstrap proposed by Cameron et al. (2008) and hence any bias from the relatively small number of clusters is likely small as well (Figures B.7 to B.9).

As a test of our identifying assumption, we conduct a placebo exercise and assign (placebo) treatments in Figure B.1. Specifically, we re-assign treatment to all overlap-

[37] The Leibniz indicator, the Post 2007 indicator, and the interaction of the two is taken up by the winner fixed effects.

[38] Table B.3 in Appendix B shows the results for the other dependent variables.

Figure 8: Effects over the Journal Quality Distribution – Triple Differences



Note: This figure shows the estimated coefficients of repeatedly estimating equation (2) for the number of publications in the top z percent of journals. Journals are ranked according to the average number of citations per paper in the three years prior. The red diamond depicts the baseline coefficient for the number of publications of all types, so encompassing also publications outside of journals. 90 percent confidence intervals are based on standard errors clustered on the year of prize reception. In all regressions, we use the weights suggested by Iacus et al. (2012) to identify the average treatment effect on the treated, calculated within each broad scientific field, year of birth, and gender stratum.

ping four-year periods, i.e. treatment is first re-assigned to Leibniz Prize recipients between 1986 and 1989, then 1987 and 1990, and so on. We replicate the analysis for the entire quality distribution of publications and plot the resulting coefficients in Figure B.1. The placebo coefficients are centered around zero and much smaller in absolute magnitude than the actual treatment coefficients.[39]

In Figure B.5 we investigate how results change if we normalize publication counts by the number of authors (Panel a) or weight publications by their forward citations (Panel b).[40] Normalizing publication counts with the (square root of the) number of authors does not alter the results. The citation-weighted results have the same qualitative pattern as the main results. However, only the increase in publications in top ranked journals is statistically significant.

Figure B.4 estimates the baseline model using a fixed effects poisson or negative binomial model. The effects are qualitatively very similar. Similarly, in Figure B.6 each prize cohort is dropped in turn to assess whether the results are driven by outliers. By and large, the overall pattern is very similar independent of which prize cohort we exclude.

5. Mechanism

The analysis so far has always focused on the interaction of increased funding amount and funding duration. Yet, it is unclear whether it is the additional funding amount, the additional time to spend the money, or a combination of the two that caused the decline in overall publications and increase in top ranked publications. In order to shed light on this issue, we exploit the fact that the *nominal* funding amount of the Leibniz Prize remained constant from 1986 to 2006, whereas money lost 45 percent of its value in *real* terms.

We look at three groups of prize recipients: the (treatment) group of 2007 to 2010 prize recipients, the 2000 to 2006 prize recipients, and the 1986 to 1992 prize recipients.

[39] To avoid overcrowding the graph, we do not plot the associated confidence bounds. However, almost all coefficients are statistically insignificant.

[40] However, the last observations of the treatment group are then dropped as these publications have not yet had time to garner forward citations.

In real terms, the 2007 to 2010 prize recipients received €2.54m, the 2000 to 2006 recipients received €1.74m, and the 1986 to 1992 prize recipients €2.35m.[41] Hence, the increase in funding amount for the 2007 to 2010 recipients is much larger relative to the 2000 to 2006 recipients than relative to the earliest prize recipients from 1986 to 1992. However, both the 2000 to 2006 group and the 1986 to 1992 group had five years to conduct their research, whereas the 2007 to 2010 recipients had seven years.

We conduct three comparisons to disentangle whether it is the grant amount, the grant duration, or a combination that matters most. First, to shed light on the effect of the grant amount alone, we compare the 1986 to 1992 cohorts to the 2000 to 2006 cohorts. We use the 1986 to 1992 group as treatment group, as these researchers had a much larger monetary amount than the 2000 to 2006 control group. Importantly, both groups had a constant five years to conduct their research. Second, to investigate the effect of grant duration alone, we compare the 2007 to 2010 cohorts to the 1986 to 1992 prize recipients. The two groups receive rather similar amounts of (real) funding (€2.54m vs. €2.35m), but the 2007 to 2010 researchers had seven years to conduct their research, whereas the 1986 to 1992 group only had five years. Third, to look at the interaction of funding amount and duration, we compare the 2007 to 2010 recipients to the 2000 to 2006 recipients. Here, the increase in real funding amount is largest (€2.54m vs. €1.74m) *and* the 2007 to 2010 recipients have two more years to conduct their research.

Figure 9 shows the results of this exercise.[42] In Panel (a), we vary only the funding amount by comparing the pre 1992 prize recipients to those researchers awarded a Leibniz Prize between 2000 and 2006. The estimated coefficients are smaller in absolute magnitude than the baseline coefficients and are positive and statistically significant at the lower parts of the journal quality distribution, suggesting an increase in the number of publications across the board in response to an increase in funding

[41] All real monetary amounts are in 2010 Euros and deflated using the consumer price index published by the Federal Statistical Office. Note that this is based on the maximum prize funding of (nominal) €1.55m and does not take into account that prior to 2002, theoretical researchers usually only received half the amount. Unfortunately, there is no data available on their share of theoretical researchers prior to 2000. In 2000 to 2002, the share is around 25 percent, implying that the 1986 to 1992 control group may have had average real funding of €2.05m and the 2000 to 2006 cohorts €1.64m. Even under this conservative estimate, the change in funding is still much larger relative to the 2000 to 2006 prize winners than relative to the very early recipients.

[42] Table B.4 shows the results for the other dependent variables.

amount, but none in funding duration. This is in line with previous research that has found small positive effects of receiving a grant on subsequent publication output (e.g. Myers, 2019). In Panel (b), we do the converse and focus on the increase in grant duration by comparing the 2007 to 2010 prize recipients to the 1986 to 1992 recipients. Similarly to Panel (a), the coefficients are small in absolute magnitude, but now are almost all statistically insignificant from zero. Lastly, in Panel (c) we only look at those receiving a Leibniz Prize after 2000. Here, the treatment group (2007 to 2010 recipients) has both a larger funding amount and a longer duration than the control group (2000 to 2006 recipients). In this comparison, the pattern is very similar to the baseline estimates, indicating that funding amount and funding duration are complementary.[43]

6. Discussion and Conclusion

At the first Leibniz Prize ceremony in 1986, the president of the DFG, Hubert Markl, chose the words *truly legendary freedom* to describe the Leibniz Prize in a nutshell. Anecdotally, many Leibniz Prize recipients also viewed the prize as giving them more freedom to conduct the type of research they wanted to do in the way they deemed appropriate. Leibniz Prize winner Herbert W. Roesky (1988) said that it *freed him from the writing of annoying grant proposals and the lecturing comments of reviewers*; Manfred Schmidt (Leibniz Prize 1995) described it as a *research paradise* (Finetti, 2010). The reform of 2007 with its increase in funding amount and period can be viewed as an increase of this *truly legendary freedom*. Due to this increase in freedom, scientists can focus on research projects resulting in publications in top ranked journals instead of the more "bread and butter" work which would be necessary otherwise to attract grant funding. In line with this interpretation, we provide tentative evidence in Appendix C that the treatment group has fewer other, non Leibniz Prize grants from the DFG compared to the control group.

We can rule out many other potential explanations for the productivity effects based

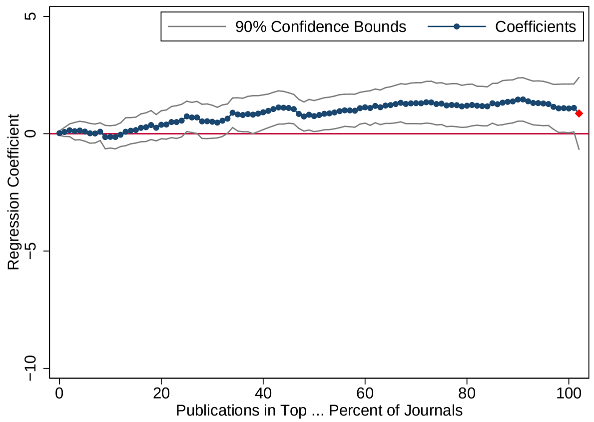
[43] In addition, this provides further evidence that our results are not an artifact of more general time trends in science. These time trends should have a much larger effect in the comparison with the 1986 to 1992 cohort, which is not borne out in the data.

on the within Leibniz Prize analysis. For example, a form of "Matthew effect" where Leibniz Prize recipients find it easier to publish their work in top ranked journals cannot explain our results, as we are conducting our comparison solely amongst Leibniz Prize recipients. The same holds for the notion that receiving a Leibniz Prize may raise the bar for research that is submitted for publication or that it raises the demand for other activities such as being asked to advise, give speeches, or sit on boards and committees.

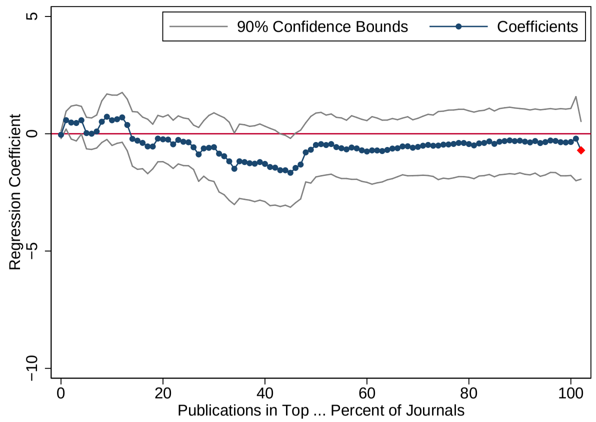
This paper sheds light on how the amount and structure of funding affect scientific output. Despite the importance of basic and applied research for long-run economic growth, we know fairly little about this issue. We use a natural experiment in the context of German academia and show how elite scientists react to an increase in *truly legendary freedom* to conduct their research. The response to an increase in grant amount of €1m and grant duration of two years is to reduce the number of publications overall, but increase publications in top ranked journals. It is most likely the combination of the increased grant amount and grant duration that drives these effects, and additional funding or time alone would not have sufficed.

The dogma of "publish or perish" is increasingly under attack. Academics and funding bodies alike bemoan that most effort is spent on publishing as many papers as possible without taking time to focus on few high quality publications. The German Research Foundation has tried to steer its grants towards more freedom and flexibility, a development called *Leibnizization* by DFG president (and Leibniz Prize recipient) Matthias Kleiner (Finetti, 2010, p.9). Although this is a promising development in light of our results, it may also require to be accompanied by an increase in the amount of funding to impact scientific productivity.

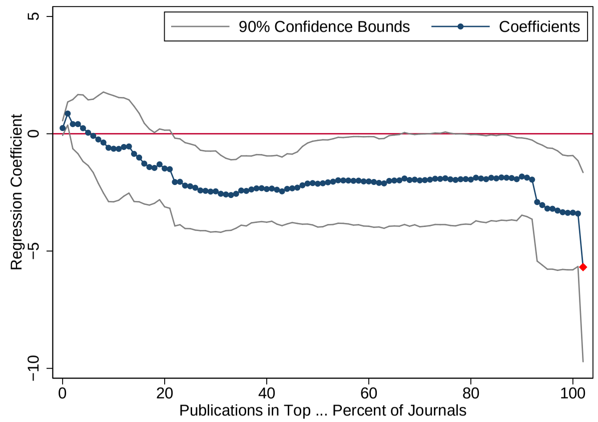
Figure 9: Mechanism: Funding Amount and Funding Duration



(a) Only Money



(b) Only Time



(c) Time and Money

Note: This figure shows the estimated coefficients of repeatedly estimating equation (1) for the number of publications in the top z percent of journals. Journals are ranked according to the average number of citations per paper in the three years prior. In Panel (a), we re-assign treatment to the prize recipients from 1986 to 1992 and use the recipients from 2000 to 2006 as control group. In Panel (b) we use the prize recipients from 1986 to 1992 as control group, whereas in Panel (c) we use only the cohorts from 2000 to 2006 as control group (the treatment group remains the same). In all regressions, we use the weights suggested by Iacus et al. (2012) to identify the average treatment effect on the treated, calculated within each institution type at appointment and broad scientific field stratum. 90 percent confidence intervals are based on standard errors clustered on the year of prize reception.

References

Alfred-Wegener-Institut für Polar- und Meeresforschung in der Helmholtz-Gemeinschaft (2012): "Geschäftsbericht 2010/2011," Bremerhaven.

Angrist, Joshua D. and Jörn-Steffen Pischke (2008): *Mostly harmless econometrics: An empiricist's companion*, Princeton University Press.

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

Benavente, José Miguel, Gustavo Crespi, Lucas Figal Garone, and Alessandro Maffioli (2012): "The impact of national research funds: A regression discontinuity approach to the Chilean FONDECYT," *Research Policy*, 41(8), 1461–1475.

Bian, Bo, Rainer Haselman, and Vikrant Vig (2016): "Incentives, Hiring and Productivity: Evidence from Academia," Mimeo.

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

Bush, Vannevar (1945): *Science, the endless frontier: A report to the President*, US Govt. print. off.

Cameron, A. Colin, Jonah B. Gelbach, and Douglas L. Miller (2008): "Bootstrap-based Improvements for Inference with Clustered Errors," *The Review of Economics and Statistics*, 90(3), 414–427.

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

Finetti, Marco (2010): "Von Märchenhafter Freiheit: 25 Jahre Spitzenforschung im Gottfried Wilhelm Leibniz-Programm," Deutsche Forschungsgemeinschaft.

Goolsbee, Austan (1998): "Does government R&D policy mainly benefit scientists and engineers?" *The American Economic Review*, 88(2), 298–302.

Gush, Jason, Adam Jaffe, Victoria Larsen, and Athene Laws (2018): "The effect of public funding on research output: the New Zealand Marsden Fund," *New Zealand*

Economic Papers, 52(2), 227–248.

Iacus, Stefano M., Gary King, and Giuseppe Porro (2012): "Causal inference without balance checking: Coarsened exact matching," *Political analysis*, 20(1), 1–24.

Jacob, Brian A. and Lars Lefgren (2011): "The Impact of Research Grant Funding on Scientific Productivity," *Journal of Public Economics*, 95(9-10), 1168–1177.

Langfeldt, Liv, Carter Walter Bloch, and Gunnar Sivertsen (2015): "Options and limitations in measuring the impact of research grants–evidence from Denmark and Norway," *Research Evaluation*, 24(3), 256–270.

Lee, You-Na, John P. Walsh, and Jian Wang (2015): "Creativity in Scientific Teams: Unpacking Novelty and Impact," *Research Policy*, 44(3), 684–697.

Lerchenmueller, Marc (2018): "Does More Money Lead to More Innovation? Evidence From the Life Sciences." *Academy of Management Proceedings*, 2018(1), 16372.

MacKinnon, James G. and Matthew D. Webb (2017): "Pitfalls when Estimating Treatment Effects Using Clustered Data," *The Political Methodologist*, 24(2), 20.

Merton, Robert K. (1968): "The Matthew effect in science: The reward and communication systems of science are considered," *Science*, 159(3810), 56–63.

Mokyr, Joel (2016): *A culture of growth: the origins of the modern economy*, Princeton University Press.

Myers, Kyle R. (2019): "The Elasticity of Science," Mimeo.

Nicholson, Joshua M. and John P.A. Ioannidis (2012): "Research grants: Conform and be funded," *Nature*, 492(7427), 34.

Sarewitz, Daniel (2016): "The pressure to publish pushes down quality," *Nature*, 533(7602).

Sinha, Arnab, Zhihong Shen, Yang Song, Hao Ma, Darrin Eide, Bo-june Paul Hsu, and Kuansan Wang (2015): "An overview of microsoft academic service (mas) and applications," in *Proceedings of the 24th international conference on world wide web*, ACM, 243–246.

Statistisches Bundesamt (2014a): "Bildung und Kultur – Personal an Hochschulen," Wiesbaden.

——— (2014b): "Monetäre hochschulstatistische Kennzahlen," Wiesbaden.

STEPHAN, PAULA (2012): *How economics shapes science*, Cambridge, MA: Harvard University Press.

STEPHAN, PAULA E. (2010): "The economics of science," in *Handbook of the Economics of Innovation*, Elsevier, vol. 1, 217–273.

TANG, JIE, JING ZHANG, LIMIN YAO, JUANZI LI, LI ZHANG, AND ZHONG SU (2008): "Arnetminer: extraction and mining of academic social networks," in *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*, ACM, 990–998.

THE ECONOMIST (2011): "MIT and the art of innovation," Babbage Blog, `http://www.economist.com/blogs/babbage/2011/01/mit_and_art_innovation`, retrieved 13 March 2019.

UZZI, BRIAN, SATYAM MUKHERJEE, MICHAEL STRINGER, AND BENJAMIN F. JONES (2013): "Atypical Combinations and Scientific Impact," *Science*, 342(6157), 468–472.

WANG, JIAN, YOU-NA LEE, AND JOHN P. WALSH (2018): "Funding Model and Creativity in Science: Competitive versus Block Funding and Status Contingency Effects," *Research Policy*, 47(6), 1070–1083.

WANG, JIAN, REINHILDE VEUGELERS, AND PAULA STEPHAN (2017a): "Bias against Novelty in Science: A Cautionary Tale for Users of Bibliometric Indicators," *Research Policy*, 46(8), 1416–1436.

WANG, YANBO, JIZHEN LI, AND JEFFREY L. FURMAN (2017b): "Firm performance and state innovation funding: Evidence from China's innofund program," *Research Policy*, 46(6), 1142–1161.

WEITZMAN, MARTIN L. (1998): "Recombinant growth," *The Quarterly Journal of Economics*, 113(2), 331–360.

WHALLEY, ALEXANDER AND JUSTIN HICKS (2014): "Spending Wisely? How Resources Affect Knowledge Production in Universities," *Economic Inquiry*, 52(1), 35–55.

WHITLEY, RICHARD (2014): "How do institutional changes affect scientific innovations? The effects of shifts in authority relationships, protected space, and flexibility," in

Organizational transformation and scientific change: The impact of institutional restructuring on universities and intellectual innovation, Emerald Group Publishing Limited, 367–406.

Wuchty, Stefan, Benjamin F. Jones, and Brian Uzzi (2007): "The Increasing Dominance of Teams in Production of Knowledge," *Science*, 316(5827), 1036–1039.

Ytsma, Erina (2017): "Career Concerns in Knowledge Creation," Mimeo.

Appendix for Online Publication

A. Appendix to Section 3: Data

The publication data is from Microsoft Academic (Tang et al., 2008, Sinha et al., 2015).[44] We process the roughly 160m publications in the following steps to arrive at the final prize winner-year panel:

1. We retrieve all publications by matching the last name and first initial of each prize winner (e.g. A Falk for Armin Falk). For some winners we additionally search for different spellings (e.g. maiden names or umlauts).
2. We manually validate which author names actually match the Leibniz Prize winners and keep only matches.
3. All publications containing comments, replies, letters, editorials, errata, and book reviews are dropped, if they are characterized as such in the title.
4. The remaining ca. 60,000 publications are manually checked for further inconsistencies.
5. Each publication is assigned to a prize winner. Some prize winners co-author with each other, each publication is then counted equally for each winner.
6. Years without publications in Microsoft Academic are assigned a publication count of zero.
7. Furthermore, we count publications by quality of the journal the paper is published in. To this end, we rank journals by the average number of forward citations per paper published in the three years prior to the focal year and then split the ensuing ranking into percentiles. We then count publications in the top 1%, top 2%, top 3%, and so on. This cumulative sum deals with the fact that the Leibniz Prize recipients do not publish in all percentiles.

[44] The data can be downloaded from `https://aminer.org/open-academic-graph`, last accessed 08 March 2019. We use version v1.

The resulting publication counts were manually checked with a sample of publication lists on prize winner's websites and were qualitatively similar.

The data on novelty and conventionality is constructed following Uzzi et al. (2013) and Lee et al. (2015) in the following steps:

1. For all publications in Microsoft Academic, all pairwise combinations of cited journals are formed. For example, if a paper references the *American Economic Review*, the *Quarterly Journal of Economics* and the *Journal of Political Economy*, this yields 2 choose 3 combinations: AER and QJE, AER and JPE, QJE and JPE.

2. The frequency of all of these combinations are counted for each year and the commonness of these combinations is calculated according to the following formula for journals j_1 and j_2: $\frac{N_{j_1,j_2,t}}{\frac{N_{j_1,t}}{N_t} \cdot \frac{N_{j_2,t}}{N_t} \cdot N_t}$ where $N_{j_1,j_2,t}$ is the number of times journal j_1 and journal j_2 are referenced together in year t. N_t, $N_{j_1,t}$, and $N_{j_2,t}$ are the number of all journal pairs, the number of journal pairs containing j_1 and the number of pairs containing j_2 in year t, respectively.

3. On the *paper* level this then yields a distribution of these commonness values for all the journal combinations referenced in a given paper

4. The negative logarithm of the tenth percentile of this distribution is then assigned as the novelty score for a paper

5. The logarithm of the median of this distribution is assigned as the conventionality score for a paper

The text similarity of abstracts is calculated in the following steps:

1. We collect all available abstracts from Microsoft Academic. Abstracts are available for around two thirds of all publications in our sample.

2. We remove stop words and stem all words and then construct a document term matrix where each abstract is a document.

3. Next, the document term matrix is tf-idf weighted, i.e. weighted by the following

factor for each term t and document D:

$$\text{tf.idf}(t, D) = \frac{\text{Frequency of term } t \text{ in document } D}{\text{Max. Frequency of a term } t' \text{ in document } D} \cdot \log \frac{\text{Number of Documents}}{\text{Number of Documents with term } t}$$

4. Between each tf-idf vector (i.e. each document in the document-term matrix) we calculate the cosine similarity.

5. For the similarity measure relative to the early stock of publications, all abstracts from relative years -10 to -6 are aggregated into a single document. The similarity of subsequent publications is then calculated relative to this document.

This publication data is complemented by further information on the prize winners:

1. Data on birth years, year of Ph.D., field and type of institution at appointment are from Finetti (2010), recipient CVs and the DFG. In case a winner has multiple affiliations, we give preference to research institutes (if e.g. an individual is a director at a Max-Planck institute and also an affiliated professor at a university, we classify her as working at a research institute).

2. Data on the number of individual research grants (*Sachbeihilfen*) is scraped from the DFG's GEPRIS database (`gepris.dfg.de`). Grants are matched to prize winners using last names and first initial.

B. Appendix to Section 4: Further Robustness

In this section, we provide additional evidence for the plausibility of the identifying assumption. Furthermore, our results are not driven by the estimation method or individual prize cohorts. The normalization of publication counts with the number of authors does not change the results substantively and weighting publications with forward citations yields results in line with the proposed mechanism. Lastly, we show that inference is robust to using the wild cluster bootstrap and show results for text similarity, novelty, and conventionality for the placebo, triple difference, and mechanism analyses.

Additional Evidence on the Plausibility of the Identifying Assumption

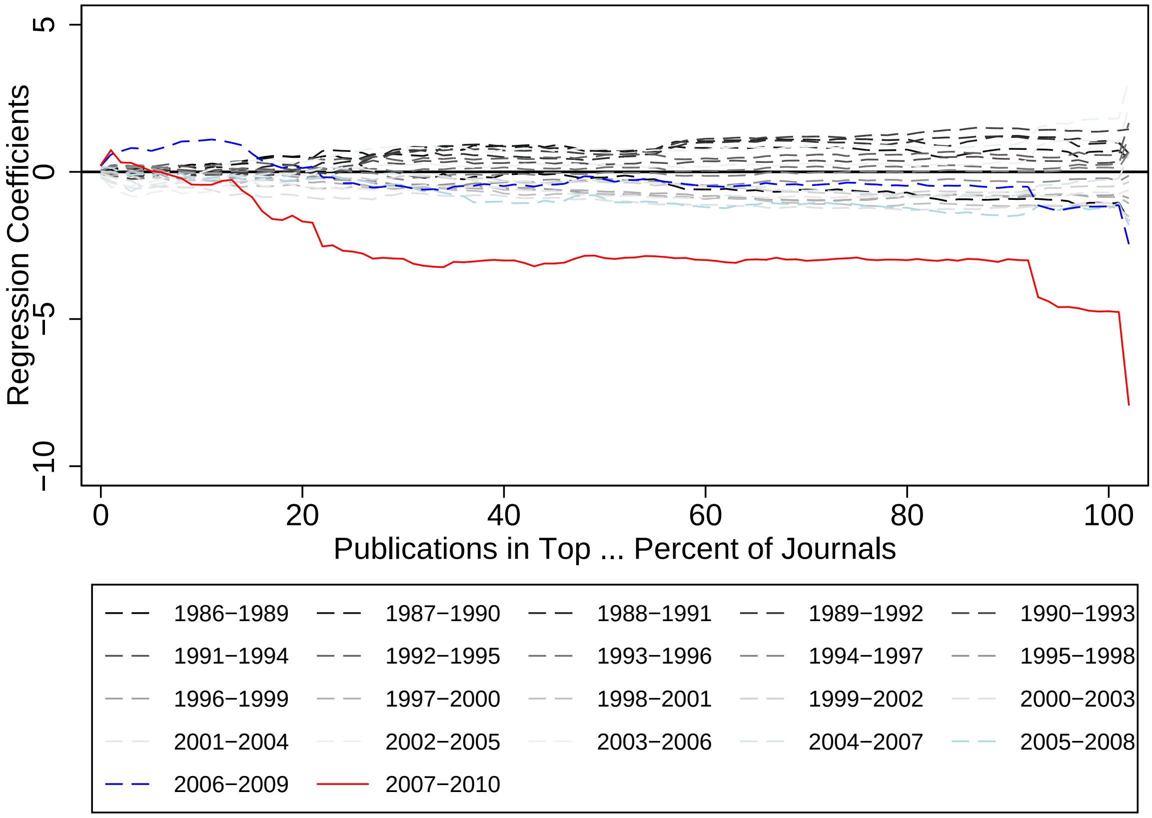
Figure B.1 shows the results from a simple falsification exercise. We re-assign the treatment indicator to each four-year span between 1986 and 2010. Specifically, we first assign treatment to the Leibniz Prize winners from 1986 to 1989 and re-estimate equation (1) for publication counts differentiated by quality. Next, we assign treatment to the winners between 1987 and 1990 and use all other winners as controls, etc. The diff-in-diff coefficients for the entire quality distribution for each of these placebo treatments is shown in Figure B.1. It is clearly visible that the actual treatment assignment has by far the largest effects in absolute magnitude and that all placebo regression coefficients are centered closely around zero.

In addition to the time-varying coefficients in Figure 2, Figure B.2 presents the time-varying coefficients for the other dependent variables of interest. To conserve space, we only show coefficients for publications in the top 1 percent of journals and not for the entire distribution. For all dependent variables, almost all coefficients in the period prior to the prize are small and centered around zero. In terms of timing, the increase in the publication in the top 1 percent of journals, is significant only four years after receiving the prize.

The last piece of evidence for the plausibility of the identifying assumption pertains to the prestige or standing of the prize. We proxy prestige by using how interested the general public is into the prize.[45] To measure the interest of the general public,

[45]The Leibniz Prizes are usually covered in national newspapers and the award ceremony is usually attended by a high-ranking representative of the government.

Figure B.1: Falsification Exercise



Note: This figure shows the estimated coefficients of repeatedly estimating equation (1) for the number of publications in the top *z* percent of journals. Journals are ranked according to the average number of citations per paper in the three years prior. For each line, the treatment indicator has been reassigned to a four year interval. For example, for the case of the 1986-1989 coefficients, the "treated" prize recipients are those who received their prize between 1986 and 1989. All other prize cohorts are then considered untreated. In all regressions, we use the weights suggested by Iacus et al. (2012) to identify the average treatment effect on the treated, calculated within each institution type at appointment and broad scientific field stratum. 90 percent confidence intervals are based on standard errors clustered on the year of prize reception.

we use the relative frequency of Google searches for the terms "Leibniz Prize" and "Gottfried-Wilhelm-Leibniz Prize" (in German). Figure B.3 shows these frequencies which are normalized by Google Trends such that they range from zero to 100. The data is monthly from January 2004 to December 2010. In general, there is a strong seasonality in interest, corresponding to the time when the prizes are handed out. Apart from that, there is no discontinuity in the interest of the general public into the Leibniz Prize around the reform in 2007. If anything, the interest seems to slightly decline over time.

Additional Robustness

All publication count variables are by definition count variables. Especially for the publications in top journals, many observations have zeros as publishing in these prestigious journals is fairly rare, even for accomplished researchers such as Leibniz Prize winners. Hence, it may be more appropriate to use count data estimation methods such as a poisson regression or negative binomial regression. We re-estimate our main specification for all count variables using both of these methods in Figure B.4. The overall pattern is very similar to our baseline results with an increase at the top of the distribution and a reduction for the overall number of publications. In terms of magnitude, the results are somewhat smaller. The reduction in the overall number of publications is 11 percent and 24 percent relative to the mean for the negative binomial and the fixed effects poisson model, respectively. The increase in publications in the top 1 percent of journals is around 90 percent relative to the mean for both estimation methods.

Next, we investigate how our results change if we weight publication counts by the number of authors or forward citations. Using forward citations is difficult in our context since the publications of the treatment group have had little time to accrue citations. We focus on forward citations in a three year window following the publication. We drop the last years of the treatment group (relative years 5 to 7 for the 2010 cohort, years 6 and 7 for the 2009 cohort and year 7 for the 2008 cohort) as the publications in these years have not yet had three years to accrue citations.

Panel (a) of Figure B.5 shows the results for the publication counts across the journal quality distribution. The pattern is very similar to the baseline pattern. Furthermore, the coefficient for the number of publications of all types corresponds to a reduction of 65 percent relative to the mean, close to our baseline estimate. Panel (b) shows the results if publication counts are weighted with their forward citations in the three years following publication. While the qualitative pattern is the same as the baseline results, the reduction at the bottom of the quality distribution is no longer statistically significant. However, given that the last prize cohorts have not had enough time to accrue forward citations we drop the last years from the analysis. Figure 2 shows that the reduction in the number of publications gets stronger over time, so we are dropping the years with the largest reduction in the number of publications. Hence, we do not want to over-interpret the finding of no reduction in the overall number of citation-weighted publications.

Last, we investigate whether the results are driven by a specific prize cohort. To this end, each prize cohort is dropped in turn and the main regression for publication counts re-estimated in a leave-one-out fashion. The weights suggested by Iacus et al. (2012) are re-calculated in every iteration. Figure B.6 shows the results of this exercise and plots the corresponding coefficients. The pattern is similar to the baseline pattern irrespective of which cohort we drop and in terms of magnitude the coefficients are also close to each other, indicating that individual cohorts do not drive the results.

Adjusting for Small Number of Clusters

One potential issue in our setting could be the fairly small number of clusters for the calculation of the standard errors. In the main specification we have 25 clusters (prize years), well below the cutoff of 42 recommended by Angrist and Pischke (2008). This issue is exacerbated in the mechanism analysis, as the number of prize years drops to as low as 11. To mitigate any concerns regarding a small cluster bias, we use the wild cluster bootstrap method proposed by Cameron et al. (2008) to deal with exactly this issue. We re-run our analyses using both the restricted and unrestricted wild cluster bootstrap in Figures B.7 to B.9 and for brevity focus on the publication counts. The standard errors get somewhat larger, such that some of the coefficients lose significance on standard levels in the main difference-in-differences specification, while they mostly retain their significance in the triple-differences specification (except at some of the publication counts in top journals in the unrestricted wild cluster bootstrap). Re-assuringly, the increase in top ranked publications and decrease in overall publications remains significant in the mechanism analysis when we focus on an increase in time and money (Panel (c) of Figure B.9). In addition, the confidence intervals are similar using both the restricted and the unrestricted wild cluster bootstrap. This mitigates concerns voiced by MacKinnon and Webb (2017) that the wild cluster bootstrap may be inappropriate with very different cluster sizes or few treated clusters. The former is not an issue here, but the latter may be of more concern a priori. However, MacKinnon and Webb (2017, p.14) conclude that *agreement between WCR [restricted wild cluster bootstrap] and WCU [unrestricted wild cluster bootstrap] seems to rule out really severe errors of inference*, which is the case here.

Additional Outcome Variables

We additionally present the results for the additional outcome variables (text similarity, novelty, and conventionality) not discussed in the main text for the heterogeneity by broad academic field, the placebo exercise, the triple-difference approach, and the mechanism section.

In Table B.1, the coefficient for text similarity relative to the early stock of publications is of similar magnitude and significance as in the main specification across all fields, except for the natural sciences. In the life sciences, there is a significant decrease in the novelty of journal combinations and increase in the conventionality of research. Although there are some significant effects in the placebo exercise using the matched control group (Table B.2), there are no significant effects in the triple differences exercise (Table B.3). Lastly, in the mechanism analysis in Table B.4, the pattern when funding amount and duration are changed is most similar to the baseline pattern, as for the publication counts. The only other significant coefficient is for the text similarity within a given year when only the funding amount is varied.

C. Appendix to Section 6: Grant Applications

We provide additional suggestive evidence that the increase in *truly legendary freedom* of the Leibniz Prize reform allows scientists to spend less time on activities that positively affect their research budget. We focus on other, non Leibniz-Prize grants at the DFG and show that they decline for both groups and appear to do so more strongly

Table B.1: Heterogeneity of Effects by Broad Academic Field

	Text Sim. I	Text Sim. II	Novelty	Conventionality
Post Prize	-0.00	-0.01^{*}	0.10^{**}	-0.01
	(0.00)	(0.00)	(0.05)	(0.03)
Post Prize × Post 2007 ×				
Social Sciences	0.01^{*}	-0.01^{***}	0.29	-0.22
	(0.00)	(0.01)	(0.33)	(0.20)
Engineering	-0.01	-0.03^{**}	0.23	-0.01
	(0.00)	(0.01)	(0.14)	(0.13)
Life Sciences	-0.01	-0.03^{***}	-0.30^{***}	0.20^{***}
	(0.01)	(0.01)	(0.05)	(0.07)
Natural Sciences	-0.00	-0.01	-0.13	0.01
	(0.01)	(0.01)	(0.10)	(0.11)
Fixed Effects	Year	Year	Year	Year
Mean Dep.	0.08	0.13	-0.70	1.87
R^2	0.06	0.39	0.03	0.02
Winners	252	248	256	256
Observations	3536	2536	3974	3974

Note: This table shows the results from a difference-in-differences estimation with ten years before receiving the Leibniz Prize as pre-period and seven years after as post-period. The treatment indicator is interacted with an indicator for each scientific field (social sciences, engineering, life sciences, natural sciences) and the estimation equation is $y_{i,t} = \beta_1 \cdot \text{Post Prize}_t + \beta_{\text{Social Sciences}} \cdot \text{Post 2007}_i \cdot \text{Post Prize}_t \cdot \mathbb{1}\{\text{Field} = \text{Social Sciences}\} + \beta_{\text{Engineering}} \cdot \text{Post 2007}_i \cdot \text{Post Prize}_t \cdot \mathbb{1}\{\text{Field} = \text{Engineering}\} + \beta_{\text{Life Sciences}} \cdot \text{Post 2007}_i \cdot \text{Post Prize}_t \cdot \mathbb{1}\{\text{Field} = \text{Life Sciences}\} + +\beta_{\text{Natural Sciences}} \cdot \text{Post 2007}_i \cdot \text{Post Prize}_t \cdot \mathbb{1}\{\text{Field} = \text{Natural Sciences}\} + \text{Winner FE} + \text{Year FE} + \epsilon_{i,t}$. In the table, the field-specific treatment coefficients are labeled with the respective field. The classification of fields and the mapping of prize recipients to fields is from the DFG. In column (1), the dependent variable is the text similarity of abstracts to each other within a given year. In column (2), text similarity is calculated relative to the early stock of publications of an author. Note that the number of observations drops as not every researcher publishes in every year. In columns (3) and (4), novelty and conventionality are defined as in Lee et al. (2015). In all regressions, we use the weights suggested by Iacus et al. (2012) to identify the average treatment effect on the treated, calculated within each institution type at appointment and broad scientific field stratum. Standard errors in parentheses are clustered on the level of the year of the prize. *, ** and *** denote significance on the 10 percent, 5 percent and 1 percent level, respectively.

Table B.2: Effects on Text Similarity and Novelty – Placebo Exercise

	Text Sim. I	Text Sim. II	Novelty	Conventionality
Post Prize	-0.00	-0.00	0.09^{**}	-0.07^{**}
	(0.01)	(0.00)	(0.03)	(0.03)
Post Prize × Post 2007	-0.04	0.01^{**}	0.08^{**}	-0.09^{*}
	(0.04)	(0.00)	(0.03)	(0.05)
Fixed Effects	Year	Year	Year	Year
Mean Dep.	0.17	0.05	-0.71	1.83
R^2	0.06	0.04	0.00	0.01
Winners	1779	1525	1750	1750
Observations	22231	15145	20689	20689

Note: This table shows the results from a difference-in-differences estimation within the sample of placebo Leibniz Prize recipients, with ten years before receiving the placebo Leibniz Prize as pre-period and seven years after as post-period. The estimation equation is as in equation (1). The dependent variable in column (1) is the average similarity of abstracts to each other within a given year. In column (2), the pre-period is limited to five years prior to prize reception and the dependent variable is the average similarity of abstracts to the abstracts of papers published between ten and six years prior to the prize. In columns (3) and (4), the dependent variables are novelty and conventionality, respectively, as defined by Lee et al. (2015). In all regressions, we use the weights suggested by Iacus et al. (2012) to identify the average treatment effect on the treated, calculated within each broad scientific field stratum. Standard errors in parentheses are clustered on the level of the year of the prize. *, ** and *** denote significance on the 10 percent, 5 percent and 1 percent level, respectively.

for the treatment prize cohorts.

The DFG is the main source of research funding in Germany, spending around €3 billion in 2017.[46] We scrape information on traditional individual research grants (*Sachbeihilfen*) from the DFG's GEPRIS database. It contains information on all grants of the DFG since the early 2000s.[47] We focus on individual research grants as they are quantitatively important (accounting for one third of the DFG's budget) and it is at the sole discretion of the individual researcher whether or not to apply for a grant.[48] The data encompasses the title of the project, the applicant's name, and the duration of the project. Unfortunately, we do not observe the grant amounts nor grant applications and hence focus only on successful grant applications.[49]

[46] `http://www.dfg.de/dfg_profil/zahlen_fakten/statistik/programmbezogene_statistiken/index.html`, last accessed on 08 August 2018.

[47] The data can be found at `http://gepris.dfg.de/gepris/OCTOPUS`, last accessed on 07 August 2018.

[48] The other main funding lines of the DFG are larger scale joint efforts such as collaborative research centers, excellence clusters, or graduate schools. These programs usually require multiple professors to apply and may span multiple institutions.

[49] This is not an issue as long as the success probability is the same across the two groups of Leibniz Prize winners. There is no indication that this would not be the case, as any "Matthew effect" where Leibniz Prize recipients find it easier (or harder) to receive a grant should affect both groups

Table B.3: Effects on Text Similarity and Novelty – Triple Differences

	Text Sim. I	Text Sim. II	Novelty	Conventionality
Post Prize	0.01**	0.00***	0.05	−0.04
	(0.01)	(0.00)	(0.03)	(0.03)
Post Prize × Post 2007	−0.11*	−0.00	0.05	−0.01
	(0.05)	(0.00)	(0.07)	(0.05)
Post Prize × Leibniz	−0.13***	−0.04***	0.09*	−0.07**
	(0.02)	(0.01)	(0.05)	(0.03)
Post Prize × Post 2007 × Leibniz	0.02	−0.00	−0.09	−0.03
	(0.03)	(0.01)	(0.08)	(0.06)
Fixed Effects	Year	Year	Year	Year
Mean Dep.	0.17	0.05	−0.62	1.77
R^2	0.05	0.09	0.01	0.01
Winners	3070	2687	3011	3011
Observations	39914	27572	37919	37919

Note: This table shows the results from a difference-in-difference-in-differences estimation with ten years before receiving the Leibniz Prize as pre-period and seven years after as post-period. Additional differences are between Leibniz Prize recipients and placebo prize winners and between prize cohorts before and after 2007. The year of prize reception for placebo prize winners is assigned based on their age. The treatment indicator, Leibniz Prize indicator, and post 2007 prize reception indicator are all taken up by the individual scientist fixed effects. The estimation equation is as in equation (2). The dependent variable in column (1) is the average similarity of abstracts to each other within a given year. In column (2), the pre-period is limited to five years prior to prize reception and the dependent variable is the average similarity of abstracts to the abstracts of papers published between ten and six years prior to the prize. In columns (3) and (4), the dependent variables are novelty and conventionality, respectively, as defined by Lee et al. (2015). In all regressions, we use the weights suggested by Iacus et al. (2012) to identify the average treatment effect on the treated, calculated within each broad scientific field, gender, and year of birth stratum. Standard errors in parentheses are clustered on the level of the year of the prize. *, ** and *** denote significance on the 10 percent, 5 percent and 1 percent level, respectively.

Table B.4: Mechanism: Text Similarity and Novelty

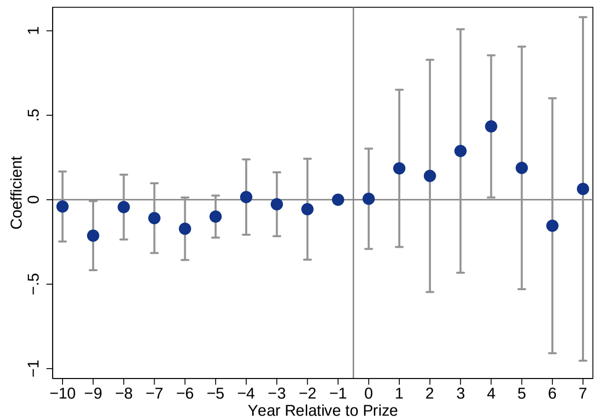
	(1) Text Sim. I	(2) Text Sim. II	(3) Novelty	(4) Conventionality
Panel A: 2000 to 2006 vs. 1986 to 1992 Cohorts – Varying Funding Amount				
Post Prize	-0.01^{*}	-0.00	0.11	0.00
	(0.00)	(0.00)	(0.09)	(0.05)
Post Prize × Pre 1992	0.02^{*}	0.00	-0.05	0.01
	(0.01)	(0.01)	(0.15)	(0.08)
Fixed Effects	Year	Year	Year	Year
Mean Dep.	0.08	0.13	-0.73	1.86
R^2	0.06	0.44	0.03	0.03
Winners	143	139	146	146
Observations	2017	1349	2254	2254
Panel B: 1986 to 1992 vs. 2007 to 2010 Cohorts – Varying Funding Duration				
Post Prize	0.00	-0.01	0.04	0.00
	(0.01)	(0.00)	(0.12)	(0.06)
Post Prize × Post 2007	-0.01	0.00	-0.05	0.07
	(0.02)	(0.01)	(0.15)	(0.13)
Fixed Effects	Year	Year	Year	Year
Mean Dep.	0.08	0.14	-0.74	1.85
R^2	0.06	0.50	0.04	0.03
Winners	110	107	113	113
Observations	1531	946	1717	1717
Panel C: 2000 to 2006 vs. 2007 to 2010 Cohorts – Varying Funding Amount and Duration				
Post Prize	-0.00	0.01	0.11	0.02
	(0.00)	(0.01)	(0.08)	(0.06)
Post Prize × Post 2007	-0.00	-0.03^{***}	-0.07	0.02
	(0.01)	(0.01)	(0.09)	(0.10)
Fixed Effects	Year	Year	Year	Year
Mean Dep.	0.08	0.14	-0.63	1.92
R^2	0.10	0.42	0.04	0.05
Winners	103	102	103	103
Observations	1614	1227	1737	1737

Note: This table shows the results from a difference-in-differences estimation with ten years before receiving the Leibniz Prize as pre-period and seven years after as post-period. The estimation equation is as in equation (1). The dependent variable in column (1) is the average similarity of abstracts to each other within a given year. In column (2), the pre-period is limited to five years prior to prize reception and the dependent variable is the average similarity of abstracts to the abstracts of papers published between ten and six years prior to the prize. Note that the number of observations drops as not every researcher publishes in every year. In columns (3) and (4), the dependent variables are novelty and conventionality, respectively, as defined by Lee et al. (2015). In Panel A, we re-assign treatment to the prize recipients from 1986 to 1992 and use the recipients from 2000 to 2006 as control group. In Panel B we use the prize recipients from 1986 to 1992 as control group, whereas in Panel C we use only the cohorts from 2000 to 2006 as control group (the treatment group remains the same). In all regressions, we use the weights suggested by Iacus et al. (2012) to identify the average treatment effect on the treated, calculated within each institution type at appointment and broad scientific field stratum. Standard errors in parentheses are clustered on the level of the year of the prize. *, ** and *** denote significance on the 10 percent, 5 percent and 1 percent level, respectively.

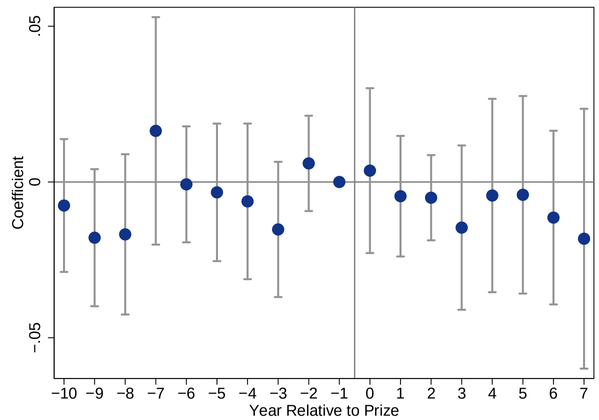
We scrape all available grants (around 60,000) and merge them to the Leibniz Prize recipients via their names. Due to the limited coverage of the data on the post 2000 period, we can only study the prize cohorts from 2004 to 2010 to observe some years prior to prize reception. Figure C.10 shows the three year moving average of the number of active grants for the two groups of Leibniz Prize winners from four years prior to the prize to seven years after. One can see that around two years after prize reception, the number of grants in both groups drops and continues to decline for the post 2007 group. However, for the 2004 to 2006 prize cohorts, the number picks up again five years after receiving the prize, right around the time when the Leibniz funding ends for this group. This suggests that the prize recipients do substitute regular grants with the Leibniz funding and that this effect is stronger for the group with a larger funding amount.

in the same fashion.

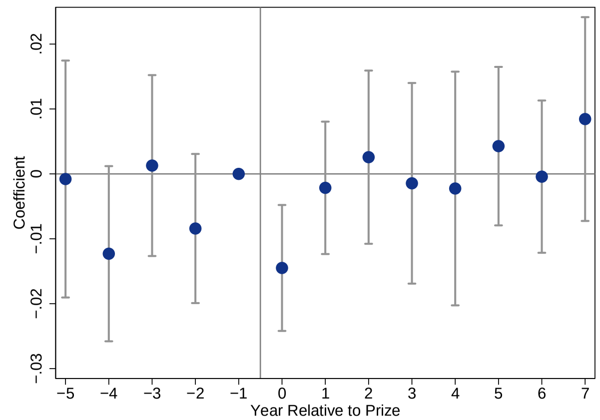
Figure B.2: Time-varying Coefficients – Other Dependent Variables



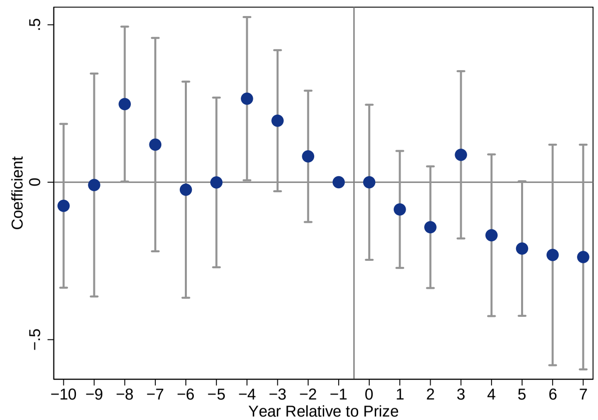
(a) Publications in top 1%



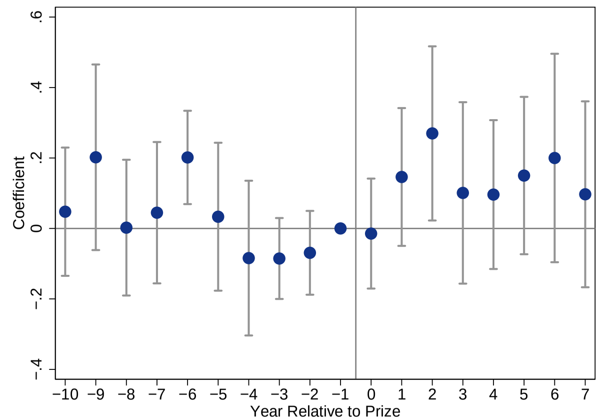
(b) Text Similarity I



(c) Text Similarity II



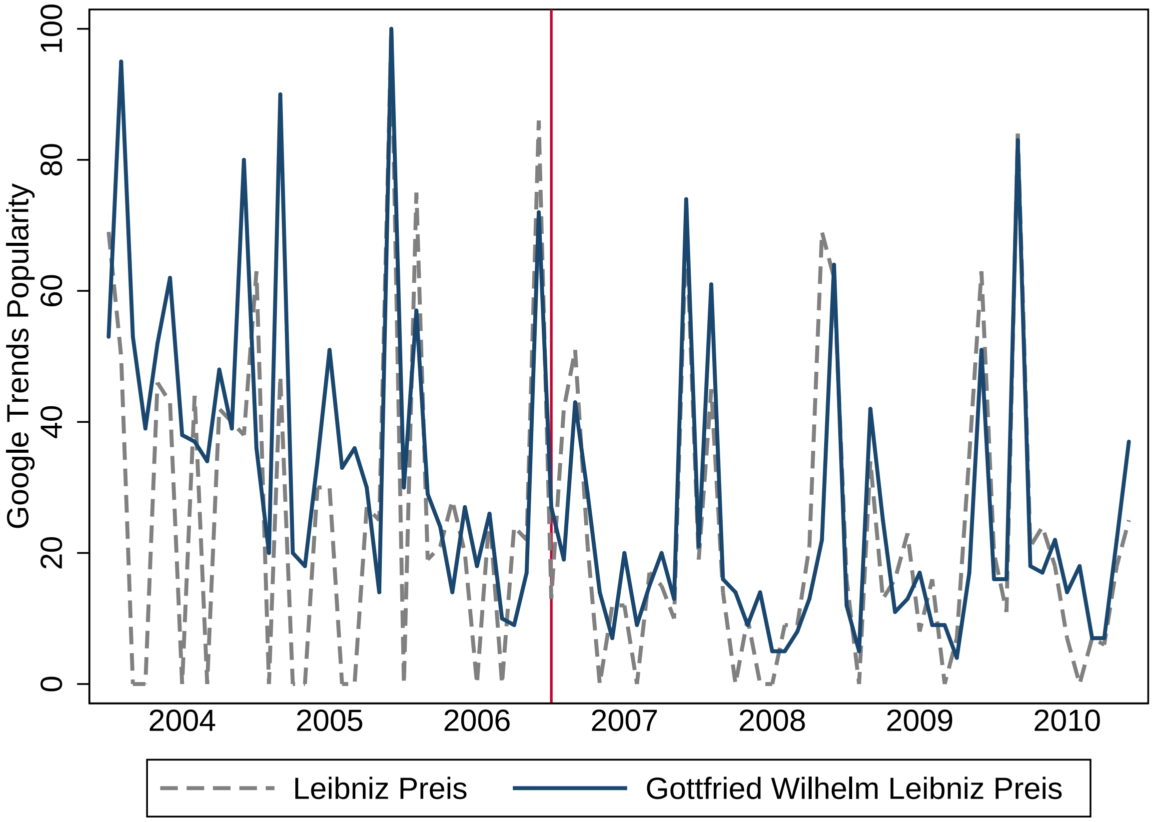
(d) Novelty



(e) Conventionality

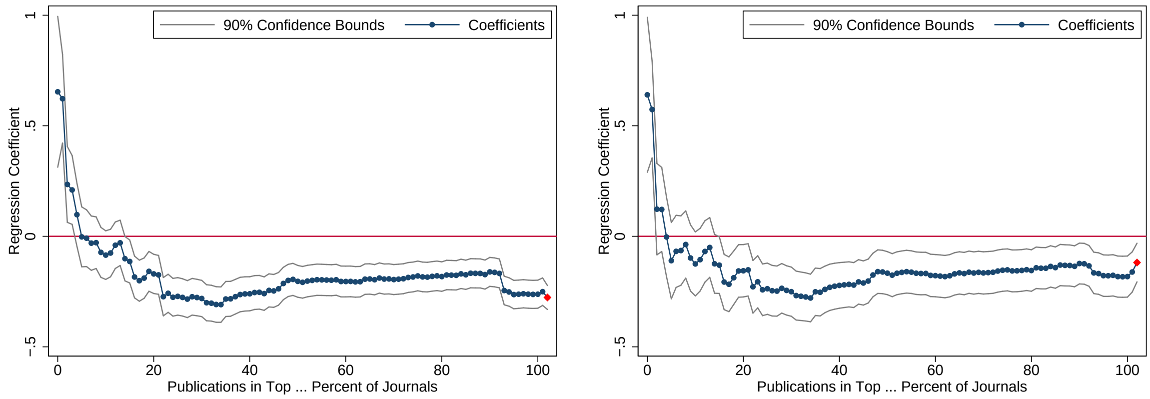
Note: This figure shows the yearly average treatment effects on the treated of receiving the Leibniz Prize in 2007 or later on a range of outcome variables relative to receiving the Leibniz Prize in 2006 or prior. In panel (a), the dependent variable is the number of publications in the top 1 percent of journals. In panel (b), the dependent variable is the average text similarity of abstracts within a given year. In panel (c), the dependent variable is the average text similarity of abstracts to the early stock of publications. As the early stock of publications is computed for ten years prior to the prize up to six years prior to the prize, the pre-period is shorter for this variable. Lastly, in panels (d) and (e) the dependent variables are novelty and conventionality, as defined by Lee et al. (2015), respectively. In all regressions, we use the weights suggested by Iacus et al. (2012) to identify the average treatment effect on the treated, calculated within each institution type at appointment and broad scientific field stratum. 95 percent confidence intervals are based on standard errors clustered on the year of prize reception.

Figure B.3: Google Search Frequencies



Note: This figure shows the popularity of the search terms "Leibniz Preis" and "Gottfried Wilhelm Leibniz Preis" on Google Trends. The data is monthly and the red line indicates the reform of 2007.

Figure B.4: Count Data Models

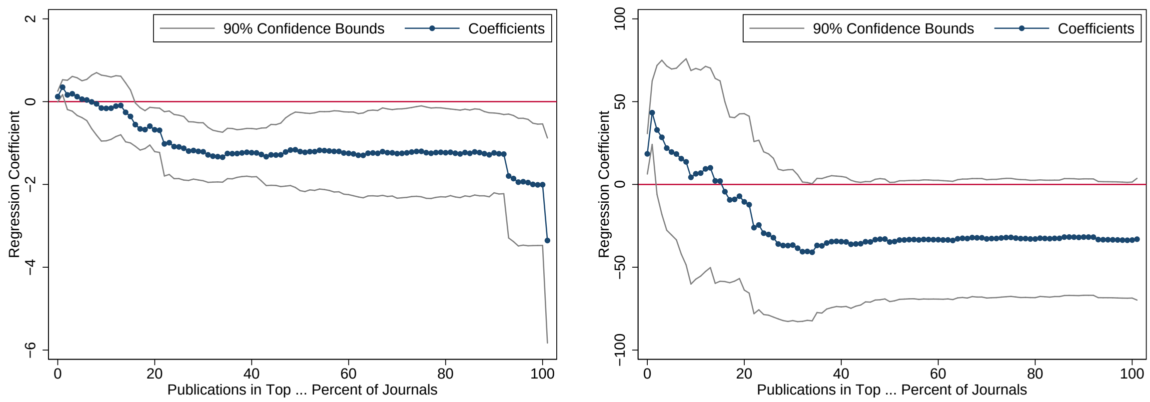


(a) Fixed Effects Poisson

(b) Negative Binomial Model

Note: This figure shows the estimated coefficients of repeatedly estimating equation (1) for the number of publications in the top z percent of journals. Journals are ranked according to the average number of citations per paper in the three years prior. In Panel (a), a fixed effects poisson model is estimated and in Panel (b) a negative binomial model. In all regressions, we use the weights suggested by Iacus et al. (2012) to identify the average treatment effect on the treated, calculated within each institution type at appointment and broad scientific field stratum. 90 percent confidence intervals are based on standard errors using the observed information matrix.

Figure B.5: Weighted Dependent Variables

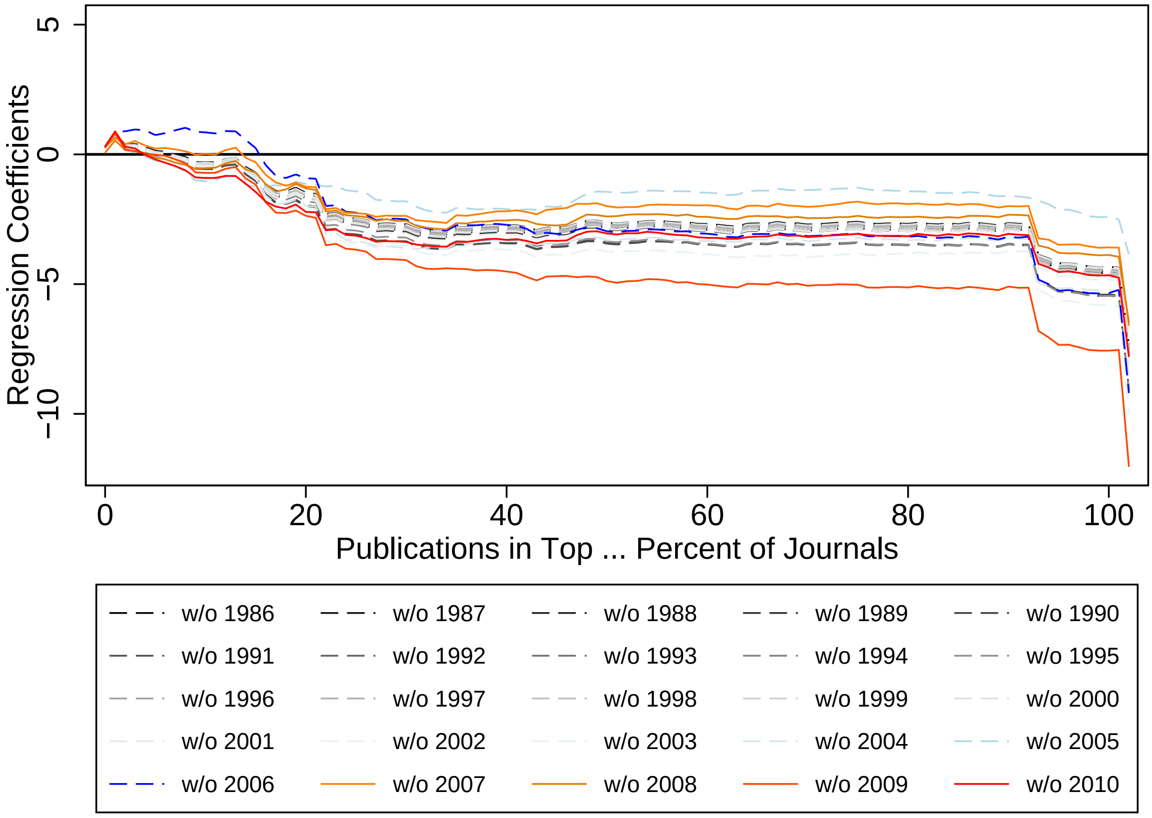


(a) Normalized with $\sqrt{\text{No. of authors}}$

(b) Weighted with three-year forward citations

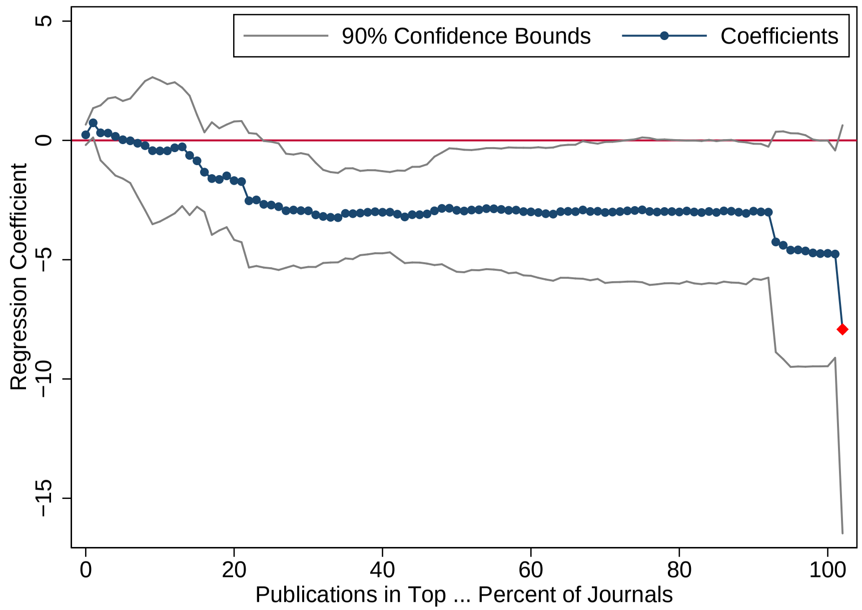
Note: This figure shows the estimated coefficients of repeatedly estimating equation (1) for the number of publications in the top z percent of journals. Journals are ranked according to the average number of citations per paper in the three years prior. In Panel (a), all publication counts are normalized with the square root of the number of authors. In Panel (b), all publication counts are weighted with the number of citations received in the three years following publication. Note that we drop the last observations in the treatment group as these publications have not had enough time to garner forward citations. In all regressions, we use the weights suggested by Iacus et al. (2012) to identify the average treatment effect on the treated, calculated within each institution type at appointment and broad scientific field stratum. 90 percent confidence intervals are based on standard errors clustered on the year of prize reception.

Figure B.6: Leave-one-out: Dropping Individual Prize Cohorts

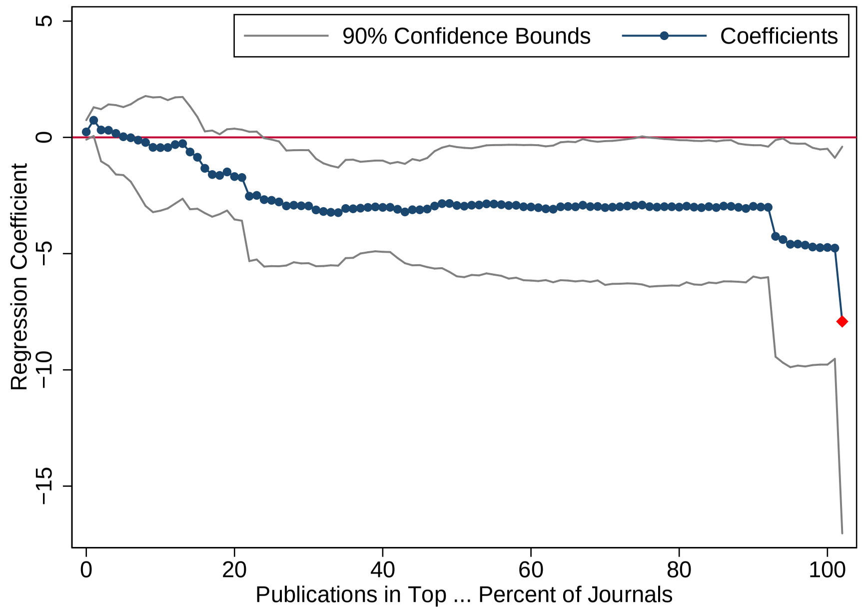


Note: This figure shows the estimated coefficients of repeatedly estimating equation (1) for the number of publications in the top z percent of journals, dropping each prize cohort one by one. Journals are ranked according to the average number of citations per paper in the three years prior. For each line, one prize cohort has been dropped and the weights of Iacus et al. (2012) are re-calculated within each institute type at appointment and broad scientific field stratum to arrive at the average treatment effect on the treated. 90 percent confidence intervals are based on standard errors clustered on the year of prize reception.

Figure B.7: Robustness of Inference I: Main Results



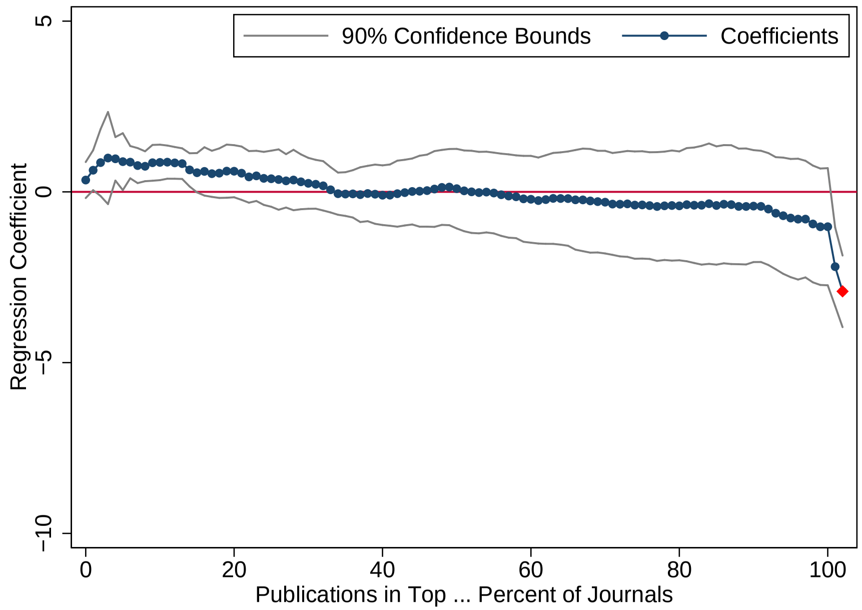
(a) Unrestricted Wild Cluster Bootstrap



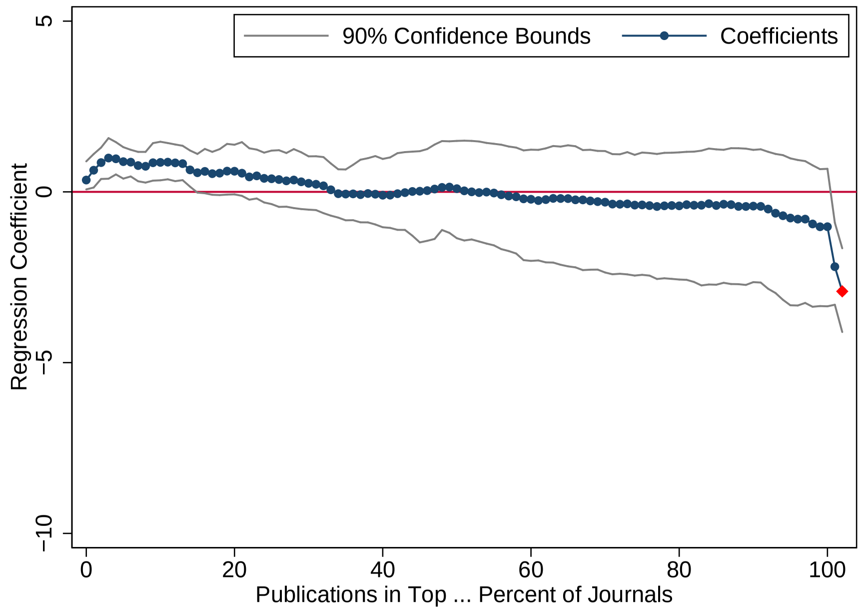
(b) Restricted Wild Cluster Bootstrap

Note: This figure shows the estimated coefficients of repeatedly estimating equation (1) for the number of publications in the top z percent of journals. Journals are ranked according to the average number of citations per paper in the three years prior. In Panel (a), inference is based on the unrestricted wild cluster bootstrap of Cameron et al. (2008) and in Panel (b) on the restricted wild cluster bootstrap. In all regressions, we use the weights suggested by Iacus et al. (2012) to identify the average treatment effect on the treated, calculated within each institution type at appointment and broad scientific field stratum.

Figure B.8: Robustness of Inference II: Triple Difference



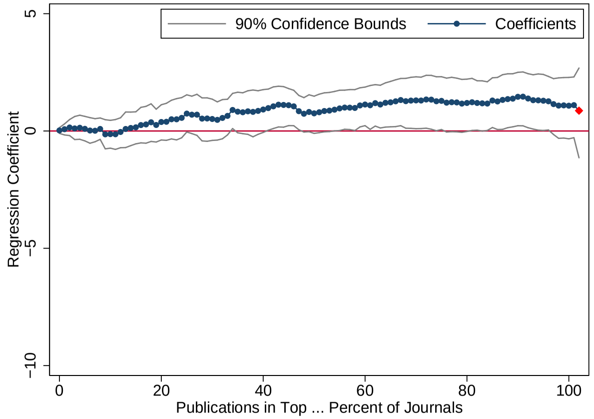
(a) Unrestricted Wild Cluster Bootstrap

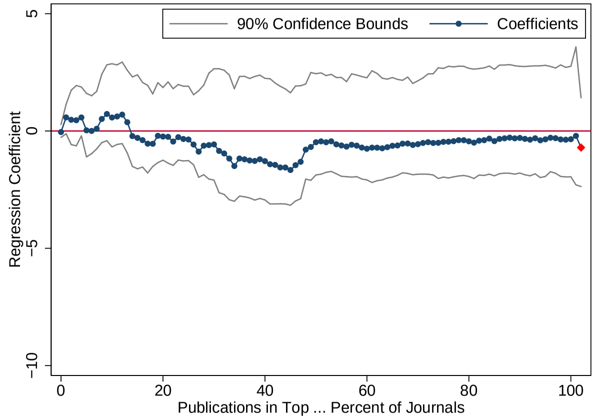


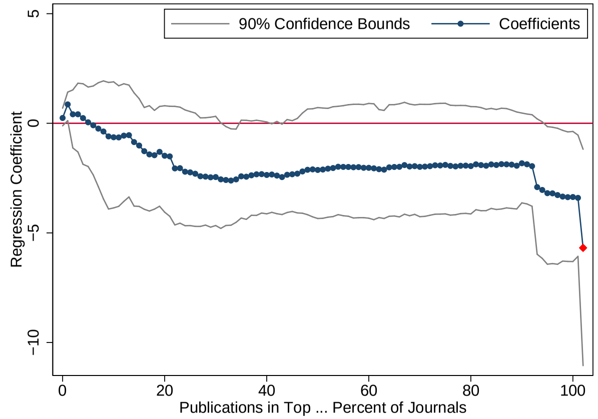
(b) Restricted Wild Cluster Bootstrap

Note: This figure shows the estimated coefficients of repeatedly estimating equation 2 for the number of publications in the top z percent of journals.Journals are ranked according to the average number of citations per paper in the three years prior. In Panel (a), inference is based on the unrestricted wild cluster bootstrap of Cameron et al. (2008) and in Panel (b) on the restricted wild cluster bootstrap. In all regressions, we use the weights suggested by Iacus et al. (2012) to identify the average treatment effect on the treated, calculated within broad scientific field, gender, and year of birth stratum.

Figure B.9: Robustness of Inference III: Mechanism



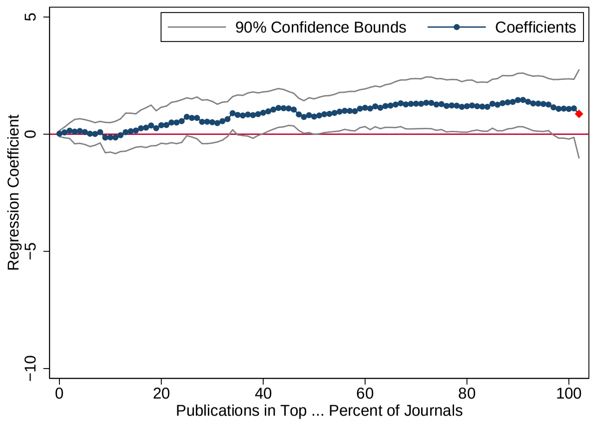


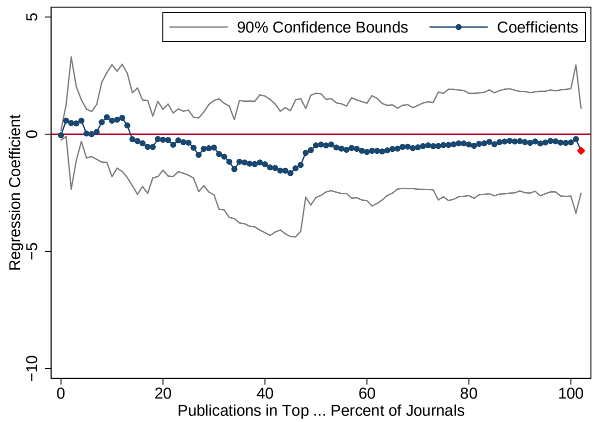


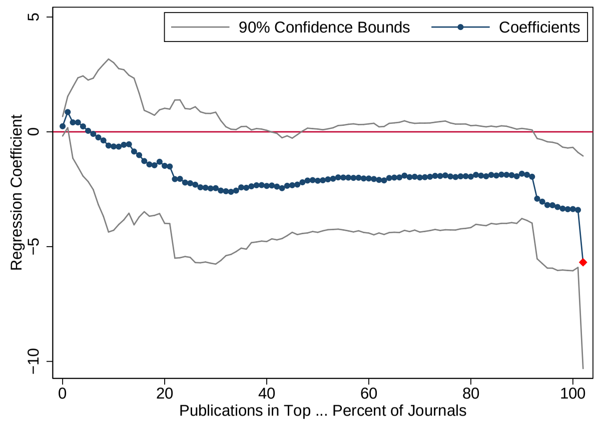
(a) Only Money

(b) Only Time

(c) Time and Money

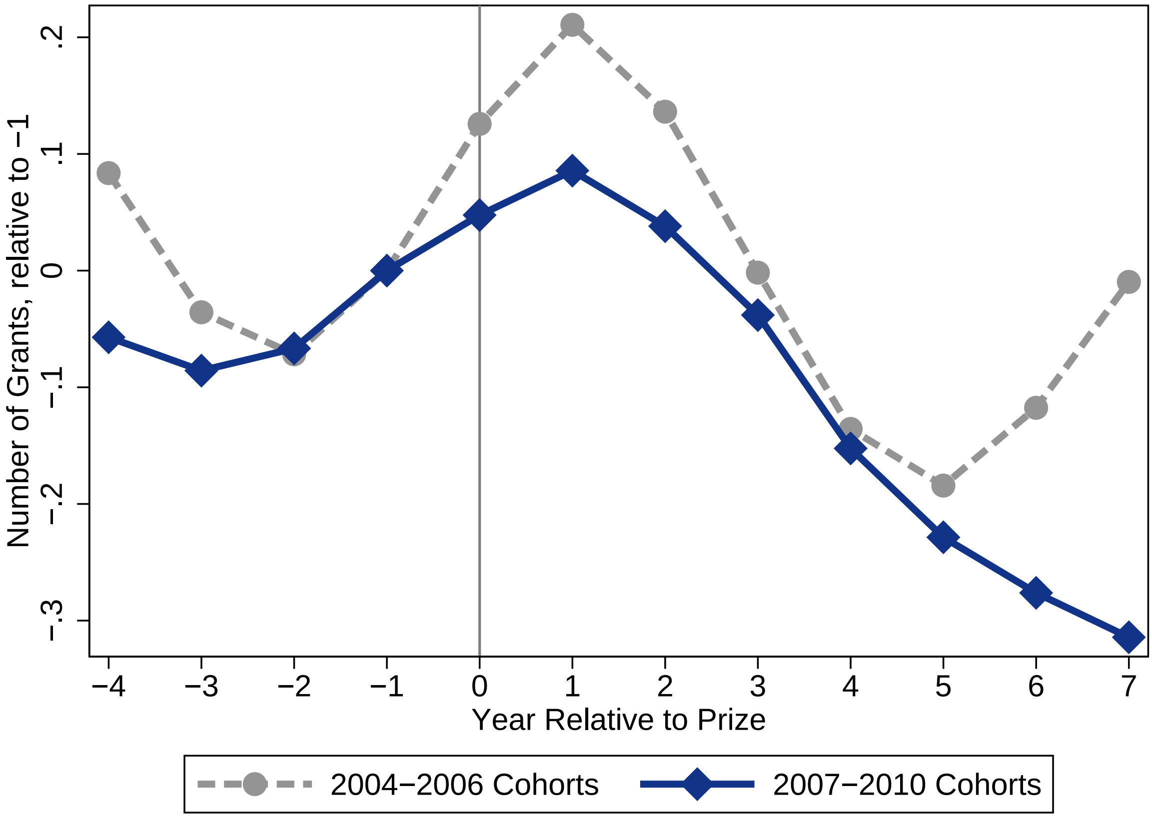






Note: This figure shows the estimated coefficients of repeatedly estimating equation (1) for the number of publications in the top z percent of journals. Journals are ranked according to the average number of citations per paper in the three years prior. In Panels (a), we re-assign treatment to the prize recipients from 1986 to 1992 and use the recipients from 2000 to 2006 as control group. In Panels (b) we use the prize recipients from 1986 to 1992 as control group, whereas in Panels (c) we use only the cohorts from 2000 to 2006 as control group (the treatment group remains the same). In the top row, inference is based on the restricted wild cluster bootstrap of Cameron et al. (2008) and in the bottom row on the unrestricted wild cluster boostrap. In all regressions, we use the weights suggested by Iacus et al. (2012) to identify the average treatment effect on the treated, re-calculated within each institution type at appointment and broad scientific field stratum.

Figure C.10: Effect of the Leibniz Prize Reform on the Number of Other Grants (Non-Parametric Evidence)



Note: Three-year moving average of the number of active grants per researcher for the Leibniz Prize winners from 2004 to 2006, relative to the year before prize reception. Averages are weighted with the weights suggested by Iacus et al. (2012), calculated within each institution type at appointment and broad scientific field stratrum. Only traditional individual research grants (*Sachbeihilfen*) are counted. The data on grants is taken from the DFG's GEPRIS online database. Relative years -4 and 7 are not averaged.