Ideas and Jobs

An Assessment of Public Policies to Promote Science, Innovation, and Employment

---

Inaugural-Dissertation
zur Erlangung des Grades
Doctor oeconomiae publicae (Dr. oec. publ.)
an der Ludwig-Maximilians-Universität München

2019

vorgelegt von
Matthias Wilhelm

Referentin:	Prof. Dr. Monika Schnitzer
Koreferent:	Prof. Dietmar Harhoff, Ph.D.
Promotionsabschlussberatung:	24. Juli 2019

Ideas and Jobs

An Assessment of Public Policies to Promote Science, Innovation, and Employment

---

Inaugural-Dissertation
zur Erlangung des Grades
Doctor oeconomiae publicae (Dr. oec. publ.)
an der Ludwig-Maximilians-Universität München

2019

vorgelegt von

Matthias Wilhelm

Referentin:	Prof. Dr. Monika Schnitzer
Koreferent:	Prof. Dietmar Harhoff, Ph.D.
Promotionsabschlussberatung:	24. Juli 2019

Tag der mündlichen Prüfung: 11. Juli 2019
Namen der Berichterstatter: Monika Schnitzer, Dietmar Harhoff, Oliver Falck

Für Cathrin

Acknowledgements

At times, pursuing a doctorate can feel like a solitary endeavor. However, I was lucky to have been accompanied and supported along the way by many people.

First and foremost, I would like to thank my supervisor Monika Schnitzer. I am deeply grateful for her guidance, motivation, and unwavering support throughout my bachelor's degree, my master's degree, and now my doctoral work. My dissertation and my academic thinking have benefited tremendously from her intellectual rigor. I am indebted to her for much of what I take away from these years of study. Furthermore, I want to thank my second advisor Dietmar Harhoff for his guidance, support, and the cooperation at the Munich Center for Internet Research. I would also like to thank Oliver Falck for completing my dissertation committee and for always being helpful and supportive.

I am very grateful to Martin Watzinger, without whom this dissertation would never have been possible. His mentoring and guidance inspired me to learn more and think deeper about economics. I am also grateful to Lukas Buchheim for all that I have learned from his rigor. I want to thank Anna Gumpert, whose thoughtful advice and sincere support were instrumental in many moments.

My colleagues at the Munich Graduate School of Economics, at the Evidence-Based Economics Program, and especially at the Seminar for Comparative Economics made graduate school much nicer. I have fond memories of the time spent together in courses, at the office, and over lunch. Henrike Steimer and Felix Montag were outstanding colleagues and dear friends. Markus Nagler deserves special mention as a faithful companion through all these years. His open ear and shared interests beyond economics were invaluable. I want to thank Dagmar Erhardt, Karin Fritsch, Ines Steinbach, Julia Zimmermann, and the student assistants at the Seminar for Comparative Economics for their outstanding help.

During my dissertation, I spent several months at MIT Sloan. I would like to thank Antoinette Schoar for making this visit possible and the Golub Center for Finance and Policy for providing me with a desk, a tremendous view of Boston, and fun moments at the office. Moreover, I am indebted to the research data center at the Bavarian Statistical Office for providing me with data for parts of this dissertation and their hospitality during my research visits. I am also grateful for the financial support from the DAAD and the Elite Network of Bavaria through the doctoral program Evidence-Based Economics. Florian Englmaier and Joachim Winter have done an outstanding job in leading the program.

Without my friends, my parents Marga and Adi, and my sister Hannah, I would not have been able to achieve any of this. More importantly, life would have been much less enjoyable. I am deeply grateful for having you. My greatest gratitude goes to Cathrin and Benedict – your love and support have been invaluable.

I owe it all to you.

Matthias Wilhelm, July 2019

Contents

List of Tables

List of Figures

Introduction

Not for the first time, as an elected official, I envy economists. Economists have available to them, in an analytical approach, the counterfactual. [...] They can contrast what happened to what would have happened. No one has ever gotten reelected where the bumper sticker said, "It would have been worse without me." You probably can get tenure with that. But you can't win office.

Former U.S. Representative Barney Frank

Two key policy objectives for governments around the world are to increase the production of new ideas and to create jobs. Whilst there is broad consensus on the goals themselves, the means to achieve them are less agreed upon. Governments of different political convictions have enacted various policies that are too numerous to count and the amount of public money spent on them is staggering. For example, in order to increase the production of new ideas, the public sector in Germany spent around €30 billion or 1 percent of GDP on research and development activity in 2016 (Statistisches Bundesamt, 2018). To save and create jobs in wake of the Great Recession, Germany and other countries enacted large fiscal stimulus programs. In Germany, close to €90 billion were spent and the American Recovery and Reconstruction Act in the US amounted to more than $800 billion (Bundesministerium für Finanzen, 2009; Congressional Budget Office, 2014).

However, whether this money is well spent and which of these policies actually work is far from clear. One of the main reasons why we know so little is that it is very difficult to approximate what would have happened in absence of the policy. In the case of public subsidies for research and development, selection issues abound. Both academic scientists and private firms need to apply for grants and subsidies, and applications are then selected on their merit. Hence, we do not fully understand whether public funding for scientists affects their productivity or whether subsidies for firms cause additional innovation (Goolsbee, 1998, Zúñiga-Vicente et al., 2014). Similar problems arise in the evaluation of fiscal stimulus policies, as worse hit regions may be disproportionately targeted by stimulus spending. For example, prior to the passage of the Recovery Act in the United States, estimates of the employment effects ranged from 1.2m to 3.6m additional jobs (Congressional Budget Office, 2009). Even ex post, the estimated employment effects in the literature differ by a factor of seven (Conley and Dupor, 2013, Chodorow-Reich et al., 2012).

This thesis exploits three quasi-experiments in Germany to further our understanding of which policies are successful at increasing science, innovation, and employment. Natural quasi-experiments allow us to construct counterfactuals and are now a standard part of the applied economist's toolkit (Angrist and Pischke, 2010). Each chapter of this thesis studies a different setting and can be read on its own. The first chapter shows that the structure and amount of funding for academic scientists has a first order effect on their productivity. The second chapter concludes that subsidizing research and development projects of private firms does not necessarily increase their amount of research and development spending. Last, the third chapter demonstrates that counter-cyclical stimulus measures and place-based policies may be a viable tool to stimulate employment.

The first chapter studies how the amount and duration of a research grant affects the scientific productivity of elite scientists. Basic scientific research would not be viable without public funding and the design of funding schemes is a key lever for governments to incentivize scientists (Stephan, 2012). However, we know fairly little about the productivity effects of various funding design choices, such as the amount or duration of funding. The main issue is that 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.

To circumvent such selection issues, we look at the Gottfried Wilhelm Leibniz Prize, Germany's most important and most prestigious research prize. Each year, around ten elite scientists from all disciplines receive this prize which comes with a large, non-renewable research grant which they can use at their full discretion. In 2007, the size of this research grant was increased by €1m and the time over which the money could be spent was lengthened by two years. This allows us to compare the recipients of the *same* prize who are exposed to a *different* funding amount and duration. We compare the scientific output of Leibniz Prize recipients who received their prize after the reform to the output of prize recipients prior to the reform in a difference-in-differences framework.

The analysis shows that the Leibniz Prize recipients who received a larger grant amount and longer grant duration reduce their overall number of publications per year by 53 percent. This *reduction* in the overall quantity of publications is accompanied by an *increase* in the number of publications in top ranked journals. For both top ranked multidisciplinary journals and top ranked field journals, the increase is around 50 percent relative to the respective mean number of publications. This effect cannot be explained by simple time trends in the life-cycle productivity profile of academics.

Additional analyses investigate whether it is the additional grant amount or grant duration that matters. By exploiting the fact that there is substantial variation in the amount of funding in real terms due to inflation, we show that additional grant amount and duration seem to be complements. The pattern of a reduction in overall publications and an increase in top-ranked research is only present when both the grant amount and the grant duration are larger. There is little to no response to an increase in funding amount or duration alone.

The chapter contributes both to the literature on the effects of the amount of funding and to the literature on the structure of funding. It is among the first to show the complementarity of funding amount and funding duration, which has implications for the design of science funding policies. Furthermore, this paper is unique in its within-prize comparison. This alleviates remaining concerns about selection on unobservables that the prior literature using matching on observables cannot rule out. Also, most papers comparing recipients of a prize or grant to non-recipients cannot discern whether it is the money or the prestige of the prize or grant that affects productivity. In the context of the Leibniz Prize, any prestige effects or "Matthew effects" where prize winners find it easier to publish their research should be constant across the two groups, isolating the causal effect of additional funding amount and duration.

The second chapter, which is based on joint work with Monika Schnitzer and Martin Watzinger, changes its focus from the knowledge creation by academic scientists to how research and development activities of firms can be incentivized. Due to the public goods nature of new ideas, the private sector may invest less in research and development than is socially optimal. In light of this fact, many governments subsidize knowledge creation of private companies, for example via subsidies for research and development projects (EY, 2018). However, it is unclear whether these subsidies have the desired effect. The subsidy may simply "crowd out" the private investment a firm would have conducted even in absence of the subsidy. In addition, as firms usually need to apply for a subsidy and applications are selected on their merit, any comparison of subsidy recipients to non-recipients is fraught with selection issues.

To circumvent these issues, we study a change in the eligibility criteria for the main research and development subsidy scheme in Germany. The *Zentrales Innovationsprogramm Mittelstand* (ZIM) has a budget of around €500m per year to subsidize research and development projects of small and midsize German firms. As part of the second stimulus package in wake of the Great Recession in 2009, the eligibility criteria were changed and firms with up to 1000 employees were allowed to apply, in contrast to prior to the reform when only firms with fewer than 250 employees could apply.

This change in eligibility criteria can be used in a regression discontinuity design. Under the assumption that the firms just above and below the threshold of 1000 employees are comparable, we can estimate the causal effect of being eligible to apply for the ZIM. This circumvents the selection issues that are inherent in a simple comparison of ZIM recipients and non-recipients. We use administrative data on the German manufacturing sector that is ideally suited for this purpose, as it includes research and development expenditures and a wide range of further firm covariates.

We find a large positive effect of being newly eligible for the ZIM on subsequent research and development expenditures. However, these differences already exist prior to the change in the eligibility criteria, implying that firms with fewer than 1000 employees already had higher research and development expenditures prior to the reform. We address this issue using a difference-in-discontinuity design which requires any differences between the two groups of firms to be constant over time. Under this assumption, we do not find any significant effects of ZIM eligibility on firm outcomes. However, the confidence intervals are very wide; hence, we cannot evaluate the effectiveness of the ZIM in a definitive fashion. Yet, a power analysis for a randomized controlled trial evaluation of the ZIM suggests that such an evaluation would be feasible within the budget currently appropriated for the ZIM.

The chapter is the first quasi-experimental evaluation of the ZIM. It contributes to the literature on measuring the effect of research and development subsidies on firm outcomes in three ways. First, we are, to our knowledge, the first to use administrative data on German manufacturing to study the effects of research and development subsidies. Second, by studying the largest German subsidy program with a difference-in-discontinuity design, we can interpret our estimates as causal. Lastly, we can trace out the effects of the ZIM on research and development inputs and firm outcomes, allowing us to assess the effects of ZIM in a comprehensive fashion.

The last chapter, which is based on joint work with Lukas Buchheim and Martin Watzinger (Buchheim et al., forthcoming), assesses whether it is easier for governments to create employment in times and places of high unemployment compared to times and places with low unemployment. Specifically, it investigates whether the fiscal multiplier is state-dependent. Assessing the state-dependence of the multiplier is challenging for three reasons: First, identifying the multiplier requires exogenous demand shocks. Second, these shocks need to vary sufficiently within each state of the economy. Otherwise there is potentially not enough statistical power to estimate the multiplier for each state. Third, the shocks have to be comparable in their composition across the different states.

We assess the state dependence of the multiplier in the context of investment in photovoltaics in Germany. This is an ideal setting as photovoltaic installations constituted an exogenous, frequent, and comparable demand shock in all states of the economy. First, the variation in photovoltaic installations over time and across space was mainly driven by factors that are exogenous to the economic circumstances. Second, there is ample identifying variation in these investments for any partition of counties into groups with slack and tight labor markets, as the installation of rooftop photovoltaic systems was profitable in all German regions. Third, the composition of investment has been constant as each photovoltaic installation of a given size constituted the same demand shock.

Our main finding is that the installation of photovoltaic systems created at least twice as many jobs in slack labor markets characterized by high unemployment than in tight labor markets with low unemployment. Our preferred specifications compare job creation in tight and slack markets using two splits: First, along the time series dimension, we compare counties at times when their unemployment rate is above or below its long-run average. Second, along the cross-sectional dimension, we compare counties with high or low unemployment relative to their state average in a given year. When we use our estimates to approximate a local labor earnings multiplier, we arrive at a multiplier of 1.1 in slack markets and below 0.5 in tight markets. These results are robust to various alternative ways of classifying slack and tight markets, and remain qualitatively unchanged when we instrument photovoltaic installations with their profitability as measured by the investments' net present value.

Furthermore, we make progress on identifying the mechanism underlying the state dependence of the effects. The evidence is most consistent with crowding-out, i.e. that in tight markets workers are drawn from other jobs whereas they are drawn from non-employment in slack markets. Consistent with this channel, there is evidence that investments lead to additional wage growth in the construction sector in tight markets, but not in slack markets.

This chapter contributes to the literature in several ways. First, it is only the second paper to investigate potential channels for state-dependence. In line with the evidence from a cross-country setting in Auerbach and Gorodnichenko (2013), we identify crowding-out as the most plausible mechanism. Second, this chapter is unique in ruling out price and composition effects as potential confounding factors for state-dependent multipliers, an issue neglected by the literature thus far. Third, we demonstrate the robustness of our results with respect to a wide array of different ways of classifying slack and tight markets.

Taking the three chapters of this thesis together, they contribute to our knowledge of which policies are successful at increasing science, innovation, and employment. This knowledge can in turn be used as a foundation for the evidence-based design of future policies. To date, Germany is lagging other countries in this domain, both in the evaluation of past policies and in the use of scientific evidence in the design of future policies (Boockmann et al., 2014; Buch et al., 2019). This thesis addresses the former and may hopefully contribute to the latter.

1

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

1.1 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 chapter 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.

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).

[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).

Hence, the grant can be spent at the full discretion of the recipient. 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 use two separate measures for top ranked journals: top ranked multidisciplinary journals (Nature, Science, PNAS, Nature Communications) and the top three journals for each scientific discipline. 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, conventional, or potentially impactful 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 by 53 percent relative to the mean, or by 5.62 publications per year. This reduction in the overall quantity of publications is accompanied by an increase in the number of publications in top ranked journals. For both top ranked multidisciplinary journals and top ranked field journals, the increase is around 50 percent relative to the respective mean number of publications. 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. However, there is an increase in the number of publications with high novelty and high conventionality (p-value 0.12).[2] Since there are only few such potential high impact publications, their increase does not affect the overall means of novelty and conventionality. These results are similar across all scientific fields (engineering, life sciences, natural sciences, social sciences) under study.

A triple differences specification shows that our results are not driven by time-varying shocks differentially affecting the treatment cohorts relative to the control cohorts. Researchers who have won a prestigious early career prize, but did not receive a Leibniz Prize, form an additional control group. These early career prize winners should be affected by the same concurrent shocks as the Leibniz Prize winners, allowing us to difference out any effects of, e.g., the introduction of ERC grants.[3] The results from this exercise are in line with our baseline estimate.

We can rule out many prominent alternative explanations for our 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

[2]These publications have been shown by Uzzi et al. (2013) to be more likely to have high impact.

[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.

the reform.

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 duration. 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.

We only find significant effects in the last comparison, indicating 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 sheds light on the question: "What is the effect of increasing both the funding amount and the period over which the funds could be spent, given that the recipient has full discretion to spend her grant?" In doing so, it speaks to two strands of the literature, which have focused on two separate, but related questions.[4] 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.[5] 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 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 chapter 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 discre-

[4]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).

[5]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.

tion in spending funds. Azoulay et al. (2011) look at the effects of becoming a Howard-Hughes Medical Institute Investigator (HHMI) on subsequent research productivity.[6] 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 chapter 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.[7]

The remainder of this chapter is structured as follows. Section 1.2 describes the Leibniz Prize and the reform of 2007 in detail. Section 1.3 explains measurement, data and identification. Section 1.4 presents the results and Section 1.5 discusses the mechanism. Section 1.6 concludes.

[6] 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.

[7] This chapter 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).

1.2 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*.[8] 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 chapter, 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.[9] 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

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

[9] 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.

projects (Finetti, 2010). This reform was announced publicly on 30 May 2006 by the DFG, prior to the communication of decisions for the 2007 prize in December 2006.[10] 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.[11] 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

[10] 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.

[11] 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).

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 eminent scientists.[12] 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.[13] 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.[14]

1.3 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.1. Second, we present the identification strategy and evidence for the validity of the identifying assumption.

[12] 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.

[13] 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).

[14] 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.

1.3.1 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; Lee2015; 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. 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, focusing on two measures. We count all publications in the four most cited multidisciplinary science journals (*Science, Nature, Proceedings of the National Academy of Sciences of the United States of America, Nature Communications*) to measure *hit publications*. In addition, we use a broader definition of top journals by counting *publications in the top three journals* by field according to the 2006 ranking of Scimago Lab.[15]

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.[16] 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

[15] The ranking can be found here: `https://www.scimagojr.com/journalrank.php?year=2006`, last accessed 02 August 2018. A list of these journals is in Appendix A.1. We do not use a time-varying journal ranking as the Scimago Ranking only goes back to 1999 and choose the last ranking prior to the reform.

[16] 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.

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 abstracts and taking the average.[17] For the second measure, we treat all early stock 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.[18] They find that papers that score high on novelty and conventionality are more likely to be im-

[17]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, $similarity = \frac{\sum_{i=1}^{n} A_{1,i}A_{2,i}}{\sqrt{\sum_{i_1}^{n} A_{1,i}^2}\sqrt{\sum_{i_1}^{n} A_{2,i}^2}}$.

[18]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.

pactful, 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.[19] 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 using the 10$^{\text{th}}$ 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$^{\text{th}}$ percentile as a measure of the *conventionality*. In addition, we classify publications by their *potential impact* by counting only those publications with high novelty and high conventionality.[20]

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

[19]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.

[20]Specifically, a paper is classified as having high novelty if it is below the 10$^{\text{th}}$ percentile of the year specific distribution of novelty and as having high conventionality if it is above the median of the year specific distribution of conventionality.

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]

1.3.2 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 five 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 five 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 from 1986 to 2010:

$$\begin{aligned} y_{i,t} = \beta_1 \cdot \text{Post Prize}_t + \beta_2 \cdot \text{Post 2007}_i \cdot \text{Post Prize}_t \\ + \text{Winner FE} + \text{Year FE} + \epsilon_{i,t} \end{aligned} \tag{1.1}$$

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

[21] Appendix A.2 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 a robustness check in Table A.3, we extend the pre period to 10 years prior to the prize and find similar results.

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 A.2.

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. This untestable assumption appears plausible, since the selection mechanism and criteria of the Leibniz Prize did not change over time.

The plausibility of this assumption is underlined by the fact that although there are differences in levels, the two groups have parallel trends prior to receiving their respective Leibniz Prizes. Summary statistics for all main variables prior to receiving the Leibniz Prize can be found in Table 1.1. 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 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

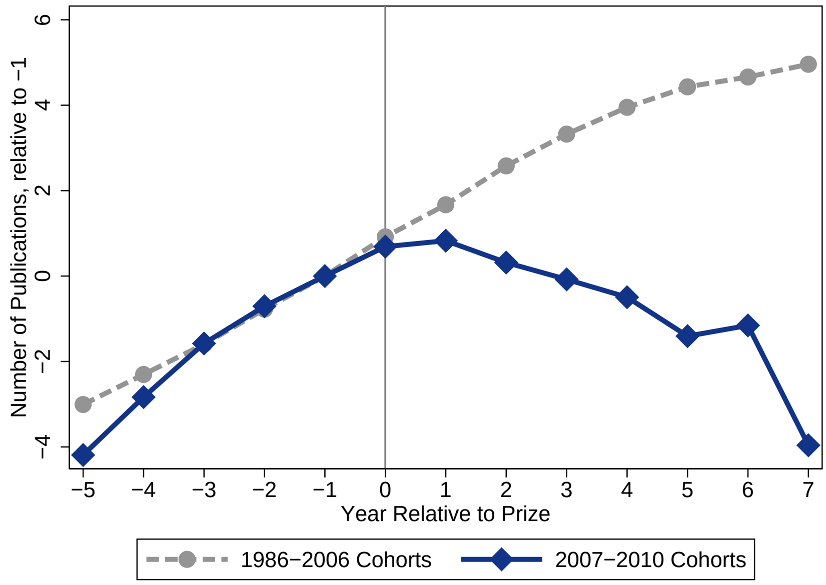
[23] Note that the baseline effect of *Post 2007* is taken up by the winner fixed effects. In addition, in robustness checks we include field by year, affiliation type by year, and gender by year fixed effects (see Table 1.3).

Table 1.1: Summary Statistics Prior to Prize

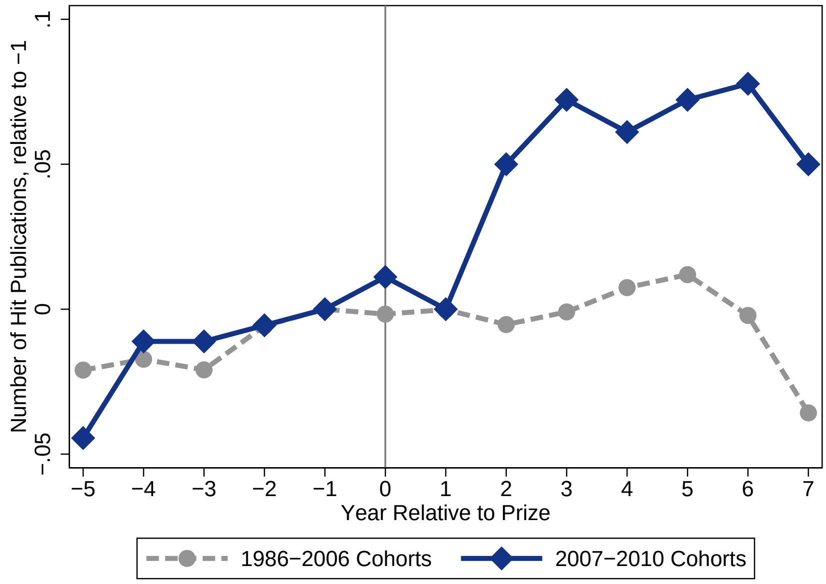
	1986-2006	2007-2010	Difference	
Age at Prize	45.20	45.03	0.18	(0.81)
Age at PhD	27.76	27.92	−0.16	(0.64)
Female	0.07	0.19	−0.13*	(0.07)
University	0.78	0.56	0.23**	(0.01)
Social Sciences	0.06	0.08	−0.02	(0.62)
Engineering	0.17	0.19	−0.02	(0.75)
Life Sciences	0.30	0.28	0.02	(0.80)
Natural Sciences	0.47	0.44	0.03	(0.77)
Number of authors per pub	3.33	4.08	−0.75***	(0.00)
Number of publications per year	7.74	13.72	−5.97***	(0.00)
Science/ Nature / PNAS	0.12	0.19	−0.06	(0.21)
Number of Top 3 pubs per year	0.47	0.70	−0.23	(0.11)
Number of pubs with potential impact	0.04	0.03	0.00	(0.91)
Text similarity within year	0.19	0.32	−0.13***	(0.00)
Text similarity rel. to early pubs	0.16	0.19	−0.02***	(0.01)
Average Novelty	1.53	2.11	−0.58**	(0.02)
Average Conventionality	−0.47	−0.80	0.32	(0.22)
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.

Figure 1.1: Effect of the Leibniz Prize Reform on Scientific Productivity (Non-Parametric Evidence)



(a) No. of Publications (all types)



(b) No. of Hit Publications

Note: Five year moving average number of publications (all types in Panel (a) and hit publications in Panel (b)) per year, relative to the year of Leibniz Prize reception. Means within groups are calculated using the weights of Iacus et al. (2012). Values are normalized with respect to relative year -1. Data for relative years 6 and 7 is not averaged.

academia, an issue we revisit below.[24] In terms of our outcome variables, there are differences in levels for the overall publication count, the text similarity of abstracts, and the average novelty prior to receiving the prize.

Figure 1.1 investigates whether trends are parallel prior to receiving the Leibniz Prize. We focus on the overall number of publications and the number of hit publications as our main dependent variables.[25] The figure shows the average number of publications per year for the two groups of prize winners, normalized to the year prior to receiving the prize. In order to minimize noise, we show a five year moving average (where the last two years are not averaged anymore) of the means.[26] As can be seen in Panel (a) for the overall number of publications, prior to receiving the Leibniz Prize, the two groups behave similarly and are virtually indistinguishable in the three years prior to the prize. Foreshadowing our treatment effect, the counts quickly diverge after receiving the prize. For the number of hit publications in Panel (b), the two groups publish on the same trend prior to the reception of the prize. About two years after the prize, the treatment group increases its number of hit publications.

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. To deal with these concerns, we estimate a demanding triple differences specification in Section 1.4.2. This incorporates winners of early career prizes who are also subject to the ERC shock as an additional control group. We find comparable results.

[24]See e.g. Wuchty et al. (2007) for evidence on increasing team sizes.
[25]Results for the other dependent variables are available from the author upon request.
[26]In Figure A.9 in the Appendix, the same data is shown without the moving average adjustment. The pattern is somewhat more noisy, especially for the number of hit publications.

1.4 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.

1.4.1 Main Findings

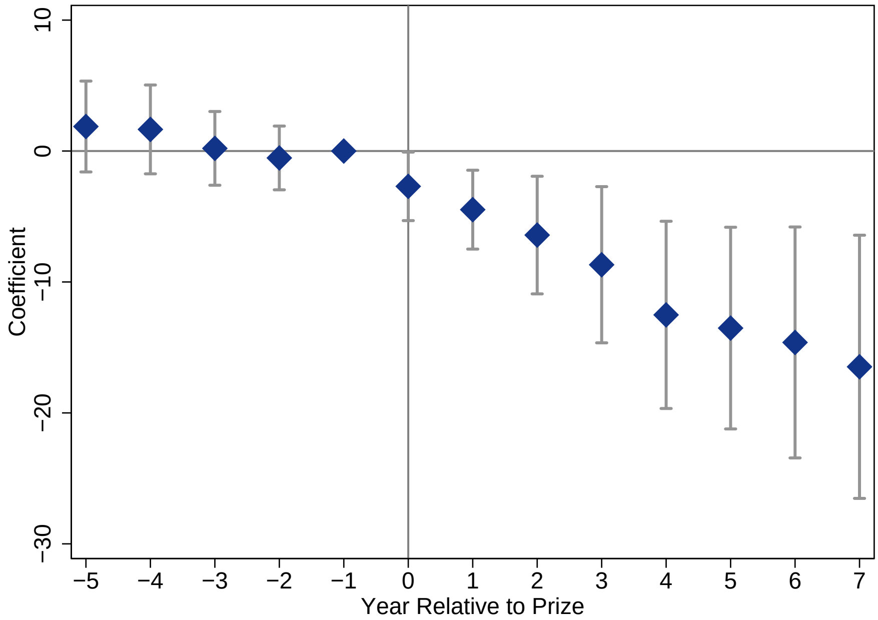
Table 1.2 presents the main results using the baseline specification from equation (1.1). In the first column, we see that prize winners after 2007 write 5.62 fewer publications than those receiving a Leibniz Prize before 2007, relative to before receiving the Leibniz Prize. This effect is large in relative terms, corresponding to a 53 percent reduction relative to the mean. In columns (2) and (3), we see that this decrease in the overall quantity of publications is accompanied by an increase in the number of publications in highly ranked journals. Irrespective of whether only publications in top multidisciplinary journals (Nature, Science, PNAS, Nature Communications) are counted or field-specific top three publications, we see a relative increase of around 50 percent (statistically significant on the 10 percent level). Columns (4) and (5) show that for both measures of text similarity of the abstracts there is a decrease in similarity. However, this is only statistically significant for the measure relative to the early stock of publications. The treatment group of Leibniz Prize winners engages in research more distant from their early work than the control group, after receiving the Leibniz Prize. This effect is smaller in relative terms than the effect for publication counts, with a 15 percent reduction in similarity. Last, we turn to the measures of novelty, conventionality, and potential high impact publications of Uzzi et al. (2013) and

Table 1.2: Effect of the Leibniz Prize Reform on Scientific Productivity (Diff-in-diff Estimates)

	(1) All Pubs	(2) Top Multidisc. Pubs	(3) Top Field Pubs	(4) Abstract Sim. I	(5) Abstract Sim. II	(6) Novelty	(7) Convent.	(8) High Impact Pubs
Post Prize	1.57**	−0.03	−0.08	0.02	−0.01	0.15***	−0.08*	0.01
	(0.59)	(0.03)	(0.09)	(0.02)	(0.00)	(0.05)	(0.04)	(0.02)
Post Prize × Post 2007	−5.62**	0.08*	0.29*	−0.06	−0.02***	−0.13	0.15	0.13
	(2.24)	(0.04)	(0.16)	(0.04)	(0.01)	(0.10)	(0.09)	(0.08)
Mean Dep.	10.64	0.15	0.54	0.27	0.13	1.88	−0.69	0.05
R^2	0.15	0.03	0.03	0.10	0.39	0.06	0.05	0.02
Winners	257	257	257	252	248	256	256	257
Observations	3341	3341	3341	2719	2536	2964	2964	3341

Note: This table shows the results from a difference-in-differences estimation with five years before receiving the Leibniz Prize as pre-period and seven years after as post-period. The estimation equation is as in equation (1.1). The dependent variable in column (1) is the count of all types of publication per year. In column (2) only publications in top general interest journals (Science, Nature, PNAS, Nature Communications) are counted. Column (3) uses publications in field specific top 3 journals (See Table A.2 for the full list of journals). In column (4), the dependent variable is the text similarity of abstracts to each other within a given year. In column (5), 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 (6) and (7), novelty and conventionality are defined as in Lee et al. (2015). Lastly, in column (8), the dependent variable is the count of publications with potential high impact, i.e. having high conventionality and novelty. In all regressions, we use the weights suggested by Iacus et al. (2012) to identify the average treatment effect on the treated. 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.

Figure 1.2: 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. 95 percent confidence intervals are based on standard errors clustered on the year of prize reception. I use the weights of Iacus et al. (2012) to arrive at the average treatment effect on the treated.

Lee et al. (2015). Whereas the abstract similarity is measured within a scientist, novelty and conventionality are defined relative to science as a whole. Neither novelty nor conventionality are statistically significant and the relative magnitude of the point estimates is comparatively small. In terms of sign, the results indicate a reduction in novelty and an increase in conventionality. The number of publications with potential high impact (those with high novelty and high conventionality) increases, although the coefficient is not significant on standard levels (p-value = 0.12).

To investigate the timing of the effect, we estimate yearly treatment effects by interacting the treatment dummy with an indicator for each year before and after the prize.[27] For brevity, we once again focus on the overall number of pub-

[27] The estimation equation is: $y_{i,t} = \sum_{\tau=-5}^{7} \beta_\tau \cdot \text{Post } 2007_i \cdot \mathbb{1}\{t = \tau\} + \text{Winner FE} + \text{Year FE} +$

lications as our main dependent variable.[28] As can be seen in Figure 1.2, 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. After receiving the prize, 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.[29]

Due to the inclusion of winner and year fixed effects, these results cannot be explained by constant differences between the researchers or common shocks to all researchers, such as an increase in the pressure to publish. However, it is evident in the summary statistics that the later cohorts differ with respect to certain demographics from the earlier control cohorts, primarily with the larger share of female prize winners and the lower share of prize winners at universities. It might be that for scientists in specific fields, at research institutes, or female scientists, the increase in the pressure to publish might be smaller. This group-specific time shock could then drive our results. To rule out any shocks that disproportionately affect scientists in specific fields, at research institutes, or female scientists, we use different year fixed effects in Table 1.3. In Panel A, we repeat the baseline results for comparison. In Panel B, we include field (social sciences, natural sciences, engineering, life sciences) by year fixed effects and find similar results. The coefficient for the overall number of publications is somewhat smaller. In Panels C and D, we interact the year dummies with the university and gender dummy, respectively. Results are again similar. Thus, shocks disproportionately

$\epsilon_{i,t}$.

[28]The results for the other dependent variables are available from the author upon request.

[29]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 1.3: Robustness: Including Different Fixed Effects

	(1) All Publications	(2) Top Multidisciplinary Publications	(3) Top Field Publications	(4) Abstract Similarity I	(5) Abstract Similarity II	(6) Novelty	(7) Conventionality	(8) High Impact Publications
Panel A: Baseline								
Post Prize	1.57**	−0.03	−0.08	0.02	−0.01	0.15***	−0.08*	0.01
	(0.59)	(0.03)	(0.09)	(0.02)	(0.00)	(0.05)	(0.04)	(0.02)
Post Prize × Post 2007	−5.62**	0.08*	0.29*	−0.06	−0.02***	−0.13	0.15	0.13
	(2.24)	(0.04)	(0.16)	(0.04)	(0.01)	(0.10)	(0.09)	(0.08)
Panel B: Field × year fixed effects								
Post Prize	1.16**	−0.03	−0.11	0.02	−0.01	0.12**	−0.05	0.00
	(0.47)	(0.02)	(0.07)	(0.01)	(0.00)	(0.05)	(0.04)	(0.01)
Post Prize × Post 2007	−4.16**	0.09**	0.32*	−0.04	−0.02***	−0.06	0.09	0.15*
	(1.63)	(0.04)	(0.16)	(0.03)	(0.00)	(0.11)	(0.09)	(0.08)
Panel C: University / Research Institute × year fixed effects								
Post Prize	1.59***	−0.03	−0.09	0.02	−0.01	0.16***	−0.08*	0.01
	(0.55)	(0.04)	(0.08)	(0.02)	(0.00)	(0.05)	(0.04)	(0.02)
Post Prize × Post 2007	−6.01**	0.08**	0.31*	−0.08	−0.02***	−0.12	0.14	0.12
	(2.87)	(0.04)	(0.16)	(0.05)	(0.01)	(0.10)	(0.10)	(0.08)
Panel D: Gender × year fixed effects								
Post Prize	1.42**	−0.04	−0.10	0.02	−0.01*	0.16***	−0.08*	0.01
	(0.57)	(0.03)	(0.08)	(0.02)	(0.00)	(0.05)	(0.04)	(0.02)
Post Prize × Post 2007	−4.95**	0.08**	0.31	−0.06	−0.02***	−0.09	0.12	0.13
	(2.01)	(0.04)	(0.19)	(0.04)	(0.01)	(0.10)	(0.10)	(0.09)
Mean	10.64	0.15	0.54	0.27	0.13	1.88	−0.69	0.05
Winners	257	257	257	252	248	256	256	257
Observations	3341	3341	3341	2719	2536	2964	2964	3341

Note: This table shows the results from a difference-in-differences estimation with five years before receiving the Leibniz Prize as pre-period and seven years after as post-period. The estimation equation is as in equation (1.1). The dependent variable in column (1) is the count of all types of publication per year. In column (2) only publications in top general interest journals (Science, Nature, PNAS, Nature Communications) are counted. Column (3) uses publications in field specific top 3 journals (See Table A.2 for the full list of journals). In column (4), the dependent variable is the text similarity of abstracts to each other within a given year. In column (5), 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 (6) and (7), novelty and conventionality are defined as in Lee et al. (2015). Lastly, in column (8), the dependent variable is the count of publications with potential high impact, i.e. having high conventionality and novelty. Panel A repeats the baseline specification with year fixed effects. Panel B controls for field (Social Sciences, Life Sciences, Natural Sciences, Engineering) by year fixed effects. Panel C has university or research institute (at prize reception) by year fixed effects and Panel D includes gender by year fixed effects. In all regressions, we use the weights suggested by Iacus et al. (2012) to identify the average treatment effect on the treated. 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.

affecting certain fields, institutions or genders do not drive the results.

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 Table 1.4 we explore heterogeneity across fields by interacting the treatment indicator with an indicator for each field.[30] Focusing on the number of publications in column (1), the effects are similar in engineering, natural sciences, and the life sciences. They are somewhat smaller in the social sciences, which may be due to the fact that journal publications are less important in these disciplines. For the number of publications in top ranked journals (both multidisciplinary and field journals) in columns (2) and (3), the coefficients are all positive and mostly insignificant, but tend to be smaller in the social sciences. In columns (4) and (5), the effect on text similarity is negative across all fields (except for the within year similarity in the social sciences) and similar in magnitude to the baseline estimate. Columns (6) and (7) show a similar pattern for novelty and conventionality, respectively. The social sciences have an opposing sign to the other fields (except the engineering effect on novelty which is positive but very small). Lastly, in column (8), the point estimate for potential high impact publications is positive for all fields, but only significant in engineering. 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 with

[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 does not play a role.

Table 1.4: Heterogeneity of Effects by Broad Academic Field

	(1) All Publications	(2) Top Multidisciplinary Publications	(3) Top Field Publications	(4) Abstract Similarity I	(5) Abstract Similarity II	(6) Novelty	(7) Conventionality	(8) High Impact Publications
Post Prize	1.56**	−0.03	−0.08	0.02	−0.01	0.15***	−0.08	0.01
	(0.59)	(0.03)	(0.08)	(0.02)	(0.00)	(0.05)	(0.04)	(0.02)
Post Prize × Post 2007 ×								
Social Sciences	−3.71	0.07	0.03	0.02	−0.02***	0.23	−0.12	0.47
	(2.28)	(0.05)	(0.24)	(0.06)	(0.00)	(0.40)	(0.25)	(0.36)
Engineering	−6.70**	0.06	0.20**	−0.08	−0.03***	0.02	0.16	0.22*
	(2.66)	(0.08)	(0.08)	(0.07)	(0.01)	(0.17)	(0.12)	(0.11)
Life Sciences	−5.61**	0.01	0.18	−0.10**	−0.04***	−0.26***	0.26***	0.09
	(2.53)	(0.03)	(0.25)	(0.05)	(0.01)	(0.07)	(0.06)	(0.05)
Natural Sciences	−5.49*	0.12	0.43**	−0.05	−0.01	−0.17	0.12	0.07
	(2.84)	(0.07)	(0.16)	(0.05)	(0.01)	(0.13)	(0.10)	(0.05)
Mean	10.64	0.15	0.54	0.27	0.13	1.88	−0.69	0.05
Winners	257	257	257	252	248	256	256	257
Observations	3341	3341	3341	2719	2536	2964	2964	3341

Note: This table shows the results from a difference-in-differences estimation with five 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. The dependent variable in column (1) is the count of all types of publication per year. In column (2) only publications in top general interest journals (Science, Nature, PNAS, Nature Communications) are counted. Column (3) uses publications in field specific top 3 journals (See Table A.2 for the full list of journals). In column (4), the dependent variable is the text similarity of abstracts to each other within a given year. In column (5), 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 (6) and (7), novelty and conventionality are defined as in Lee et al. (2015). Lastly, in column (8), the dependent variable is the count of publications with potential high impact, i.e. having high conventionality and novelty. In all regressions, we use the weights suggested by Iacus et al. (2012) to identify the average treatment effect on the treated. 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.

potential high impact and their number of publications in top ranked journals, both multidisciplinary ones and top field journals. The effects are also large in economic terms, with a decrease in the overall number of publications of around 50 percent and a similarly sized increase in the number of hit publications. This pattern is similar across fields and robust to including different time fixed effects.

1.4.2 Robustness: Triple Difference Relative to Early Career Prize Winners

One remaining concern would be 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.

To deal with these issues, we use a difference-in-differences-in-differences strategy. We employ a third group that does not receive a Leibniz Prize, but is exposed to the same shocks as an additional control group. We construct this group by building on the work by Azoulay et al. (2011) and use early career prize winners as an additional control group. These winners are drawn from a set of prestigious prizes and scholarships across all disciplines.[31] We focus on early career prize winners and not on, e.g., the universe of German academics in order to have a group that is as similar as possible to Leibniz Prize recipients. Most of these early career prize winners go on to a successful career in academia. How-

[31] The full list of early career prizes can be found in Appendix A.1.

ever, they may differ in important (observed and unobserved) dimensions from the Leibniz Prize winners.[32] For example, they are less prolific than Leibniz Prize winners and publish around 40 percent less per year. However, we expect them to be subject to the same shocks created by, e.g., the introduction of ERC grants.[33]

For all 461 early career prize winners, we construct a counterfactual year of Leibniz Prize reception by approximating when the researcher "should have" received a Leibniz Prize based on her age.[34] The full triple differences specification is

$$
\begin{aligned}
y_{i,t} = \beta_1 \cdot \text{Post Prize}_t &+ \beta_2 \cdot \text{Post Prize}_t \cdot \text{Leibniz}_i \\
&+ \beta_3 \cdot \text{Post Prize}_t \cdot \text{Post 2007 Cohort}_i \\
&+ \beta_4 \cdot \text{Post Prize}_t \cdot \text{Leibniz}_i \cdot \text{Post 2007 Cohort}_i \\
&+ \text{Winner FE} + \text{Year FE} + \epsilon_{i,t}
\end{aligned}
\tag{1.2}
$$

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 (1.1).[35] 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 early career prize winners before and after receiving the Leibniz Prize. Due to data constraints, we restrict attention to the prize years from 2000 to 2010.[36] Results for this specification can be found in Table 1.5. Focusing on the coefficient of

[32] Some Leibniz Prize recipients also received an early career prize early in their career. We exclude these Leibniz Prize recipients from the sample of early career prize winners and only include them as Leibniz Prize winners.

[33] Each year there have been around four times as many ERC advanced grants as Leibniz Prizes, making it plausible that also researchers not as successful as Leibniz Prize recipients may respond to the potential incentive effects of the ERC grants.

[34] We do not observe actual age of the early career prize winners. Hence, we use a prize-specific time lag between reception of the early career prize and the counterfactual Leibniz Prize.

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

[36] Many early career prizes are not as long running as the Leibniz Prize. In addition, in the case of the Heisenberg scholarships, recipients prior to 1999 are unavailable due to data protection regulations.

interest (Post Prize $\times$ Leibniz $\times$ Post 2007), the first takeaway is that qualitatively the effects are similar to the baseline specification across all dependent variables. However, only the treatment effects on the overall number of publications and the similarity of abstracts within a year are statistically significant on conventional levels. Reassuringly, though, the relative decrease in the overall number of publications is 55 percent, which is almost identical to our baseline estimate of 53 percent. Hence, time-varying shocks that differentially affect the two cohorts such as the introduction of ERC grants do not seem to be a cause of concern for our identification.

1.4.3 Additional Robustness

Appendix A.2 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 in the baseline specification and only 11 in the triple difference specification. 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 (Tables A.6 to A.8).

As a test of our identifying assumption, we conduct a placebo exercise and assign (placebo) treatments in Figure A.1. The coefficient for the actual treatment is one of the few statistically significant ones and largest in absolute magnitude. Furthermore, we show that extending the pre-prize time period to ten years prior to prize reception (Table A.3 and Figure A.2) or dropping individual prize cohorts (Figures A.3 and A.4) does not change results in a meaningful way. In Table A.5 we investigate how results change if we normalize publication counts by the number of authors or weight publications by their forward citations. Normalizing publication counts with the number of authors does not alter our results. The

Table 1.5: Triple Difference Comparison Relative to Early Career Prize Winners

	(1) All Pubs	(2) Top Multidisc. Pubs	(3) Top Field Pubs	(4) Abstract Simil. I	(5) Abstract Simil II.	(6) Novelty	(7) Convent.	(8) High Impact Pubs
Post Prize	−0.88***	0.00	−0.01	−0.00	0.03**	−0.06	0.05	−0.02
	(0.26)	(0.01)	(0.06)	(0.01)	(0.01)	(0.08)	(0.05)	(0.01)
Post Prize × Post 2007	1.68***	0.00	0.06	0.01	−0.04***	0.08	−0.01	0.02
	(0.28)	(0.01)	(0.06)	(0.01)	(0.01)	(0.11)	(0.07)	(0.02)
Post Prize × Leibniz	3.84***	−0.01	−0.08**	0.03	−0.00	0.21**	−0.09	0.04*
	(1.04)	(0.03)	(0.03)	(0.02)	(0.00)	(0.09)	(0.06)	(0.02)
Post Prize × Leibniz	−4.66**	0.06	0.11	−0.05*	−0.00	−0.23	0.10	0.02
× Post 2007	(1.53)	(0.04)	(0.12)	(0.03)	(0.01)	(0.17)	(0.10)	(0.05)
Constant	5.02***	0.02**	0.14**	0.16***	0.32***	1.66***	−0.52***	−0.01
	(0.42)	(0.01)	(0.06)	(0.01)	(0.01)	(0.06)	(0.04)	(0.01)
Mean	8.50	0.06	0.23	0.22	0.13	2.20	−0.90	0.03
Winners	565	565	565	548	550	558	558	565
R^2	0.07	0.00	0.01	0.03	0.32	0.05	0.06	0.01
Observations	7345	7345	7345	5620	5794	6222	6222	7345

Note: This table shows the results from a difference-in-difference-in-differences estimation with five years before receiving the Leibniz Prize as pre-period and seven years after as post-period. Additional differences are between Leibniz Prize recipients and early career prize winners and between prize cohorts before and after 2007. The year of prize reception for Early Career Prize Winners is assigned by the average difference between receiving an early career prize and receiving the Leibniz Prize. 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 (1.2). The dependent variable in column (1) is the count of all types of publication per year. In column (2) only publications in top general interest journals (Science, Nature, PNAS, Nature Communications) are counted. Column (3) uses publications in field specific top 3 journals (See Table A.2 for the full list of journals). In column (4), the dependent variable is the text similarity of abstracts to each other within a given year. In column (5), 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 (6) and (7), novelty and conventionality are defined as in Lee et al. (2015). Lastly, in column (8), the dependent variable is the count of publications with potential high impact, i.e. having high conventionality and novelty. In all regressions, we use the weights suggested by Iacus et al. (2012) to identify the average treatment effect on the treated. 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.

citation-weighted results have the same qualitative pattern as the main results. However, only the increase in publications in top field journals is statistically significant. Lastly, we investigate further heterogeneity of results by estimating individual treatment effects (Figures A.5 to A.8).

1.5 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. 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.[37] 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.

[37] 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.

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.

Table 1.6 shows the results of this exercise. 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. Except for the similarity of abstracts within a given year, all coefficients are statistically insignificant and most are very small in economic magnitude. For example, the point estimate for a reduction in the overall number of publications is -0.4, much smaller than our baseline estimate of -5.62. Hence, increasing funding alone does not seem to suffice. 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, most coefficients are statistically insignificant (except the number of potential high impact publications) and small in economic magnitude. An increase in the time span to spend a grant of a given size does not have an effect on scientific productivity. 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. Here, we see a decrease in the number of overall publications (38 percent relative to the mean) and an increase in the number of publications in top field journals. The coefficient for the top multidisciplinary is positive, but insignificant on standard levels (p-value = 0.23). This is similar to our baseline results for the entire sample and indicates that it is indeed the combination of additional funding amount and funding period that matters.[38]

1.6 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 A.3 that the treatment group has fewer other, non Leibniz Prize grants from the DFG compared to the control group.

[38]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.

Table 1.6: Mechanism: Funding Amount and Funding Duration

	(1) All Publications	(2) Top Multidisciplinary Publications	(3) Top Field Publications	(4) Abstract Similarity I	(5) Abstract Similarity II	(6) Novelty	(7) Conventionality	(8) High Impact Publications
Panel A: 2000 to 2006 Prize Cohorts vs. 1986 to 1992 Prize Cohorts – Varying Funding Amount								
Post Prize	0.57	0.01	−0.14	−0.03	−0.00	0.06	0.03	−0.01
	(0.87)	(0.03)	(0.15)	(0.02)	(0.00)	(0.08)	(0.04)	(0.04)
Post Prize × Pre 1992	−0.40	−0.05	0.13	0.06**	0.00	0.01	−0.10	−0.01
	(1.19)	(0.06)	(0.21)	(0.03)	(0.01)	(0.13)	(0.08)	(0.05)
Mean	10.77	0.16	0.55	0.27	0.13	1.81	−0.65	0.06
Observations	1911	1911	1911	1554	1349	1684	1684	1911
Panel B: 1986 to 1992 Prize Cohorts vs. 2007 to 2010 Prize Cohorts – Varying Funding Duration								
Post Prize	0.38	−0.05	0.04	0.02	−0.01	0.06	−0.10	−0.03
	(0.76)	(0.11)	(0.23)	(0.02)	(0.00)	(0.08)	(0.09)	(0.02)
Post Prize × Post 2007	0.14	0.08	−0.10	−0.01	−0.00	−0.01	0.14	0.16*
	(0.86)	(0.12)	(0.25)	(0.04)	(0.01)	(0.10)	(0.14)	(0.07)
Mean	10.92	0.18	0.58	0.28	0.14	1.83	−0.70	0.05
Observations	1482	1482	1482	1165	946	1268	1268	1482
Panel C: 2000 to 2006 Prize Cohorts vs. 2007 to 2010 Prize Cohorts – Varying Funding Amount and Duration								
Post Prize	2.64***	0.01	−0.22	0.01	0.01	0.14*	−0.02	0.02
	(0.57)	(0.04)	(0.16)	(0.02)	(0.01)	(0.07)	(0.05)	(0.04)
Post Prize × Post 2007	−4.83***	0.05	0.38*	−0.04	−0.03***	−0.11	0.10	0.12
	(1.43)	(0.04)	(0.18)	(0.03)	(0.01)	(0.11)	(0.10)	(0.09)
Mean	12.86	0.18	0.64	0.30	0.14	2.09	−0.77	0.08
Observations	1339	1339	1339	1223	1227	1286	1286	1339

Note: This table shows the results from a difference-in-differences estimation with five years before receiving the Leibniz Prize as pre-period and seven years after as post-period. The estimation equation is as in equation (1.1). The dependent variable in column (1) is the count of all types of publication per year. In column (2) only publications in top general interest journals (Science, Nature, PNAS, Nature Communications) are counted. Column (3) uses publications in field specific top 3 journals (See Table A.2 for the full list of journals). In column (4), the dependent variable is the text similarity of abstracts to each other within a given year. In column (5), 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 (6) and (7), novelty and conventionality are defined as in Lee et al. (2015). Lastly, in column (8), the dependent variable is the count of publications with potential high impact, i.e. having high conventionality and novelty. 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. 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 can rule out many other potential explanations for the productivity effects based 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 chapter 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.[39] 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.

[39] See for example Benedictus et al. (2016), Kolata (2009), and Sarewitz (2016).

2

Do Subsidies for Research Increase Firm Innovation?

2.1 Introduction

Private sector innovation is one of the key drivers of overall economic growth. However, due to the public good nature of new ideas, the private sector might underinvest in research and development (R&D) activities. In line with this fact, many countries offer incentives for private sector R&D activity in the form of project-based subsidies (EY, 2018). The idea behind these subsidies is to lower the costs of R&D activity and ease financing constraints, thus increasing R&D investment and innovation.

In this chapter, we evaluate the largest project-based R&D subsidy program in Germany, the *Zentrales Innovationsprogramm Mittelstand*[1], or ZIM for short. With a total budget of around €500m per year, the ZIM offers companies non-refundable grants of up to €380,000 for R&D projects. The ZIM is open to all technologies and industries. All companies in Germany with fewer than 500 em-

This chapter is based on joint work with Monika Schnitzer and Martin Watzinger.

[1] "Main innovation program for small and medium sized companies", authors' translation.

ployees are eligible to apply. Its stated aim is to improve competitiveness, foster firm growth, and create jobs.[2]

In the first part of the chapter, we analyze the cross-sectional patterns of who can apply for the ZIM. Due to the differential firm size distribution across industries and regions, the size cutoff of the ZIM implies that the program is not neutral with respect to industry or region. It favors industries with small companies, such as Fabricated Metal Productions, over industries with large companies, such as Automotive. It also favors companies in the East German states over companies in the West. In the second part of the chapter, we analyze whether the program effectively increases R&D spending in eligible companies. From our analysis we cannot conclude that the program increases R&D spending or influences any other variable indicating company performance. The reason is that the variances of the outcome variables are too high to separate the signal from the noise using purely observational data.

To evaluate whether a subsidy program increases R&D spending of subsidy recipients over and above the level without the subsidy, we face two challenges: First, there is a paucity of data on R&D and innovation activity. Very few data sets contain information on variables such as R&D spending or R&D employment. Those that do, such as Compustat, usually cover mostly large, publicly traded companies. However, many subsidy programs focus on small and medium sized companies. Second, in terms of identification, firms usually self-select into applying for a subsidy program and project proposals are then selected on their merit. This double selection makes it difficult to construct counterfactuals for subsidy recipients, as there may be (unobserved) differences between recipients and non-recipients driving self-selection and differences in outcomes. Hence, any simple comparison of recipients and non-recipients is bound to be biased.

[2]The mission statement is from the website: `https://www.zim.de/ZIM/Navigation/DE/Infothek/UeberZIM/ueber-zim.html`, last accessed 23 October 2018.

In our study of the ZIM, we can address both of these challenges. First, instead of a publicly available data set, we use proprietary and restricted administrative data on German manufacturing. This data includes information on R&D expenditures and employment, in addition to firm outcomes such as sales. An additional unique firm outcome in our data is new product introduction. This outcome is particularly relevant in the context of R&D subsidies, as new product introduction is the final outcome of the innovation process. The data set "AFiD - Amtliche Firmendaten für Deutschland" ("AFiD - Official company data for Germany") is collected by the Federal Statistical Office of Germany in a consistent manner since 1995. It covers the universe of all firms in German manufacturing with more than 20 employees, irrespective of their ownership structure or location.

To address the self-selection of companies into the program, we exploit an unanticipated reform of the ZIM program in the wake of the Great Recession in 2009. This reform increased the eligibility threshold from 250 to 1000 employees. This change in the eligibility criteria allows us to estimate an intent-to-treat effect. We compare newly eligible companies to companies with more than 1000 employees that continued to be ineligible after the expansion. To make sure that companies are similar, we only compare companies within the same industry.

Using a Regression Discontinuity Design (RDD), we show that after 2009 there is a clear discontinuity in R&D spending and the number of R&D workers, but no discontinuity in sales. The average effect is an increase of around €2.24m that is statistically different from zero on the 10 percent level. Yet, when we use the same outcome variable to estimate a counterfactual treatment in the period prior to the expansion, we find effects of the same magnitude. This speaks in favor of permanent differences between the companies in the treatment and control group. Thus, the identification assumption that treatment companies would have similar outcomes without subsidies as the control sample is most likely not valid.

To address this issue, we control for time-invariant differences between the two groups of firms using a difference-in-discontinuity model (Grembi et al., 2016). When we control for firm and year fixed effects, we do not find any significant effects of eligibility for ZIM on R&D spending or employment. This is also not due to a shift in the composition of R&D spending or a shift of subsidy funds to other investment programs, as we also do not find any significant effects on sales, employment, or new product introduction.

In all these regressions, the confidence bounds encompass all plausible effect sizes for the ZIM. For example, for R&D spending, the 95 percent confidence bounds for the effect of ZIM range from €-1.8m to €1.1m. For that reason, we can neither conclude that the program did not work nor that it did work. This points to the fact that our study is underpowered. Yet, as we already analyse the universe of German manufacturing companies, it is not possible to increase the size of the estimation sample to get more precise estimates. In auxiliary analysis, we show that a randomized controlled trial with at least 520 treated companies and an equal number of control companies might have a reasonable chance of finding an effect, if there is any.

Our chapter contributes to the literature on measuring the effect of R&D subsidies on firm outcomes in three ways. First, we are, to our knowledge, the first to use administrative data on German manufacturing to study the effects of R&D subsidies. Second, by studying the largest German subsidy program with a regression discontinuity design, we can interpret our estimates in a causal fashion. Lastly, we can trace out the effects of the ZIM on R&D inputs, R&D outputs, and firm outcomes, allowing us to assess the effects of ZIM in a comprehensive fashion.

Our chapter is closely related to several studies that use regression discontinuity designs to address the issue of selection based on observables and unobservables.

Most of these studies have access to administrative data from the funding bodies and use the score for an application as a running variable and exploit the discontinuity around some cutoff. The exception is De Blasio et al. (2014) who use time as their running variable around a cutoff date when funding ran out. Bronzini and Iachini (2014), Bronzini and Piselli (2016), and De Blasio et al. (2014) look at Italy and find that effects are mostly confined to smaller firms. Zhao and Ziedonis (2012) study a subsidy program in Michigan and find increased firm survival but no increase in patenting. Howell (2017) looks at the US Department of Energy's SBIR grant program and finds positive effects on patents, financing, and firm survival. Wang et al. (2017b) study a program in China and find that although the program is good at picking winners, it does not seem to have any treatment effect on the recipients. Given that RDD estimates are inherently local and institutional details differ across countries, this may explain why the results in the literature range from no effect to quite sizable effects on innovation.

In addition, our chapter is related to the existing literature on the effects of German R&D subsidy programs. Most of these studies have used matching-on-observable techniques, but unobserved differences between firms remain a serious concern. Examples include Almus and Czarnitzki (2003), Czarnitzki and Lopes-Bento (2014), Aerts and Schmidt (2008), and Aschhoff (2009). None of these studies, however, use the administrative data we employ in this chapter. Other studies using matching-on-observables have looked, among other countries, at Spain (González and Pazó, 2008), New Zealand (Jaffe and Le, 2015), or Ireland (Görg and Strobl, 2007). Becker (2015) and Zúñiga-Vicente et al. (2014) are two fairly recent surveys of this literature.[3]

Lastly, we are closely related to Dechezleprêtre et al. (2016) who study the R&D tax credit in Great Britain from a methodological point of view.[4] They exploit

[3] An earlier survey is David et al. (2000).

[4] The same reform of the R&D tax credit in Great Britain is studied by Guceri and Liu (2019).

a similar change in eligibility criteria and use assets as their running variable. They find very strong effects of the R&D tax credit. As Germany does not have a system of R&D tax credits, their results do not directly speak to the German case. We complement their study by looking at project-based subsidies in Germany.

In Section 2.2 we describe the institutional background and details of the reform. In Section 2.3 we present our data. Section 2.4 presents the cross-sectional patterns of which firms are eligible to apply and discusses the results from the RDD and the differences-in-discontinuities design. Section 2.5 concludes.

2.2 Institutional Background

The *Zentrales Innovationsprogramm Mittelstand* (ZIM) is the main program that subsidizes research and development activities in Germany. This program offers project-based subsidies that cover between 25 percent and 55 percent of the eligible costs of an R&D project for small and medium-sized companies. The maximum subsidy is around €160,000 per project. The subsidized share varies with the size of the firm and is higher for firms in East Germany. Eligible costs encompass mostly wages, but other costs and payment for third-party services may also be included.[5] The program is open to small and medium enterprises from all regions, industries, and technologies.[6]

In our empirical strategy, we exploit the exogenous variation caused by the shift in the size cut-off for the eligibility to the ZIM program during the Great Recession. As part of the second stimulus package from February 2009, funding

[5]For details, see the respective versions of the regulations governing the ZIM, available at `https://www.zim.de/ZIM/Redaktion/DE/Publikationen/Richtlinien/richtlinien-archiv.html`, last accessed 13 March 2019.

[6]The two predecessor programs, INNO-WATT and NEMO, focused solely on firms located in East Germany and Pro Inno II and INNONET had less stringent eligibility criteria for firms located in East Germany. All programs were discontinued with the start of the ZIM except INNO-WATT, which ended in December 2008.

for ZIM was increased by €900m and eligibility was extended to firms with up to 1000 employees. Projects approved under the extended eligibility had to be finished by the end of 2011. Prior to this and beginning with ZIM's inception in July 2008, a firm had to meet the small and medium enterprise definition of the European Union. Firms meet this definition if they have less than 250 employees and either less than €50m turnover or a balance sheet of less than €43m (Commission Recommendation 2003/361/EC).

After the stimulus package ended, the eligibility bar dropped again to 250 employees in January 2011. In the following years there were several further changes in the eligibility criteria: From July 2012 to December 2013, eligibility was extended to firms with up to 500 employees; from January 2014 to April 2015 it dropped again to 250; lastly, it increased again in April 2015 to include firms with up to 500 employees that fulfill the other two EU criteria (turnover and balance sheet size).

According to the Federal Ministry for Economic Affairs and Energy, 440 newly eligible firms undertook 670 projects during the expansion due to the stimulus package. Similar to the overall distribution of firms in space, most recipients are located in the states of Bavaria, Baden-Württemberg and North Rhine-Westphalia. Lastly, the vast majority of firms is active in manufacturing, especially in the manufacturing of machinery (Depner et al., 2011, p. 22-26). On average, each project received around €80,000 in subsidies. In our empirical exercise we aim to measure the effect of these €53m in subsidies on R&D spending of these companies.

2.3 Data

For our study, we use data on the universe of German manufacturing firms with more than 20 employees. This data is collected by the Federal Statistical Office and the Statistical Offices of the German states. Firms are required by law to respond. The data contains information on employment, turnover, location, industry affiliations, and which products they produce in which quantity. A subset of these companies, which are representative for the population of firms, are obligated to give detailed cost information in a more detailed survey, the *Kostenstrukturerhebung*. Due to the high administrative burden of collecting the data, small companies are usually surveyed only for four years in row. Companies with more than 500 employees are surveyed every year.[7]

A key advantage of this data set relative to commonly used data sets, such as Compustat and Amadeus, is that it contains data on R&D expenditures and R&D employees for companies of almost all sizes. R&D expenditures encompasses all types of spending (labor costs, material costs, investment) for R&D activity conducted within the boundaries of the firm.[8] R&D employment are all employees that are either directly tasked with R&D activity or perform directly related services, such as R&D managers or office assistants.

For this data set, each company also has to report which products it produces using a 9-digit code for each product category. In total, there are around 5,000 different product codes with an accompanying high level of granularity. For example, in the classification of 2009, there are different product codes for regular and diet cola products (product codes `1107 19 301` and `1107 19 302`) or cars with

[7]For further details please see Statistisches Bundesamt (2017).

[8]We convert all nominal values into 2010 Euros using the consumer price index of the Federal Statistical Office (`https://www.destatis.de/DE/ZahlenFakten/GesamtwirtschaftUmwelt/Preise/Verbraucherpreisindizes/Tabellen_/VerbraucherpreiseKategorien.html?cms_gtp=145110_slot\%253D2&https=1`), last accessed 16 May 2018.

petrol or diesel motors of different sizes (product codes `2910 22 301`, `2910 22 302`, `2910 23 100`, `2310 23 303`, `2310 23 305`). Hence, we are able to measure very accurately whenever a company changes its product mix, either by adding or dropping a product. Since the development of novel products is a central goal of the ZIM, it seems natural to assess whether there are any positive effects on new product introduction.[9]

For our main analysis, we restrict attention to firms that are in the data in all years to arrive at a balanced sample.[10] In total, we have information on 1,007 firms with 500 to 1,500 employees that report in all years.[11]

2.4 Results

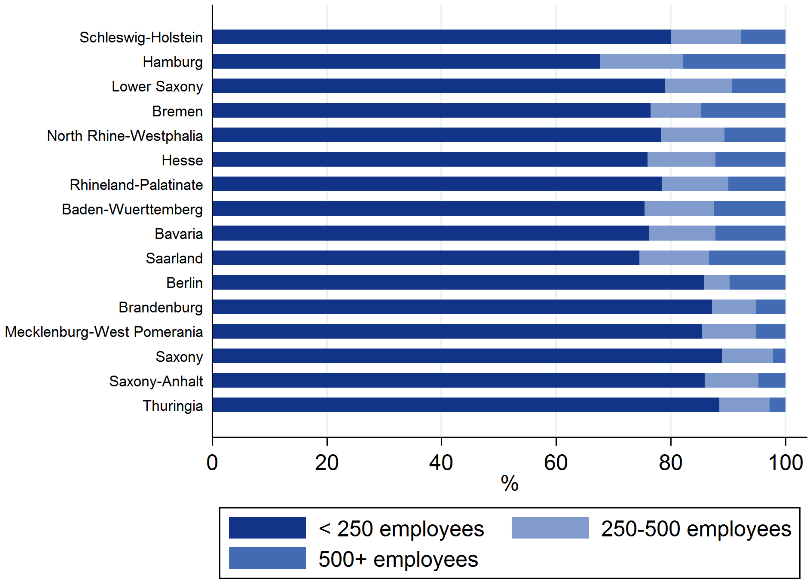
In this section, we first discuss the cross-sectional patterns of the eligibility criteria of the subsidy program under consideration. We show that due to the size limit for companies, industries and regions with on average smaller companies can benefit more than industries and regions with larger companies. In a second step, we use a regression discontinuity design to evaluate the effectiveness of the subsidy program in inducing additional R&D spending. Lastly, we use a differences-in-discontinuities design to address a likely violation of the identifying assumption of the RDD. The results of these exercises are inconclusive.

[9] One issue in accurately determining product entry and exit is the introduction of a new product classification scheme in 2009. Appendix B.1 gives details on how we deal with this issue.

[10] This is due to the fact that if we cease to observe a firm, this can be due to multiple reasons. For example, the company may move its operations away from Germany, shut down their business, or switch sectors. As we cannot disentangle these reasons in the available data, we circumvent this issue by focusing on a balanced panel.

[11] This roughly corresponds to the universe of German manufacturing firms of this size. This is also borne out by two separate sources. First, the firm registry of the Federal Statistical Office lists around 4,200 firms in German manufacturing with more than 250 employees and the Amadeus database by Bureau van Dijk lists 857 companies in German manufacturing with 500-1500 employees.

Figure 2.1: Share of Firms by State and Size Group



Note: This figure shows the share of the number of firms in the manufacturing sector falling into each size group, by state. The number of firms in a state ranges from 102 in Bremen to 3749 in North Rhine-Westphalia. Source: RDC of the Federal Statistical Office and Statistical Offices of the Federal States, *AFiD Panel Industrieunternehmen [2008-2011]*, own calculations.

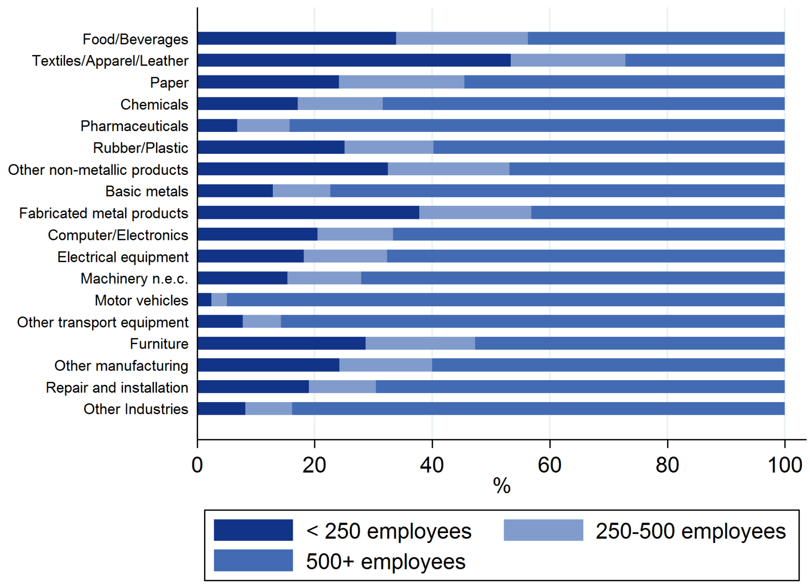
2.4.1 Cross-sectional Patterns of ZIM Eligibility – Who Can Apply?

In 2018, companies with fewer than 500 employees are eligible to apply for the ZIM. As the firm size distribution is not the same in every industry and in every region, this size criteria steers funding from industries and regions with larger companies to regions and industries with smaller companies.

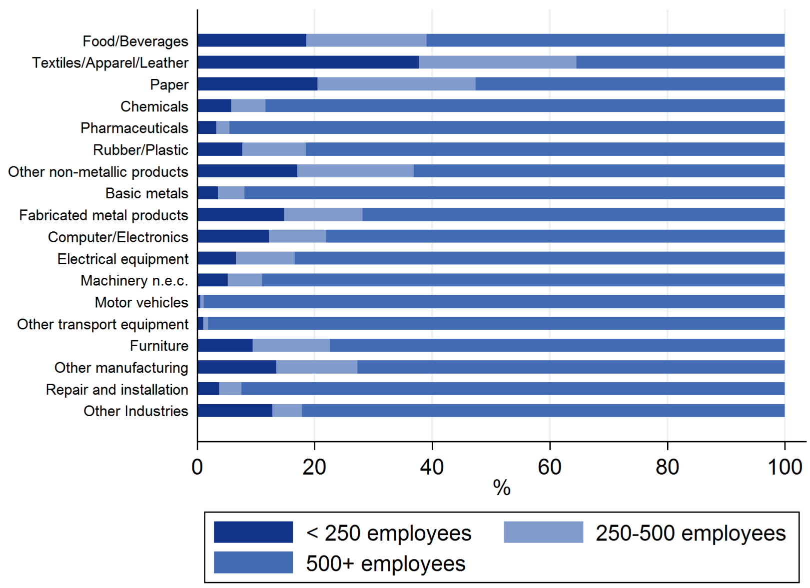
In Figure 2.1 we present the size distribution of firms across German states. Figure 2.2 shows the distribution of sales across industries and firm sizes (Panel (a)), and the distribution of R&D by industry and firm size (Panel (b)). The data is for the year 2008, the last year before the onset of the Great Recession and prior to the beginning of our treatment period in the next section.[12]

[12]Since the firm size distribution, the distribution of firms across space, and the industry structure are all very stable over time, these patterns should still be broadly present today. Unfortu-

Figure 2.2: Cross-Sectional Patterns of Sales and R&D Spending



(a) Distribution of Sales Across Sectors and Size Groups



(b) Distribution of R&D Across Sectors and Size Groups

Note: Panel (a) depicts the share of total R&D spending in 2008 falling into each size group, by two digit industry. Panel (b) shows the share of total firm sales in 2008 falling into each size group, by two digit industry. The "other industries" category encompasses the following two-digit industries: Manufacturing of tobacco products, wood products, printing, coke and petroleum. Source: RDC of the Federal Statistical Office and Statistical Offices of the Federal States, *AFiD Panel Industrieunternehmen [2008-2011]*, own calculations.

Figure 2.1 shows that the eligibility criteria favors East German Firms over West German firms. The share of firms below 500 employees is much smaller in West German states (the first 10 bars) than in East German states and Berlin (the latter 6 bars). The ZIM subsidy program follows in the footsteps of its predecessor programs that were specifically focused on aiding East German firms.

Notionally, the ZIM is open to firms from all sectors. Yet, its limitation to companies with below 500 employees makes it non-neutral with respect to industry. Panel (a) in Figure 2.2 shows the share of sales of companies with different sizes. The sales share of companies with less than 500 employees varies widely from around 10 to 20 percent in the Motor vehicles and Pharmaceuticals industries to over 50 percent for companies in the Food/Beverages and Textiles/Apparel/Leather industries.

The effective focus of the ZIM program is on companies that are actually doing R&D. When looking at the share of R&D across industries, the picture is even more skewed. In the Motor vehicles and Pharmaceutical industries, virtually no R&D spending is undertaken by firms that are eligible to apply for the subsidy. In contrast, in the Fabricated metals, Paper, and Computer/Electronics industries, more than 20 percent of all R&D is done by companies with fewer than 500 employees (Panel (b) of Figure 2.2).

Is it optimal that the considered subsidy program favors some regions, industries and thus technologies over others? The literature provides several arguments under which circumstances the direction of research might be improved by government intervention (Akcigit et al., 2013; Hopenhayn and Squintani, 2016; Bryan and Lemus, 2017). For example, companies might do too little basic research. Yet, as the program is intentionally neutral, these conditions are at best fulfilled by accident.

---

nately, we do not have data for 2018 to directly assess these patterns for the current eligibility criteria of the ZIM.

For a general R&D subsidy program, the usual justification is two-fold. First, R&D activities produce new knowledge that spills over to other companies. As a consequence, the social return of R&D is larger than the private return and companies underinvest in R&D relative to the social optimum. Second, small companies might face borrowing constraints and therefore cannot implement the optimal investment plans. These borrowing constraints are more likely to be binding for R&D investment than for investment in real assets, because real assets can be pledged as security for a loan.

Neither of these justifications maps neatly into a size cut-off of 500 employees. There is no evidence available that the R&D activities of small companies produce more knowledge spillovers than the R&D programs of large companies. To the contrary, Bloom et al. (2013) show that small companies are more active in technological niches and therefore their R&D program have fewer knowledge spillovers. Schnitzer and Watzinger (2017) show that spillovers from venture capital-backed start-ups are larger than for established companies. Yet, next to none of the companies applying for ZIM are venture capital financed. Taken together, while it might be sensible to subsidize the research of small firms, spillovers are not a good argument to restrict it only to small firms.

Similarly, while smaller companies are thought to face more financing constraints than larger companies, it is unclear why these constraints should change discontinuously at a size of 500 employees or any other arbitrary point in the firm size distribution. In a representative sample of 4400 German companies in August 2006, around 10 percent of companies with more than 250 employees reported restrictive credit policies, while 30 percent of companies with less than 50 employees reported a restrictive credit policy. In contrast, at the height of the crisis in August 2009, around 45 percent of companies of all size classes reported problems in arranging a loan (Kunkel, 2010). Therefore, the employment size of a company seems to be a rather imprecise measure for financing constraints.

Taken together, it is unclear why an optimal R&D subsidy program should use a size cut-off. Additionally, even if the size cut-off is necessary for legal reasons, the program could be improved by targeting credit constrained companies directly, e.g. companies with few pledgeable assets (Almeida and Campello, 2007). An alternative would be to link subsidies to firm age instead of firm size. Haltiwanger et al. (2013) show that young and small companies create more jobs and grow faster than companies that are old and small. Therefore, such a subsidy program would directly reward job-creation and foster the entry of new companies.

2.4.2 Did ZIM Increase R&D Outlays?

In this section, we analyze whether the ZIM subsidies increased R&D spending of companies. If companies use the subsidy for the intended purpose, we should observe more R&D spending, more R&D workers, and more new products. If companies substitute their own R&D spending with the government funds, the effect of the subsidy on R&D spending might be zero.

Regression Discontinuity Design

From observational data alone, it is difficult to learn whether the ZIM R&D subsidy is effective because firms need to apply for the subsidy program and project proposals are then selected on their merit. This double selection makes it difficult to construct counterfactuals for subsidy recipients, as there may be (unobserved) differences between recipients and non-recipients driving differences in outcomes. Hence, any simple comparison of recipients and non-recipients is bound to be biased.

To address this identification problem, we exploit the unanticipated reform in 2009 in wake of the Great Recession. This reform increased the eligibility threshold from 250 to 1000 employees. The new threshold of 1000 employees made the group of firms between 250 and 1000 employees newly eligible, whereas firms with more than 1000 employees continued to be ineligible. In order to ensure comparability of firms, we limit attention to firms with more than 500 employees and less than 1500 employees.[13] The change in the cutoff lends itself to a (fuzzy) regression discontinuity design with employment as the forcing variable.

Using this variation in eligibility, we can estimate the intent-to-treat effect of the subsidy. By focusing on the intent-to-treat effect instead of the average treatment effect, we circumvent the bias resulting from self-selection and proposal selection on merit. Also, arguably, the intent-to-treat effect is the policy relevant effect as politicians can change eligibility but cannot force firms to apply for the grant.[14]

We estimate the following specification:

$$y_{i,t} = \gamma_0 + \gamma_1 T_{i,2008} + \gamma_2 E_{i,2008} + \gamma_3 E_{i,2008} \cdot T_{i,2008} + \epsilon_{i,t} \tag{2.1}$$

where $y_{i,t}$ is the dependent variable (R&D spending, R&D employees, sales) of firm i in year t. $T_{i,2008}$ is a treatment indicator whether employment of company i is below the threshold of 1000 employees in 2008. We use employment in 2008 prior to the enactment of the second stimulus package as forcing variable to alleviate any concerns that companies might shed employees to become eligible for the subsidy program.[15] $E_{i,2008}$ is employment in 2008. As controls we include employment in 2008 linearly on both sides of the threshold ($E_{i,2008} \cdot T_{i,2008}$).[16] We cluster standard errors on the firm level. To make sure that companies on the left and right of the cut-off are similar, we only compare companies within the same industry using coarsened exact matching (CEM) as proposed by Iacus et al.

[13] In addition, this ensures that all firms are surveyed in all years.

[14] However, if we had data on which firms actually apply and which ones receive funding from ZIM, we could also estimate the local average treatment effect by using the change in eligibility as an instrumental variable. In absence of this data, we focus on the intent-to-treat effect.

[15] Also, the stimulus package was passed after minimal debate and very quickly, so we do not expect firms to be able to anticipate or influence the new threshold.

[16] Including higher order polynomials does not change our results substantively. These are available from the authors upon request.

(2012).[17] We choose a symmetric bandwidth of 500 employees, i.e. we include firms with 500 to 1500 employees.[18] Appendix B.2 shows summary statistics for the treatment and control group.

Implementing this RDD strategy, Figure 2.3 shows a binned scatter plot for average R&D expenditures from 2009 to 2011, the treatment period. Our forcing variable, employment in 2008, is grouped into bins of 50 employees and we show fitted linear regressions on either side of the cutoff.[19]

There is a clear discontinuity at 1000 employees with firms with 500 to 999 employees spending significantly more on R&D between 2009 and 2011 than firms with 1000 to 1500 employees, conditional on employment in 2008. The corresponding regression results are shown in Column (7) of Table 2.1. From 2009 to 2011, firms eligible for ZIM spend around €2.2m more on R&D, a very large difference of over 50 percent relative to the mean of the treatment group prior to the expansion.

This effect is significantly larger than what we would expect given the size and structure of the program. The maximum eligible project size is €380,000, which would constitute a 12 percent increase relative to the pre-treatment mean of €3.2m. As we only observe the intent-to-treat effect, we have to scale this number by the likelihood that a company gets funding. To calculate the expected magnitude, we assume full directionality, i.e. that the entire project is only undertaken due to the ZIM. This seems plausible, as 99 percent of firms that received the subsidy stated in Depner et al. (2011) that funding via ZIM was relevant for

[17]Our results are also fairly robust to varying the strata used for the CEM procedure. For example, using three or four digit industries or two-digit industries $\times$ East/West Germany do not alter our results to a great extent. The results are available by the authors upon request.

[18]We validate this ad-hoc bandwidth by showing in the robustness section that our results are not too sensitive to varying this bandwidth. The Imbens-Kalamanyaran optimal bandwidth algorithm suggests an optimal bandwidth of only 28. For this small bandwidth we are left with only 28 firms.

[19]The bin for 500 employees includes all firms with 500 to 549 employees. The number of underlying companies per bin varies from nine to 120.

their project. Given that 440 firms with more than 250 employees received ZIM funding and the firm size distribution, we arrive at a scaling factor of 5 to 57 percent, implying an effect of between 0.6 percent to 6.8 percent of additional R&D spending. A reasonable middle-ground scaling factor implies an increase of 3.9 percent, or around €124,000. This expected effect is much smaller than our estimated effect of €2.2m.

We find a similar discontinuity for R&D employment, as can be seen in Panel (b) of Figure 2.3. The point estimate in the second row of Column (7) in Table 2.1 is large and positive, but not statistically significant. In terms of magnitude, we estimate an intent to treat effect of around 14 additional employees.

In contrast to the results on R&D outcomes, we do not find a discontinuity for the sales of a company. In Panel (c) of Figure 2.3 there is no visual difference between treatment and control group, indicating that the ZIM had no effect on overall firm sales. This might be due to our time horizon, as many projects only ended in 2011. Many firms indicated in Depner et al. (2011, p.50) that they have not yet monetized the results of their R&D project and need additional time to do so. The regression coefficient in row 3 of column (7) in Table 2.1 is small and not statistically different from zero. Given the time lags between R&D activity and introduction in the market, our findings of no discontinuity in sales from 2009 to 2011 could also be seen as an indication that firms around 1000 employees are comparable to each other, conditional on employment.

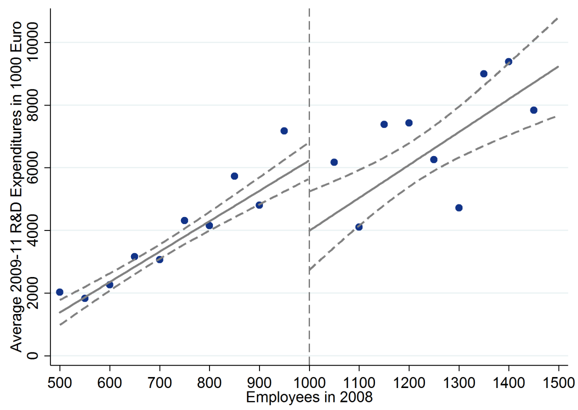
The key identifying assumption in a RDD is that all pre-determined and unrelated variables vary smoothly around the threshold. To see whether this is the case, we redo our main analysis for the period *prior* to the stimulus package expansion, i.e. from 2007 to 2008, where none of the companies were eligible for ZIM. In Figure 2.4, we plot the pre-treatment outcomes for R&D expenditures, R&D employees, and sales for the period from 2007 to 2008. The picture looks

Table 2.1: Effect of ZIM Eligibility on R&D Inputs, RDD Estimates

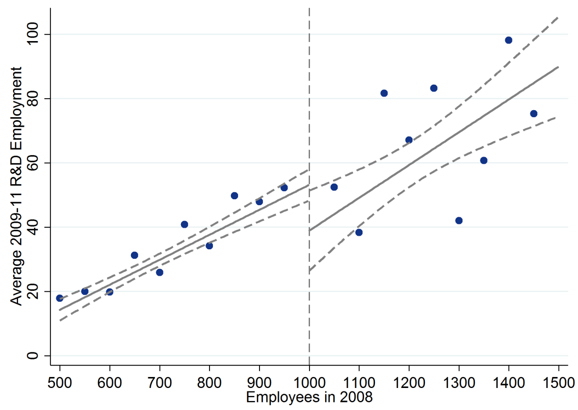
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Coefficient on < 1000 employee dummy						
	Pre-policy		Post-policy			Pre	Post
	2007	2008	2009	2010	2011	2007-08	2009-11
Dependent Variable							
R&D expenditures	2176.52	3101.10**	2670.70	1576.67	2530.64**	2638.81*	2245.46*
	(1421.74)	(1491.16)	(1226.37)	(1323.56)	(1235.30)	(1407.77)	(1198.66)
R&D employment	26.07	11.44	15.57	11.13	16.05	18.75	14.26
	(18.46)	(10.48)	(10.17)	(11.08)	(11.07)	(12.97)	(10.57)
Indicator for positive R&D expenditures	0.16**	0.16**	0.17**	0.14*	0.14**	0.18**	0.11
	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)
N	989	989	989	989	989	1978	2967

Note: Firms between 500 and 1500 employees in 2008, covering 2007 to 2011. Diff-in-disc estimate of the impact of becoming eligible for the ZIM. Regression includes employment in 2008 on both sides of the threshold as in equation (2.1). Standard errors in parentheses are clustered on the firm level. *, ** and *** denote significance on the 10 percent, 5 percent and 1 percent level, respectively. Source: RDC of the Federal Statistical Office and Statistical Offices of the Federal States, *AFiD Panel Industrieunternehmen [2008-2011]*, own calculations.

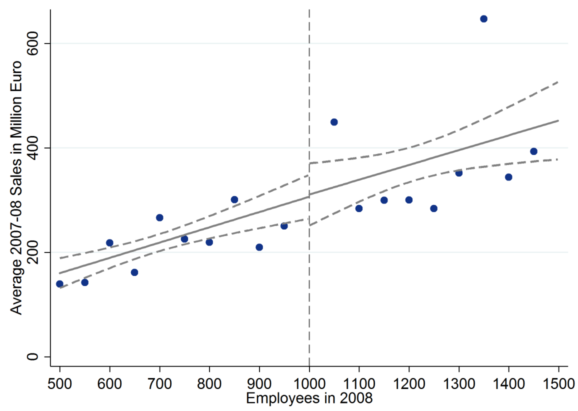
Figure 2.3: Outcomes in Post-period: 2009 to 2011



(a) R&D Expenditures



(b) R&D Employees



(c) Sales

Note: In Panel (a) we show average R&D expenditures from 2009 to 2011 (in 2010 Euros) by employment bins in 2008. In Panel (b) we show average R&D employees and in Panel (c) the average sales for the same period. The fitted lines and 95 percent confidence intervals are estimated on the full data. The number of firms in each bin ranges from 9 to 120. Source: RDC of the Federal Statistical Office and Statistical Offices of the Federal States, *AFiD Panel Industrieunternehmen [2008-2011]*, own calculations.

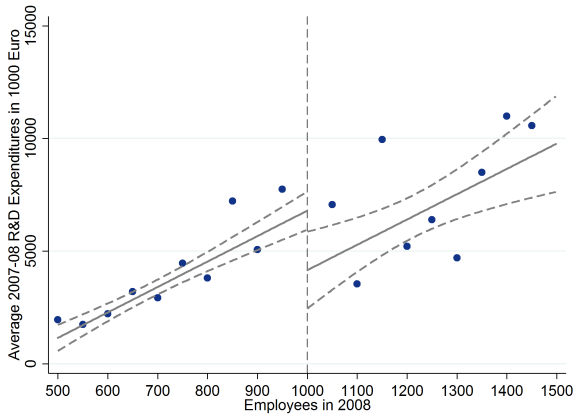
virtually identical to Figure 2.3. There is a sizable discontinuity for R&D expenditures and R&D employees, but none for sales. In column (6) in Table 2.1, we repeat the RDD regression for the pre-period. The difference in R&D expenditures and R&D employees has the same magnitude as in the post-period.

The results using the counterfactual treatment in the pre-period therefore indicate a clear violation of our identification assumption underlying the RDD, namely that companies to the left and to the right of the cut-off would have comparable outcomes without the subsidy. In unreported analyses, we have tested a battery of further dependent variables such as wage bill, average wages, total costs, investment in physical capital, and exports, and consistently found smooth variation around the threshold. Hence, a straightforward explanation would be an additional R&D focused policy using this cutoff. However, to the best of our knowledge, no such policy exists and unobserved firm heterogeneity is the most likely remaining explanation. The full results can be found in Table 2.1, where we follow Dechezleprêtre et al. (2016) and also present results for each year separately.

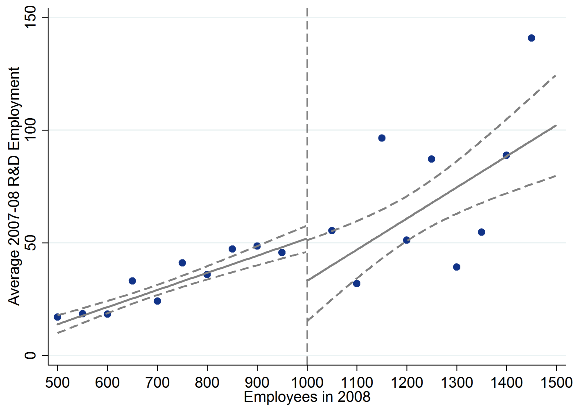
Difference in Discontinuities Design

Under the assumption that this unobserved firm heterogeneity does not vary over time, we can nevertheless estimate the causal intent-to-treat effect by employing a *difference-in-discontinuities* ("diff-in-disc") specification as in Grembi et al. (2016). This approach deals with any confounding factors that vary at the same cutoff as our policy of interest, but do not vary over time. Grembi et al. (2016) provide a detailed discussion of the properties of this estimator and the underlying assumptions. The key assumption is that the effect of a confounding policy at the threshold may not vary over time. This assumption seems reasonable in our setting, given that only the cutoff for the ZIM was changed to 1000 employees

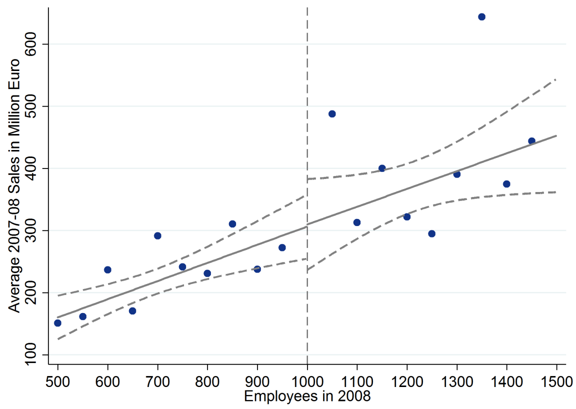
Figure 2.4: Outcomes in Pre-period: 2007 to 08



(a) R&D Expenditures



(b) R&D Employees



(c) Sales

Note: In Panel (a) we show average R&D expenditures from 2007 and 2008 (in 2010 Euros) by employment bins in 2008. In Panel (b) we show average R&D employees and in Panel (c) the average sales for the same period. The fitted lines and 95 percent confidence intervals are estimated on the full data. The number of firms in each bin ranges from 9 to 120. Source: RDC of the Federal Statistical Office and Statistical Offices of the Federal States, *AFiD Panel Industrieunternehmen [2008-2011]*, own calculations.

but to our knowledge all other policies around this threshold stayed in place. In the diff-in-disc specification, we estimate the following equation:

$$\begin{aligned} y_{i,t} = {} & \beta_0 + \beta_1 T_{i,2008} \cdot P_{i,t} + \beta_2 T_{i,2008} + \beta_3 P_{i,t} + \beta_4 E_{i,2008} \\ & + \beta_5 E_{i,2008} \cdot T_{i,2008} + \beta_6 E_{i,2008} \cdot P_{i,t} + \beta_7 E_{i,2008} \cdot P_{i,t} \cdot T_{i,2008} + \eta_{i,t} \end{aligned} \tag{2.2}$$

where $P_{i,t}$ is a post-period indicator equal to one for the years 2009 to 2011. All other variables are defined as in equation (2.1). As we focus on a balanced panel of firms, this specification is equivalent to the inclusion of firm and year fixed effects. The coefficient of interest is now β_1. It yields the causal intent-to-treat effect of being eligible to apply for the ZIM, relative to the pre-existing difference between the eligible and ineligible group.

We present the results for our diff-in-disc estimates in Table 2.2. The first column shows the results of estimating equation (2.2) with R&D expenditures as the dependent variable and column (2) with R&D employees, respectively. In column (3) we focus on the extensive margin and use an indicator for positive R&D expenditures as our dependent variable. In all specifications, we find statistically insignificant effects for the interaction term of after 2009 and smaller than 1000 employees. The point estimate is negative for all three dependent variables. For R&D expenditures, the point estimate would imply a reduction of €380,000 due to the ZIM and 95 percent confidence bounds range from €-1.8m to €1.1m.

One benign explanation for us not finding an effect on R&D inputs such as expenditures or R&D employees might be that firms keep their overall R&D efforts the same, but shift the composition of their R&D projects. For example, they might pursue riskier projects with potentially higher upsides. Alternatively, they might use the received subsidy for other productive investment that is not classified as R&D. Such a change might lead to differences in firm outcomes such as higher growth in employees or sales, an entry into exporting, or changes in the

Table 2.2: Effect of ZIM Eligibility on R&D Inputs, Diff-in-Disc Estimates

	(1) R&D exp.	(2) R&D empl.	(3) R&D ind.
$<$ 1000 employees	2638.81*	18.75	0.16**
	(1407.70)	(12.97)	(0.07)
After 2009	−182.22	5.79	0.01
	(619.86)	(8.39)	(0.02)
$<$ 1000 employees $\times$ After 2009	−380.15	−4.51	−0.01
	(768.74)	(8.57)	(0.03)
R^2	0.07	0.07	0.09
Number of firms	989	989	989
Observations	4945	4945	4945

Note: Firms between 500 and 1500 employees in 2008, covering 2007 to 2011. Diff-in-disc estimate of the impact of becoming eligible for the ZIM. Regression includes employment in 2008 on both sides of the threshold in pre- and post-policy period as in equation (2.2). Standard errors are clustered on the firm level. *, ** and *** denote significance on the 10 percent, 5 percent and 1 percent level, respectively. Source: RDC of the Federal Statistical Office and Statistical Offices of the Federal States, *AFiD Panel Industrieunternehmen [2008-2011]*, own calculations.

Table 2.3: Effect of ZIM Eligibility on Firm Outcomes, Diff-in-Disc Estimates

	(1) Empl.	(2) Sales	(3) Export	(4) Add prod.	(5) Drop prod.
$<$ 1000 employees	−4.57	−3.30	0.08	0.02	−0.00
	(7.73)	(47.64)	(0.05)	(0.01)	(0.01)
After 2009	−9.44	−35.57**	0.00	−0.00	0.01
	(21.13)	(16.20)	(0.00)	(0.01)	(0.01)
$<$ 1000 employees	3.17	17.41	0.01	−0.02	−0.01
$\times$ After 2009	(24.69)	(17.27)	(0.01)	(0.02)	(0.01)
R^2	0.87	0.03	0.01	0.00	0.01
Number of firms	989	989	989	989	989
Observations	4945	4945	4945	4945	4945

Note: Firms between 500 and 1500 employees in 2008, covering 2007 to 2011. Diff-in-disc estimate of the impact of becoming eligible for the ZIM. Regression includes employment in 2008 on both sides of the threshold in pre- and post-policy period as in equation (2.2). Standard errors are clustered on the firm level. *, ** and *** denote significance on the 10 percent, 5 percent and 1 percent level, respectively. Source: RDC of the Federal Statistical Office and Statistical Offices of the Federal States, *AFiD Panel Industrieunternehmen [2008-2011]*, own calculations.

product mix of firms.

To see whether this is the case, we use sales, employees (overall, not only employees tasked with R&D), an exporter dummy, and dummies for adding or dropping a product as outcome variables. Especially studying adding and dropping of products is of interest, as over 70 percent of firms stated in Depner et al. (2011, p. 31) that the goal of their project was product development. Given the granularity of product codes in our data, we expect to measure new product development as long as it is not merely an upgrade of an existing product. In Appendix B.1 we explain in detail how we define entry and exit.

Using our specification from equation (2.2) once more, we also find statistically insignificant effects for these firm outcomes, with positive point estimates for all outcomes except product entry and exit (Table 2.3). Taken together, we cannot find any statistically significant impact of becoming eligible for the ZIM on firm innovation investments or firm outcomes.

Given the wide range of our confidence intervals, we cannot conclude anything about the effects of the ZIM on firms using observational data at the cut-off of 1000 employees. Due to the large variability of the outcome variables in the data and our sample size, we do not have enough statistical power. However, we cannot increase the sample size as we are already studying the universe of German manufacturing firms. Theoretically speaking, one could expand the sample to non-manufacturing firms. Yet, from Depner et al. (2011) we know that the vast majority of the ZIM recipients are manufacturing firms and from Stifterverband für die deutsche Wissenschaft (2018) we know that manufacturing accounts for 96 percent of private sector R&D expenditures. Hence, manufacturing firms seem the right population to study.

Table 2.4: Inputs for Power Calculation

Parameter	Value
Share of treated units	0.5
Minimum detectable effect	190,000
Standard deviation of R&D spending	1,206,592
t-value corresponding to desired significance level	1.96
t-value corresponding to desired power	0.84
R^2 of regression of R&D spending on covariates	0.1774

Note: All variables are measured in 2011. Covariates are size as measured by employment and three digit industry dummies. Source: RDC of the Federal Statistical Office and Statistical Offices of the Federal States, *AFiD Panel Industrieunternehmen [2008-2011]*, own calculations.

Power Calculation for a Randomized Controlled Trial

The results of our study support the idea that the ZIM should be evaluated with a randomized controlled trial. If we take the current structure of the ZIM as given, i.e. the focus on firms with fewer than 500 employees, our power calculations suggest that the randomized controlled trial should have at least 1040 companies, split equally into a treatment and control group.

To arrive at this number, we consider a very simple experimental design with randomization on the individual firm level and an equal number of treated and control units. We assume that the minimum detectable effect size should be €190,000, i.e. half of the maximum eligible project size of €380,000. We follow convention and set the desired power to 80 percent. From the data we estimate the standard deviation of R&D spending and how much covariates explain in the outcome variable (R^2 of 18 percent). As covariates, we use industry fixed effects and firm size as measured by employment. Table 2.4 lists all of the inputs.

Using these inputs, the required sample size is 520 treated firms and the same number of control firms. The manufacturing sector covered in our data has 13,800 firms with fewer than 500 employees. Hence, one would need to randomly subsidize 3.7 percent of the universe of all manufacturing companies to study whether the ZIM is effective. Given the average subsidy size of €135,000 (Depner et al., 2018, p.33), the required funds for this study would be €70m. Given that the budget for ZIM is around €500m per year and around 2500 projects are conducted by firms each year (Depner et al., 2018, p.19), such a randomized controlled trial would be feasible within the budget currently appropriated for ZIM.

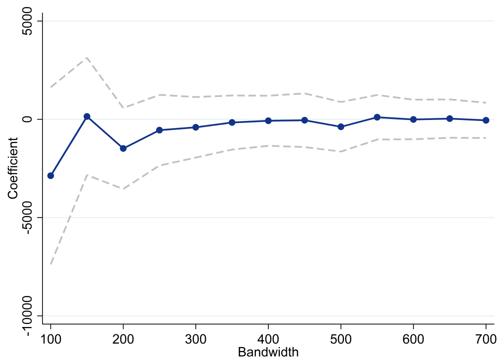
Robustness Checks

In this section, we show additional results to document that our conclusions are robust to changes in the empirical specification. First, we repeat our main estimations, but include industry × year interactions or state × year interactions (Table 2.5) to see whether these soak up some of the firm variability. Second, we vary the bandwidth of our running variable and run some placebo checks by varying the threshold (Panels (a) and (b) in Figure 2.5). Furthermore, we show that the results do not change if we selectively drop single industries or firms (Panels (c) and (d) in Figure 2.5).

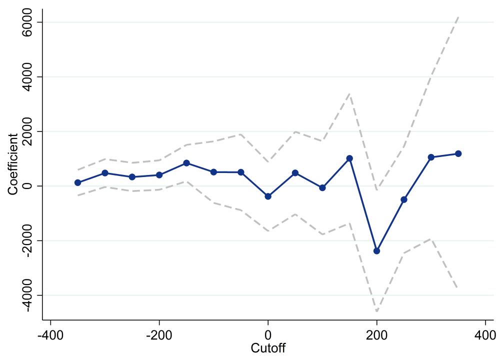
In Table 2.5, we include state × year interactions in the odd numbered columns and industry × year interactions in the even numbered columns. Overall, the results are very similar to our baseline results with our coefficient of interest having a negative sign and large standard errors (except in column (6), where the sign flips). In addition, the coefficients are also very close to the baseline estimate in terms of magnitude. Looking at the R^2, we see that the industry × year interactions substantially add explanatory power, with little change for the state × year interactions. However, the standard errors are unaffected by the inclusion of these interactions, indicating that there is still substantial variability between firms even within industries.

In Panel (a) of Figure 2.5, we vary the bandwidth around our threshold of 1000 employees between 100 and 700 employees in 2008 and plot the resulting coefficient on the below 1000 employees and post 2009 interaction. This implies

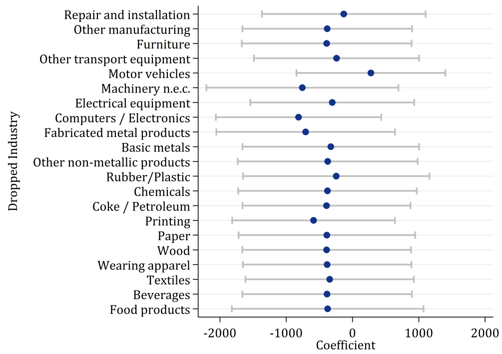
Figure 2.5: Robustness Checks



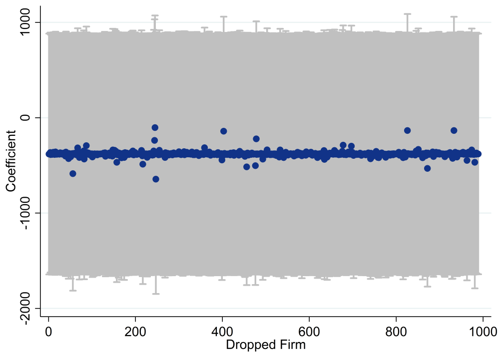
(a) Varying Bandwidths



(b) Placebo Check with Varying Thresholds



(c) Dropping Industries One-by-One



(d) Dropping Firms One-by-One

Note: In Panel (a) we vary the bandwidth between 100 and 700 employees in 2008 around the threshold of 1000 employees. Bandwidth of 500 corresponds to the baseline estimate. Reported are the coefficient and the 90 percent confidence intervals on the post-treat interaction in regression (2.2). In Panel (b) we show the results using different placebo thresholds. Cutoff of 1000 corresponds to the baseline estimate. A constant bandwidth of 500 is chosen for all cutoffs. In Panel (c) we sequentially drop each of the two-digit NACE industry one-by-one. In Panel (d) we sequentially drop each firm one-by-one. Matching is done each iteration and sample size varies from 837 to 988 firms. Source: RDC of the Federal Statistical Office and Statistical Offices of the Federal States, *AFiD Panel Industrieunternehmen [2008-2011]*, own calculations.

Table 2.5: Robustness to Inclusion of Controls, Diff-in-Disc Estimates

	R&D expenditures		R&D employees		Indicator for R&D Activity	
	(1)	(2)	(3)	(4)	(5)	(6)
< 1000 employees	2579.68*	2320.35*	17.88	15.47	0.15**	0.14**
	(1414.96)	(1309.85)	(13.34)	(11.74)	(0.07)	(0.07)
After 2009	4427.07	959.27	19.51	18.58	0.25	−0.04
	(3252.01)	(1248.23)	(22.02)	(17.02)	(0.15)	(0.10)
< 1000 employees × After 2009	−460.53	−390.08	−5.11	−4.16	−0.01	0.00
	(773.67)	(764.00)	(8.88)	(8.41)	(0.03)	(0.03)
State × Year	yes	no	yes	no	yes	no
Industry × Year	no	yes	no	yes	no	yes
R^2	0.09	0.19	0.11	0.20	0.04	0.21
Number of firms	989	989	989	989	989	989
Observations	4945	4945	4945	4945	4945	4945

Note: Firms between 500 and 1500 employees in 2008, covering 2007 to 2011. Diff-in-disc estimate of the impact of becoming eligible for the ZIM. Regression includes employment in 2008 on both sides of the threshold in pre- and post-policy period as in equation (2.2). Industry denotes two digit NACE codes. Standard errors are clustered on the firm level. *, ** and *** denote significance on the 10 percent, 5 percent and 1 percent level, respectively. Source: RDC of the Federal Statistical Office and Statistical Offices of the Federal States, *AFiD Panel Industrieunternehmen [2008-2011]*, own calculations.

that for the smallest bandwidth we only include firms with between 900 and 1100 employees, whereas for the largest bandwidth we include firms from 300 to 1700 employees. We face a trade-off between reducing bias by shrinking our bandwidth and choosing more similar firms and reducing variance by increasing our bandwidth to obtain a larger number of observations. The number of firms ranges from 104 for a bandwidth of 100 to 1686 firms for a bandwidth of 700. As expected, the standard errors get smaller as we increase the bandwidth and the estimates are very noisy for the smallest bandwidth choices. Overall, all coefficients are close to zero with large standard errors.

Next, we keep the bandwidth constant at 500 employees, but re-estimate our baseline model using placebo cutoffs in 50 employee intervals between 650 and 1350 employees. Results can be found in Panel (b) of Figure 2.5.[20] The number of firms varies between 420 for a cutoff of 1350 employees and 2370 for a cutoff of 650 employees. The estimated coefficients are once again around zero and tend to be positive for other placebo cutoffs than our real cutoff of 1000. However, the confidence intervals include zero for almost all cutoffs and get much larger as our number of firms drops for higher cutoffs. In unreported regressions, we conducted the same exercise for our simple RDD model. Here we find that a positive significant effect is only found for our actual cutoff of 1000 employees and not the placebo cutoffs. This supports the underlying RDD assumption, but unfortunately this pattern also exists for the pre-treatment period. Hence, we prefer the diff-in-disc estimates presented here. Since our baseline estimates were insignificant, the absence of a significant effect for placebo cutoffs is only limited support for our identifying assumption.

In the last two panels, we investigate whether our results are driven by any industries or firms in particular. In Panel (c) of Figure 2.5, we do a leave-one-out exercise and re-estimate our baseline model and drop each industry one-by-one.

[20] Note that the horizontal axis is relative to our actual cutoff of 1000 employees.

We then plot the resulting coefficient of interest and its associated confidence intervals. The coefficients are mostly fairly close to each other and are all within the confidence intervals of each other. All coefficients are below zero, except when we drop the motor vehicles industry and obtain a coefficient of 277.61, which, however, is also not statistically significant. Lastly, we drop each firm in turn in Panel (d) of Figure 2.5. There is some variability in the estimates, but regardless of which firm we leave out, the coefficient is negative and statistically insignificant.

Taken together, our robustness checks imply that the large variability in the data cannot be absorbed by industry $\times$ year interactions and regardless of the exact specification we obtain statistically insignificant effects of the ZIM on R&D inputs. Our results are also not driven by particular outliers nor by the exact choice of our bandwidth.

2.5 Discussion and Conclusion

We study the largest German R&D subsidy program and assess its causal effect on firm innovation outcomes. We first document that changes in the employment-based eligibility criteria have implications regarding which sectors and regions can disproportionately apply for the subsidy. By exploiting a change in the eligibility criteria due to the stimulus package in wake of the Great Recession, we estimate the causal intent-to-treat effect of being able to apply for ZIM. We find inconclusive results due to the large variability of the outcome variable in the data.

This chapter is the first study to assess the effectiveness of ZIM in a causal fashion. Whereas so far ZIM has only been evaluated by asking recipient companies about their experience, the current wave of ZIM is supposed to be evaluated with a matched control group design.[21] However, such a design does not mitigate concerns about selection on unobservables. An alternative way to evaluate the ZIM would be to randomly assign ZIM grants to eligible firms, e.g. via an oversubscribed lottery. However, as our back-of-the-envelope power calculation indicates, the sample size for such an effort needs to be quite large and should encompass at least around 1000 firms.

[21]See the Appendix to the Evaluation Plan for ZIM, available at `www.evergabe-online.de/tenderdetails.html?id=189659`, last accessed 06 March 2018.

3

Job Creation in Tight and Slack Labor Markets

3.1 Introduction

The Great Recession has sparked renewed interest in the effects of stimulus spending. One focus of the debate is whether the size of the multiplier depends on the state of the economy, i.e., whether additional demand shocks have larger effects on output in times of slack resources than in times of full capacity utilization. This question is still unresolved: The growing empirical literature using either variation on the national level over time or cross-sectional variation on the regional level has not come to a consensus yet.[1]

Assessing the state-dependence of the multiplier is challenging for three reasons: First, identifying the multiplier requires exogenous demand shocks. Second, these shocks need to vary sufficiently within each state of the economy. Otherwise there is potentially not enough statistical power to estimate the multi-

This chapter is based on joint work with Lukas Buchheim and Martin Watzinger, which is accepted for publication in the Journal of Monetary Economics (Buchheim et al., forthcoming).

[1]For example, Auerbach and Gorodnichenko (2012, 2013), Bachmann and Sims (2012), Fazzari et al. (2014), Nakamura and Steinsson (2014), Shoag (2015), and Dube et al. (2018) find evidence for state dependent multipliers, while others find little differences in the multiplier by the state of the economy (e.g., Owyang et al., 2013; Caggiano et al., 2015; Biolsi, 2017; Ramey and Zubairy, 2018).

plier for each state. Third, the shocks have to be comparable in their composition across the different states.

The literature to date has addressed the first and second challenge. Shocks that are plausibly exogenous to the economic circumstances have been identified, for example, from structural VARs, Ramey's (2011) military news, and a variety of instrumental variables on the subnational level.[2] In order to have a sufficient amount of shocks across different states of the economy, prior work has extended time series data to the late 19th century (Ramey and Zubairy, 2018), combined data from multiple countries (Auerbach and Gorodnichenko, 2013), and focused on the regional level (Dube et al., 2018). In contrast to addressing the issues of exogeneity and statistical power, the literature has been silent about the third challenge: a differential composition of spending across the states of the economy. Specifically, one issue could be that the primary goal of public investments in booms is to increase the long-run growth potential with projects characterized by long planning horizons. Spending during recessions, in contrast, may be designed to deliver a quick boost to the economy. In this case, the state-dependent composition of demand may lead to differential short-run effects on economic activity, even for demand shocks of the same size.

In this chapter, we are the first to tackle all three challenges at once. We do so by studying local demand shocks from investment in rooftop photovoltaic (PV) systems in Germany between 2003 and 2012. In this time period, over €60 billion were invested in PV systems due to a guaranteed price for photovoltaic electricity. Using the cross-sectional and time-series variation resulting from this program, this chapter seeks to answer the following question: Did PV installations create more jobs in slack than in tight labor markets?

[2]Examples for instrumental variables used in the literature are windfall profits from state pension plans (Shoag, 2015), exposure to military spending (Nakamura and Steinsson, 2014), or shocks to population counts from the decennial census (Suárez Serrato and Wingender, 2016).

The setting of PV investment in Germany is ideal to estimate state-dependent multipliers as PV installations constituted an exogenous, frequent, and comparable demand shock in all states of the economy. First, the variation in PV installations over time and across space was mainly driven by factors that are exogenous to the economic circumstances: Over time, investment was largely determined by the world market price of solar modules in relation to the amount of the guaranteed price. The spatial variation in investments across the 400 German counties depends on the local solar radiation and the availability of suitable rooftops. Second, there is ample identifying variation in these investments for any partition of counties into groups with slack and tight labor markets, as the installation of rooftop PV systems was profitable in all German regions. Third, the composition of investment has been constant as each PV installation of a given size constituted the same demand shock. Also, we directly observe the physical amount of investments. Hence, variation in regional prices, which may be a function of labor market tightness, cannot affect our results.

Our main finding is that the installation of PV systems created at least twice as many jobs in slack labor markets characterized by high unemployment than in tight labor markets with low unemployment. Our preferred specifications compare job creation in tight and slack markets using two splits: First, along the time series dimension, we compare counties at times when their unemployment rate is above or below its long-run average. Second, along the cross-sectional dimension, we compare counties with high or low unemployment relative to their state average in a given year. For these sample splits, the installation of PV systems with a capacity of one megawatt peak (MWp) led to about 37 new jobs in slack labor markets, but created only around 13 jobs or less in tight labor markets. This implies that €100,000 of investment created 1.2 job-years in slack and fewer than 0.5 job-years in tight markets. Moreover, using these estimates to approximate a local labor earnings multiplier, we arrive at a multiplier of 1.1 in slack markets and below 0.5 in tight markets.[3] These results are robust to various alternative ways of classifying slack and tight markets, and remain qualitatively unchanged when we instrument PV installations with their profitability as measured by the investments' net present value.

Furthermore, we make progress on identifying the mechanism underlying the state dependence of the effects. For example, higher crowding out in tight labor markets may explain our findings, consistent with the model of Michaillat (2014).[4] Additional demand for PV installations might draw workers from other jobs and put upward pressure on wages in tight markets, while it creates new jobs in slack markets. Another mechanism could be that in a tight market, companies substitute labor for capital. For example, they might use a machine instead of labor to move solar panels to roofs. Lastly, in slack markets additional demand in a county might be accommodated by an increase in employment in the same county, while in tight markets it might lead to an increase in employment mainly in other regions.

The evidence is most consistent with the crowding-out mechanism put forward by Michaillat (2014). First, a county's employment is largely unaffected by additional demand in surrounding regions, irrespective of the state of the county's labor market. This excludes regional spillovers as a mechanism. Second, the differential employment gains in slack and tight markets are driven both by the directly exposed sector installing PV systems as well as indirectly exposed local

[3] Abusing terminology slightly, a local labor earnings multiplier of x means that a *local* investment of the yearly gross median wage delivers x additional local jobs. This figure accounts for the fact that of the total costs of installing rooftop PV systems only one third accrue locally, while two thirds are spent on tradable parts (solar panels and components).

[4] In the search and matching framework of Michaillat (2014), higher crowding out in tight markets is a consequence of a convex quasi-labor supply function. Roulleau-Pasdeloup (2017) suggests a similar mechanism based on search frictions and downward-rigidity in prices. Recent work by Rendahl (2016) also emphasizes the role of the labor market for generating differential effects of fiscal spending in booms and recessions, albeit with a different channel. The mechanism in this paper relies on differential expectations regarding the path of the entire economy at and away from the zero lower bound. For most of our sample these differential expectations are likely not present and, hence, cannot explain the results.

service sectors. This rules out a simple substitution of labor and capital. Finally, consistent with crowding-out, there is evidence that investments lead to additional wage growth in the construction sector in tight markets, but not in slack markets.

Taken together, we provide evidence that the employment effects of PV investment in Germany are state-dependent and that crowding-out is the mechanism behind this finding. Yet, it is important to keep in mind that we estimate local multipliers from relative employment gains across regions. In consequence, our estimates do not account for the general equilibrium effects of monetary policy, Ricardian equivalence, and regional spillovers. For this reason, local multipliers tend to be larger than the national multiplier in theory (Nakamura and Steinsson, 2014; Farhi and Werning, 2016), and converting local into national multipliers requires additional assumptions (see, e.g., Chodorow-Reich, forthcoming).[5]

This chapter is thus complementary to the time-series literature on the state-dependence of multipliers (Auerbach and Gorodnichenko, 2013, 2012; Bachmann and Sims, 2012; Fazzari et al., 2014; Caggiano et al., 2015; Biolsi, 2017; Ramey and Zubairy, 2018). Specifically, in addition to this chapter, only Auerbach and Gorodnichenko (2013) investigate potential channels for state-dependence. In a cross-country setting, they show that at times of low output growth, employment responds stronger to demand shocks while wage growth is muted. The opposite is true at times of high output growth, also pointing towards crowding-out as a likely mechanism.

This chapter adds more directly to the literature on local multipliers. So far, this literature is primarily concerned with estimating the unconditional multiplier.[6] Our chapter and recent work by Dube et al. (2018) are the first to focus on its

[5]See, however, Dupor and Guerrero (2017), who show empirically that the difference between the local and aggregate multipliers of military spending is small.

[6]See Fuchs-Schündeln and Hassan (2016) and Chodorow-Reich (forthcoming) for recent reviews.

state-dependence, although some papers evaluate the state-dependence of their estimates in auxiliary analyses (Cohen et al., 2011; Brückner and Tuladhar, 2014; Nakamura and Steinsson, 2014; Shoag, 2015; Suárez Serrato and Wingender, 2016; Adelino et al., 2017).[7] The majority of these papers find—as we do—that multipliers are larger if production inputs are slack.

In relation to these papers, we make progress on three fronts. First, this chapter is unique in ruling out price and composition effects as potential confounding factors for state-dependent multipliers. Ruling out local price effects is important, as otherwise the same nominal amount of spending may constitute a different real shock. Ruling out composition effects is important, as the state of local economies may directly affect the composition of public funds.[8] Second, we are the first to provide evidence for crowding-out as the mechanism for larger multipliers in slack than in tight labor markets. Third, we demonstrate the robustness of our results with respect to a wide array of different ways of classifying slack and tight markets.

While PV installations and their constant composition are ideal to test the state-dependence of the multiplier, it is unclear whether they are an effective tool to stimulate local demand: The scope for installing PV systems is limited; once one system is mounted to the roof there is little economic rationale to add another one. Furthermore, two thirds of the investment costs accrue to the tradable system components which are mostly produced overseas. Nevertheless, to the extent that our findings are informative about the job gains from small-scale construction activity, they suggest that such undertakings may be a viable stimulus at times of slack (see also Buchheim and Watzinger, 2017).

[7] Specifically, Dube et al. (2018) show that the projects financed by the American Recovery and Reinvestment Act during the Great Recession had a stronger effect on employment and earnings in counties hit harder by the recession.

[8] This is an important concern if local governments have discretion in how to best use available funds. This applies to the work by Shoag (2015), Suárez Serrato and Wingender (2016), and Adelino et al. (2017), who all instrument spending with changes in the local government's budget constraints.

The remainder of this chapter is structured as follows. Section 3.2 describes the institutional background and the data. We lay out the empirical approach in Section 3.3, where we also discuss the identifying variation in PV installations. Section 3.4 reports the main results. Section 3.5 discusses potential mechanisms and Section 3.6 concludes.

3.2 Institutional Background and Data

The German Renewable Energy Act (*Gesetz für den Vorrang Erneuerbarer Energien*) went into effect on April 1st 2000 with the aim to increase the share of renewable energy in German energy production. The current target is that 80 percent of German electricity consumption stems from renewable energy sources by 2050. In order to achieve this, the law rests on two key mechanisms: First, the law mandates grid operators to connect all (household) renewable energy systems to the grid and to purchase the produced electricity. Second, this electricity is remunerated with a guaranteed feed-in tariff specified in the law. The relevant feed-in tariff for a given source of renewable energy is determined at the time at which it is connected to the grid and remains fixed for 20 years thereafter. In other words, for an existing renewable power plant (e.g., a PV system), the feed-in tariff cannot be changed retroactively. The feed-in tariff is independent of the market price for electricity, but has always been considerably higher than the market price during the sample period.

3.2.1 Physical Investments in Rooftop PV Systems

An additional provision of the Renewable Energy Act is that until July 2014 it mandated grid operators to collect and publish data on all renewable energy power plants. The data provided by the grid operators has been aggregated, cleaned, and validated by the *Deutsche Gesellschaft für Sonnenenergie* (DGS), which is the German branch of the International Solar Energy Society. We use the data up to and including 2012 as data postings have become more sketchy in 2013 ahead of the change in the data publishing requirements.

Every entry in the DGS database contains the type of the renewable energy power plant (photovoltaics, biomass, wind, hydropower, geothermal), the exact street address, the date of commissioning, and the power output capacity. The date of commissioning determines the applicable feed-in tariff (which, thereafter, is fixed for 20 years). For this reason, plant operators usually commission each system as soon as it is connected to the grid, as the feed-in tariff has been falling over time. We can thus exactly pinpoint when the installation of a PV system was finished. The power output capacity, in turn, is a measure of the size and, hence, the physical investment volume of a PV system. For PV systems, output capacity is measured by its peak energy production under ideal working conditions, denominated either in kilowatt peak (kWp) or megawatt peak (MWp, with 1 MWp equal to 1,000 kWp).

We restrict attention to PV systems with an output capacity of less than 500 kWp that the DGS has not deemed to have errors such as wrong addresses.[9] This restriction is due to the fact that very large systems are usually mounted on the ground by few specialized companies, so that their installation does not affect local labor demand.[10] In addition, our instrumental variable is based on the net present value of rooftop systems and does not predict systems mounted on the ground.

From these raw data entries, we construct yearly physical investment in PV systems at the county level as the sum of installed capacity within a county in a

[9]A rule-of-thumb is that 1 kWp of power output capacity requires around 8-10 m^2 of space, implying that 500 kWp require around 4,000 to 5,000 m^2 (43,056 to 53,820 square feet) of space.

[10]This restriction affects fewer than 1 percent of PV systems installed between 2000 and 2012.

given year. The main advantage of measuring physical investments, as opposed to monetary investments, is that physical investments capture real labor demand irrespective of variation in the prices of the production factors over time or across space. This is particularly relevant here, given that the price of solar panels varies considerably over time and given that the relevant wages may vary across space conditional on the state of the local labor market.

3.2.2 Determinants of Rooftop PV Installations

The volume of PV installations over time and across space is determined by five main factors: total costs, the feed-in tariff, the prevailing interest rate, solar radiation, and rooftop space. Total costs are obtained from an industry survey (Bundesverband Solarwirtschaft e.V., 2012) that asks a representative sample of 100 companies about their total installation price per kWp. The resulting cost data are available quarterly since 2006. Prior to 2006, we use the yearly data from Janzing (2010). According to this data, the average installation costs of PV systems in our sample amount to €3,121 per kWp. We use this figure to calculate the costs of job creation based on the estimated employment effects of physical PV installations.

The second investment determinant is the feed-in tariff for electricity from photovoltaics specified by the Renewable Energy Act. The feed-in tariff typically varies by year; when there are multiple changes during a year (as in 2009 and 2012), we take the yearly average.[11]

The revenue flow from selling solar electricity at the price of the feed-in tariff accrues over time. The net present value of these revenues are calculated by discounting the expected payments using the "effective interest rates of commercial

[11] The capacity bins that determine the exact feed-in tariff for each PV system have been changed in April 2012. Between April and December 2012, we use the feed-in tariff applicable to PV systems with a capacity of less than 10 kWp. Prior to April 2012, we use the feed-in tariff applicable to systems with less than 30 kWp capacity.

banks for housing loans," published by the Bundesbank.[12]

The profitability of PV systems is driven by their energy production, which is a function of the amount of solar radiation and the available rooftop space. Data on solar radiation is taken from the Photovoltaic Geographical Information System of the European Union (Huld et al., 2012). From the grid cell data on the "yearly average global irradiance on the optimally inclined surface," we calculate the average radiation (measured in kWh per m^2) at the county level.

For estimating the rooftop potential for solar energy production (in kWp), we follow the methodology of Lödl et al. (2010), who provide a detailed estimate of rooftop potential for the state of Bavaria. Lödl et al. first classify municipalities into four categories ("very rural," "rural," "suburban," and "urban") based on five observable municipality characteristics: population, population density, settlement area, average living area per capita, and the number of apartments per building. Second, they use aerial maps of 4,500 dwellings to estimate the average rooftop potential conditional on the settlement area and the municipality's type. We apply their classification of municipalities to Germany and compute each municipality's rooftop potential using the conditional estimates from Lödl et al. (2010). Rooftop potential at the county level is given by the aggregate of these municipality-level estimates. Appendix C.1.2 provides a detailed description of the calculations.[13]

[12]This series starts in January 2003. For the calculation of the net present value in Panel (b) of Figure 3.1, we use the series "average interest rates for mortgage loans" prior to 2003. This latter series has been discontinued in June 2003.

[13]A preferable approach would be to use building-level estimates of rooftop potential like Google's project "Sunroof." Google sunroof has been rolled out in Germany in May 2017, but only includes data for major municipalities thus far (`https://www.eon-solar.de/`, last accessed on March 12th 2018).

3.2.3 Employment Data and Control Variables

The data on employment, unemployment, and wages is from the Federal Employment Agency, which collects this administrative data to determine social security contributions and eligibility. The employment data counts every employed individual who lives in a county and pays social security contributions, including part-time workers but excluding the self-employed and public servants. In our main analyses, we use the yearly mean of the quarterly available employment data measured on the last day of the quarter. The industry-specific employment data on the three-digit industry level is measured at the end of the second quarter and the monthly gross median wage in construction is measured at the end of the fourth quarter.

The covariates are either from the Federal Employment Agency or the Federal Statistical Office, unless noted otherwise. The data on county types ("non-city" or "city," where "city" is a county consisting of a single municipality) and on spatial planning regions are from the Federal Office for Building and Regional Planning.

In the empirical analyses, all variables are measured annually at the county level and normalized by a county's working-age population (between 15 and 65 years of age) in 2003 unless noted otherwise. This normalization facilitates the comparison of variables across counties. Appendix C.1.1 provides further details regarding the data.

3.3 Empirical Model

The goal of our empirical strategy is to assess whether the effect of physical investments in rooftop PV systems on employment differs conditional on the state of the labor market. We identify the effect of investment in PV systems on employment by exploiting variation in installations within German counties from 2003 to 2012 using the following model

$$\begin{aligned} \textit{Employment p.c.}_{c,t} =& \beta \, \textit{PV Installations p.c.}_{c,t} + \textit{CountyFE}_c + \textit{Controls}_{c,t} \\ &+ \delta_{c,t} \, \mathbb{1}[\textit{Year}_t \cdot \textit{State}_c \cdot \textit{CountyType}_c] + \epsilon_{c,t}, \end{aligned} \tag{3.1}$$

where the index c denotes the county, t denotes the year, and "p.c." (for "per capita") in the variable name indicates that the variable is normalized by the county's working-age population measured in 2003. PV installations are measured in megawatt peak (MWp).

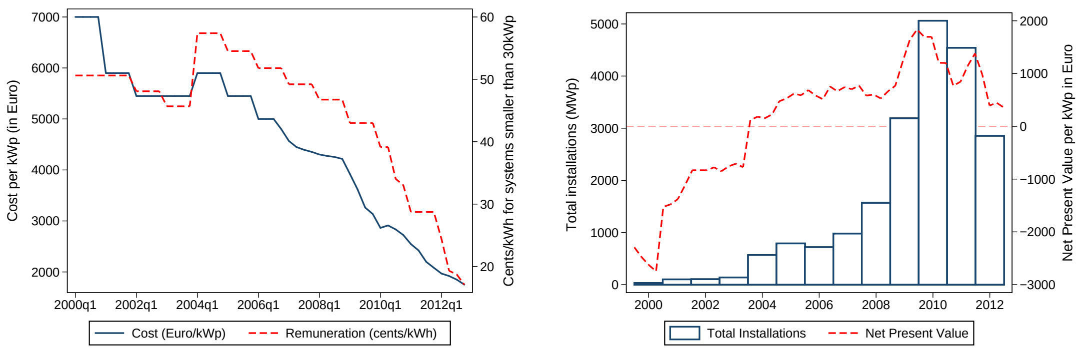
We control for county fixed effects and year fixed effects for each county type and state combination (given by $\delta_{c,t} \, \mathbb{1}[\textit{Year}_t \cdot \textit{State}_c \cdot \textit{CountyType}_c]$).[14] To adjust for labor market dynamics due to population flows, we control for population growth via the ratio of the working-age population in year t and the working-age population in 2003. We also account for construction activity as measured by the number of buildings completed in year t. Construction activity is likely to both affect the demand for rooftop PV installations and employment. We show in Appendix C.3 that our results are unchanged for different sets of covariates. The standard errors are clustered at the level of 94 German spatial planning regions to account for potential geographic and serial correlation within these regions.[15]

In our main specification, we estimate equation (3.1) using OLS, separately for

[14]Counties are defined according to their boundaries in 2012, resulting in a total of 402 counties. We omit Hamburg and Berlin, as these are city-states and their employment outcomes are fully captured by the year fixed effects at the state level.

[15]Note that model (3.1) does not allow for dynamic effects of PV installations. Because PV installations are correlated over time, empirically separating their contemporaneous and lagged effects is difficult. Yet, there are reasons to believe that dynamic effects are small: First, the measure of installed solar power capacity captures only works that have been completed in a given year, and installing a PV system rarely takes longer than a few weeks. The direct employment effect should hence be confined to the same year. Second, Appendix C.2.1 follows an approach suggested by Shoag (2015) to provide evidence suggesting that dynamic effects are indeed negligible.

Figure 3.1: Feed-in Tariff, Costs, Installations, and Net Present Value of PV Systems



(a) Feed-In Tariff and Installation Costs (b) Total Installations and Net Present Value

Note: Panel (a) shows the average costs (in Euro) per kWp for PV systems smaller than 100 kWp (solid line, left axis) as well as the feed-in tariff (fixed for 20 years, in Euro-Cents) per kWh of energy produced by a PV system smaller than 30 kWp conditional on when the system is connected to the grid (dashed line, right axis). Panel (b) displays the total installations of PV systems (measured in MWp) with less than 500 kWp capacity (bars, left axis) as well as the net present value (NPV) per kWp for a PV system with less than 30 kWp of capacity given the costs and the feed-in tariff from Panel (a) (dashed line, right axis). See the text for details.

each sample of counties with tight and slack labor markets. In a second step, we repeat this analysis using the potential profitability of PV installations in a region as an instrumental variable for investment. The next section describes the identifying variation in PV installations across time and space as well as the construction of the instrument. Section 3.3.2. explains how we classify labor markets as being tight or slack.

3.3.1 Identifying Variation

The profitability of a rooftop PV system is determined by how much electricity can be produced, how high this electricity is remunerated, and by the costs of installing and maintaining the system. The remuneration and costs exhibit substantial variation over time, but none across space, whereas the reverse is true for the potential for electricity production.

Determinants of Investment over Time

The feed-in tariff is one of the time-varying determinants of the profitability of a rooftop PV-system. The dashed line in Panel (a) of Figure 3.1 shows the feed-in tariff per kWh in Euro-Cents (right axis) of produced electricity for rooftop PV systems with less than 30 kWp, conditional on the date the system was connected to the grid. The initial feed-in tariff was 50.62 Cents in 2000, and scheduled to decrease by 5 percent each year from 2002 onwards. However, reflecting the policy goal of the government at the time to boost renewable energies, the feed-in tariff was raised to 57.40 Cents in 2004, with yearly degressions of 5 percent in 2005 and of 6.5 percent thereafter. The ensuing boom of solar energy production led to a steep increase in the cost of the policy. Further amendments of the law in 2009 and 2012 aimed to keep these costs in check, prescribing steeper degressions conditional on the volume of new installations in the previous year.

Falling costs of PV systems have also contributed to making PV installations more affordable, as illustrated by the solid line in Panel (a) of Figure 3.1. Costs have declined steeply from €7,000 per kWp in 2000 to less than €2,000 per kWp in 2012. The drop in costs mainly reflects the global decline in the price of the system components (solar modules, power inverters). This decline has been caused by technological progress and higher competition due to the market entry of Asian manufacturers.

The increased feed-in tariff combined with rapidly falling costs made it profitable to invest into rooftop PV systems in most German regions. This is illustrated in Panel (b) of Figure 3.1, which depicts the net present value for each kWp of credit-financed PV installations in a county with median solar radiation (dashed line) as well as total annual installations in MWp (bars).[16] In counties with median

[16]Most rooftop PV systems are, at least in part, credit financed (Bickel et al., various years). We assume that the relevant interest rate for financing PV installations is similar to the one for mortgages, as the PV system and its relatively risk-free income stream can serve as collateral. The

radiation and above, investing into rooftop PV systems became profitable with the increase in the feed-in tariff in 2004. The steep decline in costs in 2009 made the investment very profitable, particularly before lawmakers reacted by reducing the feed-in tariff accordingly. Yearly PV installations closely track the time variation in the profitability of these investments. In 2004, photovoltaic systems with 600 MWp were installed, more than in all previous years combined. After 2004, the upward trend continued until its peak in 2010, when yearly installations reached 5,000 MWp.[17]

Determinants of Investments across Space

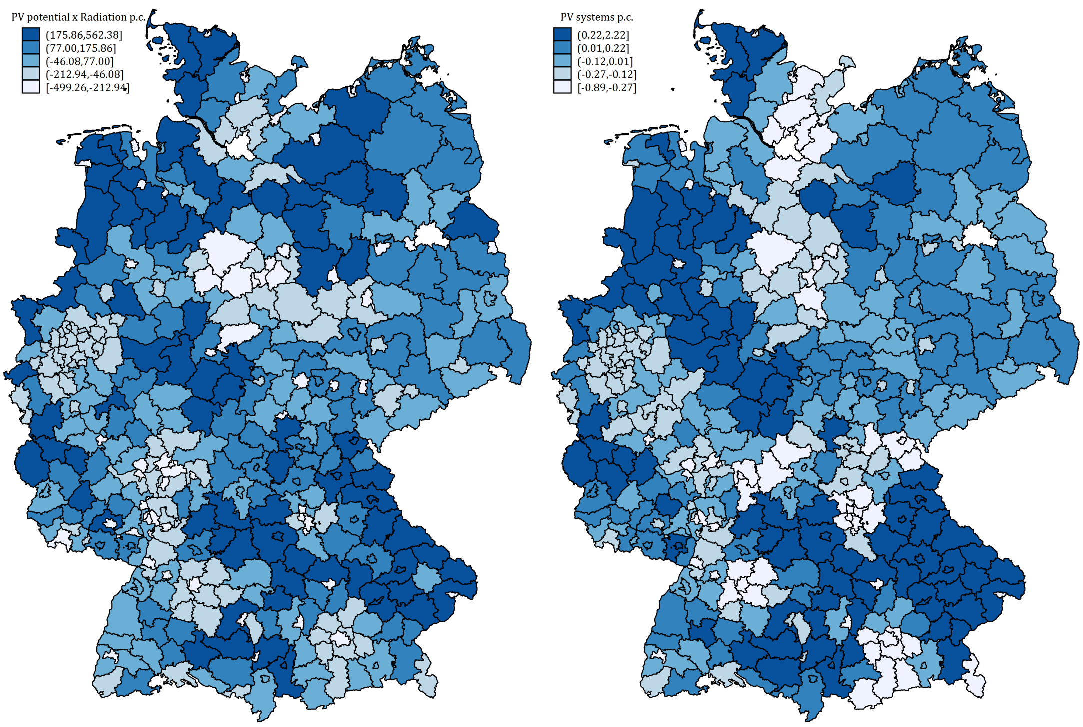
As the feed-in tariff and the costs of PV systems are (roughly) equal across German regions, the extent to which counties may benefit from installing rooftop PV systems depends on the local potential for electricity production. The latter is, in turn, a function of the local amount of solar radiation and local rooftop potential, the space available and suitable for PV installations. Because the electricity produced by a PV system is proportional to the product of radiation and the amount of space covered with solar panels, a county's potential gain from PV installations is proportional to the product of the county's yearly radiation (measured in kWh per m^2) and its total rooftop potential (measured in kWp).

Panel (a) of Figure 3.2 shows the spatial distribution of *rooftop potential* $\times$ *radiation*, normalized by the working-age population in 2003 and relative to the state-specific mean, across counties. While radiation is generally higher in the South, there is substantial variation in whether counties are more or less suitable for rooftop PV installations across all parts of Germany. Panel (b) of Figure 3.2,

average yearly interest rate on new mortgages has fluctuated between 4 and 6.5 percent between 2000 and 2009, and has been dropping to below 3 percent between 2009 and 2012. Equation (3.2) below gives the exact formula for the net present value.

[17] For comparison, one reactor of a typical nuclear power plant produces between 500 MW and 1,500 MW of electrical power.

Figure 3.2: Geographic Distribution of Total Installations and Rooftop Potential $\times$ Radiation



(a) Radiation $\times$ Rooftop Potential p.c. (b) Total Installations p.c.

Note: Panel (a) shows the geographic distribution of *rooftop potential* $\times$ *radiation* per capita (p.c.) across counties relative to their state-specific mean. Panel (b) depicts the *total power output capacity* (in kWp p.c.) installed across counties between 2003 and 2012 relative to the state-specific mean. The city-states of Hamburg (blank county in the North) and Berlin (blank county in the North-East) are excluded. The color coding scheme corresponds to quintiles of installations and *rooftop potential* $\times$ *radiation*; darker colors indicate higher values. Per capita values are normalized with the working-age population in 2003.

in turn, depicts the spatial variation of total capacity installed during the major expansion of installations between 2003 and 2012, normalized by the working-age population and relative to the state-specific mean, as before. Comparing the variation in *rooftop potential* $\times$ *radiation* and PV installations, it becomes clear that counties with a higher suitability for PV installations in general also experience a larger increase in their solar power capacity.

Remuneration Potential

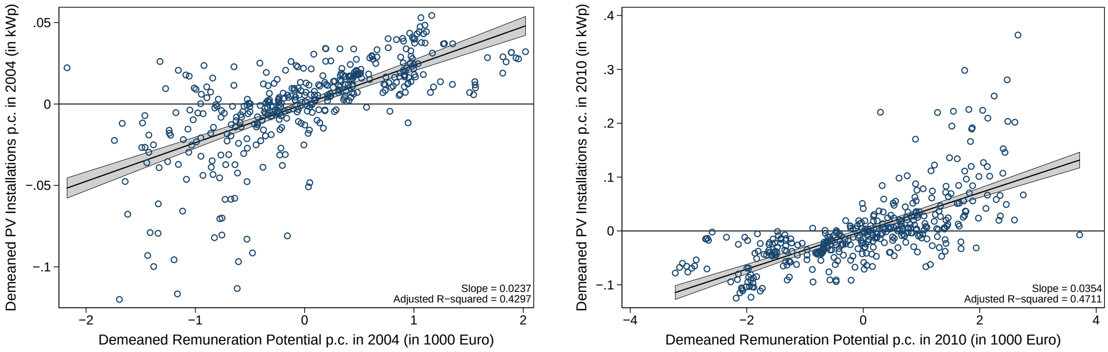
The time-varying costs and benefits of installing PV systems can be combined with the regional productivity of PV systems in producing solar energy into a single measure that captures the time-variation in local profitability of PV installations. This measure, which we call "*remuneration potential*" and which we use as an instrument for investments in Section 3.4.2, is the net present value of investing in PV systems with an output capacity equal to the county's rooftop potential in a given year t. Formally, the remuneration potential of county c in year t is defined as follows:

$$\begin{aligned}&\mathit{Remuneration\ Potential}_{c,t} = \mathit{Rooftop\ Potential}_c \cdot \\ &\left[\sum_{\tau=t}^{t+20}\left(\left(\frac{1}{1+i_t}\right)^{\tau-t}\left(\underbrace{0.995^{\tau-t}\cdot 0.75\cdot \mathit{Radiation}_c}_{\text{electricity prod. by 1 kWp system}}\cdot T_t - \underbrace{0.01\cdot C_t}_{\text{op. costs}}\right)\right) - C_t\right].\end{aligned} \tag{3.2}$$

Remuneration potential is the product of the rooftop space suitable for PV systems, measured in kWp, and the net present value of operating a PV system with output capacity of 1 kWp for 20 years from year t onwards.[18] The net present value, the term in brackets, is given by the net income stream (discounted using the interest rate at t, i_t) less the installation costs at t, C_t. The net income stream, in turn, equals the electricity production times the feed-in tariff, T_t, where we

[18]We assume the PV system to be operational for 20 years, as this is the time for which the feed-in tariff remains fixed.

Figure 3.3: Remuneration Potential and PV Installations



(a) In 2004

(b) In 2010

Note: This figure plots demeaned PV installations per capita (relative to the county and state × year specific mean) against the identically demeaned remuneration potential per capita as defined by equation (3.2). Panel (a) plots these data for the cross-section of counties in 2004, and Panel (b) shows the equivalent data for 2010.

need to adjust the power output capacity under optimal conditions for average working conditions. Here, we follow the European Union's PVGIS and assume a performance ratio of 0.75. Following Wirth (2015), we also adjust for gradual performance losses of 0.5 percent per year and annual operating costs of 1 percent of the installation costs.

Figure 3.3 shows that remuneration potential is a strong predictor of investments in PV systems.[19] It plots PV installations per capita, demeaned by their 2003 to 2012 county mean and relative to the state × year average, against the similarly demeaned remuneration potential per capita. Panel (a) shows the data for the start of the PV investment boom in 2004, and Panel (b) shows the data for the peak of the boom in 2010. In both years, variation in remuneration potential explains roughly half of the variation in PV installations. The slopes imply that a €1,000 increase in per capita remuneration potential is associated with an increase in per capita installations of 0.024 to 0.035 kWp. Given the installation costs in 2004 and 2010, the latter correspond to additional investments worth roughly €140 and €100, respectively.

[19]This echoes the formal first stage results in Appendix C.2.2.

3.3.2 Classification of Slack and Tight Labor Markets

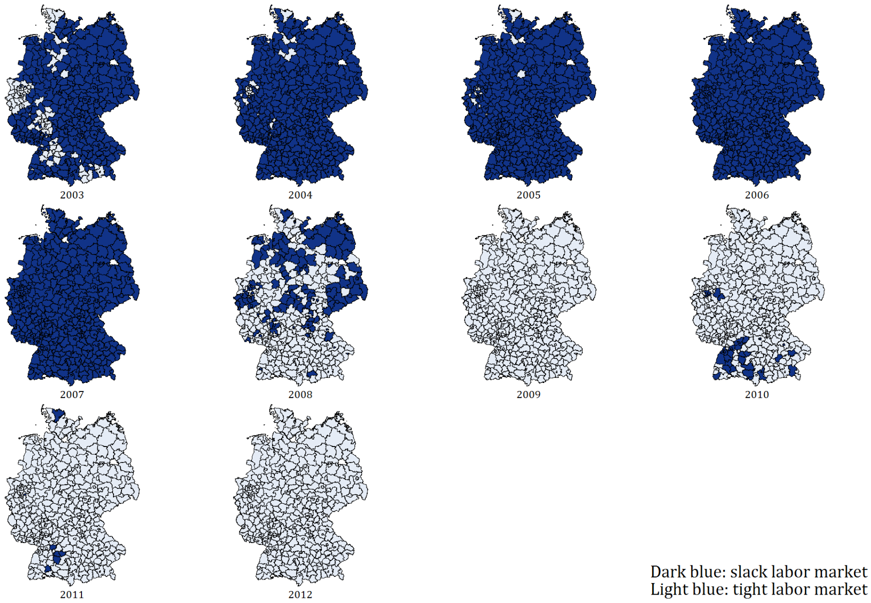
In order to investigate whether the employment effect of PV installations depends on the state of the economy, we need to define whether a labor market is "slack" or "tight." We build on the approaches of Nakamura and Steinsson (2014) and Shoag (2015). In particular, we define the state of the labor market of a county at time t as slack if the county's unemployment at time $t-1$ is above a benchmark. Otherwise, the county's labor market is defined as tight. Given this definition, the choice of the benchmark specifies in which dimension the sample is split into counties with slack and tight labor markets. For each sample split, we then estimate equation (3.1) separately for each subsample and test whether the employment effects of investments differ between the subsamples.

For the main specifications, we apply two definitions of the benchmark unemployment level that separates slack and tight labor markets. The first definition follows Nakamura and Steinsson (2014) and splits the sample along the time series dimension. This exploits the long, ten year period, during which there were stable and favorable conditions for investments in PV systems. According to this *time-series split*, a county's labor market is defined to be slack if its unemployment in the previous year is higher than the county's mean unemployment between 2003 and 2012.[20] Panel (a) of Figure 3.4 shows that, according to the time-series split, most counties' labor markets are defined to be slack prior to 2008 and tight thereafter. This reflects the downward sloping trend in German unemployment over this period.

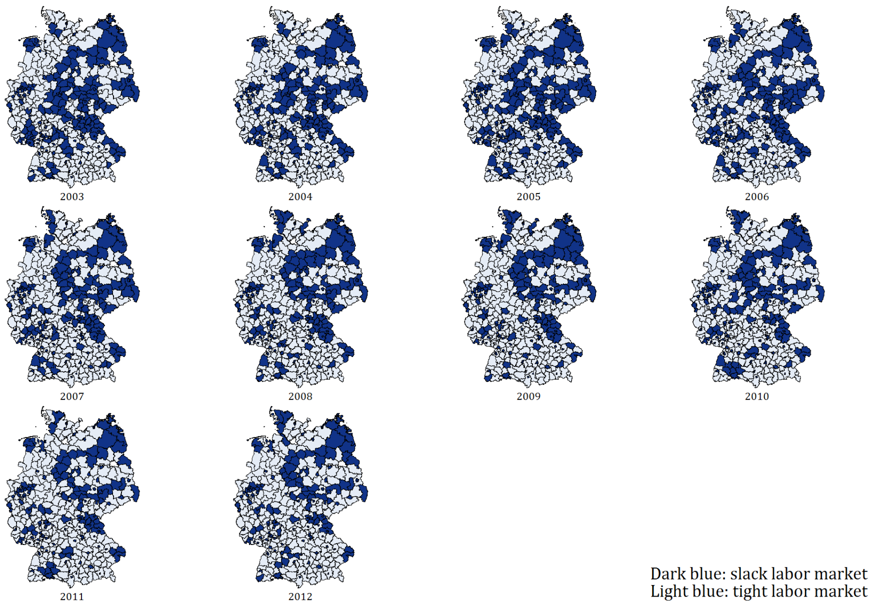
The second definition of the unemployment benchmark follows Shoag (2015) and splits the sample along the cross-sectional dimension. According to this *cross-sectional split*, a county's labor market is defined to be slack if its unemployment

[20]Formally, the labor market is said to be slack according to the time-series split if $unemployment_{c,t-1} > 1/10 \sum_{t=2003}^{2012} unemployment_{c,t}$.

Figure 3.4: Labor Market Slack across Counties



(a) Time-series Classification of Slack



(b) Cross-sectional Classification of Slack

Note: Panel (a) shows, for each year in the sample, which counties are classified as having a slack or tight labor market according to the definition of slack in the time-series dimension. Here, the labor market is defined to be slack if a county's unemployment in the previous year is higher than the mean unemployment of the county over the sample period. Panel (b) shows which counties exhibit slack / tight labor markets according to the definition of slack in the cross-sectional dimension. Here, the labor market is defined to be slack if a county's unemployment in the previous year is higher than the state average of unemployment in the previous year.

in the previous year is higher than the state mean of unemployment in the same year.[21] Panel (b) of Figure 3.4 shows that the cross-sectional split selects a similar set of counties to have slack and tight labor markets in the different years.

The two sample splits have different implications. The time-series split compares the same set of counties at different times, so that the two samples of slack and tight labor markets, respectively, share the same structural features. The cross-sectional split compares different counties at the same time, thus holding constant all factors that may affect the employment response to investments over time (such as innovation in the production technology). Hence, differential employment effects in the sample splits along both dimensions can neither be explained by time trends nor structural features alone. To further bolster the robustness of the results with respect to the definition of slack and tight labor markets, Section 3.4.3 explores the employment response to investments for a wide range of additional definitions.

3.4 Results

This section presents the main findings. The empirical analysis shows that physical investments increase employment more at times and in regions with slack labor markets compared to times and regions with tight labor markets. This result holds irrespective of whether we estimate the employment effects of investments via OLS or IV, and independent of the particular definition of slack and tight labor markets.

Table 3.1: The Effect of PV Installations on Employment (OLS)

	Employment Rate				
Split along		Time series		Cross-section	
	Baseline (1)	Slack (2)	Tight (3)	Slack (4)	Tight (5)
Installed capacity p.c.	19.98***	36.61**	2.78	37.91***	13.34**
	(6.58)	(17.09)	(3.86)	(13.83)	(6.47)
Population growth	0.37***	0.37***	0.25***	0.31***	0.40***
	(0.02)	(0.03)	(0.02)	(0.04)	(0.03)
Construction p.c.	−0.25	−0.19	0.47	0.04	−0.23
	(0.22)	(0.14)	(0.30)	(0.21)	(0.26)
County fixed effects	yes	yes	yes	yes	yes
Year fixed effects	yes	yes	yes	yes	yes
P-value slack < tight		0.017		0.036	
Jobs per €100,000	0.64	1.17	0.09	1.21	0.43
Observations	4000	2044	1956	1783	2189

Note: The dependent variable is the average yearly employment rate (employment normalized by the working-age population in 2003) between 2003 and 2012. *Installed capacity p.c.* are yearly PV installations measured in megawatt peak (MWp) normalized by the working-age population in 2003 (indicated by "p.c." for "per capita"). *Population Growth* is the ratio of the working-age population in a given year to the working-age population in 2003. *Construction p.c.* is the number of residential and non-residential buildings completed in a given year. *P-Val slack < tight* reports the p-value of the test of the null hypothesis that the employment effect of PV installations is smaller in slack than in tight labor markets. The *year fixed effects* are estimated at the level of the state × county type (rural or urban county). Note that the observations in the cross-sectional split do not sum to 4000 due to singleton groups. Standard errors (in parentheses) are clustered at the level of 94 spatial planning regions. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

3.4.1 OLS Results

Table 3.1 presents the OLS estimates of empirical model (3.1). As a benchmark, column (1) reports the average effect of physical investments in PV systems on employment for the full sample. Because both employment and installations are normalized by the working-age population, the coefficient of *installed capacity p.c.* can be interpreted as the number of additional jobs per MWp of PV installations. Hence, in the full sample, additional PV installations of 1 MWp capacity lead to, on average, around 20 additional jobs lasting for one year. Given the average installation costs of €3.121m per 1 MWp capacity, this estimate implies that investments of €100,000 created 0.64 local job-years, corresponding to costs per job-year of €156,000.

The magnitude of this estimate is in line with how much work is required to install rooftop PV systems, the direct effect of additional demand on the labor market: The industry survey of Brohm (2010) suggests that, in 2008, installing capacity of 1 MWp required on average between 15 and 20 workers. We also have access to the internal accounting data of one installation company. Their data suggests a lower bound of 9 worker-years per MWp.[22]

Columns (2) and (3) report the estimates for the sample split into slack and tight labor markets along the *time series* dimension. Here, we find that at times of economic slack, additional PV installations of 1 MWp capacity lead to about 37 additional job-years, corresponding to 1.17 job-years per investments of €100,000. This effect is 80 percent larger than the baseline effect in column (1). In contrast,

[21] Formally, the labor market is said to be slack according to the cross-sectional split if $unemployment_{c,t-1}/N_{c,2003} > (\sum_{c \in state(c)} unemployment_{c,t-1})/(\sum_{c \in state(c)} N_{c,2003})$, where $N_{c,2003}$ is county c's working-age population in 2003.

[22] This number includes the installation of the panels, the acquisition and the planning of the PV system as well as the additional jobs at the regional distributor. It does not include jobs outside of the construction sector. The 9 worker-years are within the 95 percent confidence interval of our OLS estimate in column (1). If we only consider employment in the high exposure sector—as we do in Table C.9 in Appendix C.4.2—we arrive at 15.04 job-years per MWp with a standard error of 4.74. Here, 9 job-years are within the 80 percent confidence interval.

at times of tight labor markets, the local employment effect of investments into PV systems is economically small and statistically indistinguishable from zero. Moreover, we reject the null hypothesis that the investment-induced employment gains are larger at times of tight labor markets at the five percent level; the p-value of the respective one-sided test equals 0.017.

The *cross-sectional* split in columns (4) and (5) leads to similar results as the time series split. Here, 1 MWp in PV installations leads to 38 more job-years in counties with slack labor markets, while the same additional demand creates only 13 new job-years in counties with tight labor markets. As before, the null hypothesis that employment gains in regions with tight labor markets are larger than in ones with slack labor markets is rejected at the five percent level.

The differences in the employment effects across the two sample splits cannot be explained by either counties' structural characteristics or time-varying changes in the relation between real demand and employment alone. While a correlation of local investments and structural labor market characteristics—such as higher investments in less sklerotic labor markets—may explain the difference in the employment effects in the cross-sectional split in columns (4) and (5), such an explanation cannot account for the difference in the employment creation over time in the identical sets of counties in the time series split. Conversely, we have seen in Panel (a) of Figure 3.4 that local labor markets were mostly slack in the first years of the photovoltaic investment boom and tight in the later years, so that the results in columns (2) and (3) could potentially be explained by a reduction in labor demand for new installations.[23] However, changes in technology cannot account for the results of the cross-sectional split, as the latter compares different counties with slack and tight labor markets at the same time. Additionally, the next section uses an instrumental variable approach to provide additional

[23]Contrary to this hypothesis, conversations with industry experts suggest that there was no fundamental change to the technology for installing PV systems over time.

evidence that unobserved third factors are unlikely to drive the results of Table 3.1.

To interpret the magnitude of the employment effect, we approximate the local labor earnings multiplier. To do so, we divide the median annual earnings by the costs per job-year implied by our estimates. Median earnings equaled €32,160 during the sample period. The results in column (1) of Table 3.1 imply costs per job-year equal to €100,000/0.64 = €156,000. Combined, this translates into a local labor earnings multiplier of 0.21, i.e., local earnings increased by €0.21 per €1 investment in PV systems. In slack labor markets, this multiplier is around 0.38, while in tight labor markets the multiplier is at most 0.14. Finally, it is important to note that only around a third of the total costs of PV systems, the basis for the calculation of the costs per job-year, accrue locally.[24] If we scale our estimates accordingly, the multiplier in tight labor markets is around 1.14, while in slack labor markets the multiplier is at most 0.42.

3.4.2 IV Results

One concern for identifying the effect of PV installations on employment is that investment decisions may depend on expected labor market dynamics via an unobserved third channel, such as local credit markets. As many rooftop PV systems are credit-financed, changes in local lending might influence both, employment and investments and thus bias the estimates in either direction.[25] For

[24]According to an industry survey in 2013, the local installation costs amount to about 20 percent of the total costs, while the remaining 80 percent are spent on solar panels and components (EuPD Research, 2013). According to anecdotal evidence from industry experts, installation firms charge an additional 10 percent of the total costs as a mark-up on the panels and components, so that roughly one third of the total costs contribute to local demand.

[25]Financial service provision in Germany has a strong regional focus due to the nationwide presence of local savings banks (*Sparkassen*) and credit cooperations (*Volks- und Raiffeisenbanken*). In 2012, there were 423 savings banks, the area of business of which is often defined by county borders, and more than 900 credit cooperations. Statistics on the share of debt-financing of PV systems do not exist. However, the state-owned bank *Kreditanstalt für Wiederaufbau (KfW)* reports that in the years 2007 to 2012, between 42 percent and 74 percent of the yearly investments in PV systems have been at least partially backed by their subsidized loans program (Bickel et al.,

example, OLS could overestimate the effect of PV installations on employment if favorable lending conditions drive employment growth and investment. OLS could also underestimate the effect of PV installations on employment if loans for safe investments into PV systems are particularly attractive when the local economy is on a downward trajectory.

To address such concerns, we instrument local investments into PV systems by their profitability which is captured by *remuneration potential* as defined in Section 3.3.1. *Remuneration potential* is a function of time-varying factors (the feed-in tariff, the costs of components, mortgage rates) that are determined at the global or national level and thus unrelated to the trajectories of local labor markets, as well as by pre-determined geographic characteristics (the rooftop potential, local solar radiation) that are likely fixed over time and thus unresponsive to labor market developments as well. At the same time, *remuneration potential* strongly predicts investments, as shown in Figure 3.3. Taken together, the variable *remuneration potential* hence likely meets the identifying assumptions of relevance and exogeneity. Appendix C.2.2 provides additional details regarding the first stage and further arguments for why the exclusion restriction is likely to hold.

Table 3.2 summarizes the IV estimates of the main empirical model (3.1). The IV results are qualitatively similar to the findings from OLS. The employment gains due to PV installations in slack labor markets are larger than the overall average and much larger than the employment gains in tight labor markets. At average installation costs, the estimated employment gains from physical investments imply that €100,000 in PV installations increase employment by about five job-years in slack and by about one job-year in tight labor markets, both in the time series and the cross-sectional split.[26] Due to these large differences, we reject the

various years) .

[26]The difference in the magnitude of the OLS and IV estimates could be driven by higher investment incentives in less prosperous regions that lead to a downward bias of the OLS estimates. Another part of the explanation may be that the IV estimates are, despite their strong first stage, much more noisy than the OLS estimates. Furthermore, the IV estimates a local average treatment

Table 3.2: The Effect of PV Installations on Employment (IV)

	Employment Rate				
Split along		Time series		Cross-section	
	Baseline	Slack	Tight	Slack	Tight
	(1)	(2)	(3)	(4)	(5)
Installed capacity p.c.	52.57***	148.43***	22.60**	180.11***	30.52**
	(13.60)	(45.38)	(10.86)	(37.73)	(12.01)
Population growth	0.36***	0.36***	0.25***	0.30***	0.38***
	(0.02)	(0.03)	(0.02)	(0.03)	(0.03)
Construction p.c.	−0.24	−0.20	0.48*	0.36	−0.25
	(0.23)	(0.14)	(0.29)	(0.30)	(0.27)
County fixed effects	yes	yes	yes	yes	yes
Year fixed effects	yes	yes	yes	yes	yes
P-value slack < tight		0.003		0.000	
Jobs per €100,000	1.68	4.76	0.72	5.77	0.98
F-statistic instrument	88.22	22.11	59.15	28.43	79.67
Observations	4000	2044	1956	1783	2189

Note: The dependent variable is the average yearly employment rate (employment normalized by the working-age population in 2003) between 2003 and 2012. *Installed capacity p.c.* are yearly PV installations measured in megawatt peak (MWp) normalized by the working-age population in 2003 (indicated by "p.c." for "per capita"). *Population Growth* is the ratio of the working-age population in a given year to the working-age population in 2003. *Construction p.c.* is the number of residential and non-residential buildings completed in a given year. *P-Val slack < tight* reports the p-value of the test of the null hypothesis that the employment effect of PV installations is smaller in slack than in tight labor markets. *F-statistic instrument* is the Kleibergen-Paap F-statistic of *remuneration potential p.c.* in the first stage. The *year fixed effects* are estimated at the level of the state × county type (rural or urban county). Note that the observations in the cross-sectional split do not sum to 4000 due to singleton groups. Standard errors (in parentheses) are clustered at the level of 94 spatial planning regions. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

hypothesis that employment is more responsive to investments in tight than in slack labor markets at the one percent level, even though the IV estimates are less precise than the corresponding OLS results. Finally, note that with Kleibergen-Paap F-statistics of the excluded instruments equal to 22 or higher, the first stage is strong in all specifications.

3.4.3 Alternative Classifications of Slack and Tight Labor Markets

Table 3.3 uses alternative classifications of slack and tight labor markets to show that our finding of differential employment effects does not crucially depend on the specific classification.[27] Each row of Table 3.3 reports the results of an alternative sample split. Panel A reports the OLS estimates and Panel B the corresponding IV results.

Rows (1) and (8) in boldface repeat the baseline time series and cross-sectional splits from Tables 3.1 and 3.2, respectively, for comparison purposes. Row (2) contains an alternative time series split that defines all years prior to 2007 as times of slack and all years after 2008 as times of tight labor markets. Rows (3) to (6) split the sample based on unemployment benchmarks calculated across time and space. In row (3), a county is said to have a slack labor market if its unemployment rate in year $t-1$ is above the average national unemployment rate between 2003 and 2012. Otherwise, the labor market is said to be tight. In rows (4) and (5), labor markets are classified accordingly, but with respect to the 2003 to 2012 state average and the 2003 to 2012 state $\times$ county-type average,

effect, while the OLS uses the entire variation in the data. Local average treatment effects are difficult to interpret because the compliers, for which the treatment effects are estimated, cannot be identified in the data. We thus cannot assess whether the IV results can be extrapolated to the entire population.

[27]In addition, in Appendix C.3.3 we show that for the main slack definitions, the results are comparable when we split observations into three groups instead of two.

Table 3.3: Alternative Definitions of Slack in the Labor Market

	Slack		Tight		P-Value
	Coeff	SE	Coeff	SE	
	(1)	(2)	(3)	(4)	(5)
	Panel A: OLS				
Based on time variation in unemployment					
(1) **Within county mean**	36.61	17.09	2.78	3.86	0.017
(2) 2003-2007 vs. 2008-2012	47.79	16.98	1.87	3.25	0.002
Based on aggregate variation in unemployment					
(3) 2003-2012 national mean	28.33	26.35	18.69	6.93	0.361
(4) 2003-2012 state mean	49.82	18.22	11.25	6.47	0.013
(5) 2003-2012 state $\times$ county type mean	32.09	9.38	14.16	6.99	0.049
(6) Residuals wrt. county and state $\times$ year FEs	20.46	5.60	12.50	6.01	0.058
Based on cross-sectional variation in unemployment					
(7) Yearly national mean	43.04	23.60	21.22	7.30	0.186
(8) **Yearly state mean**	37.91	13.83	13.34	6.47	0.036
(9) Yearly state $\times$ county type mean	23.56	7.58	19.61	7.17	0.296
(10) State mean in 2002	31.32	12.00	12.78	6.31	0.061
Based on aggregate variation in wage growth					
(11) National CPI growth	22.59	7.12	13.10	6.62	0.038
	Panel B: IV				
Based on time variation in unemployment					
(1) **Within county mean**	148.43	45.38	22.60	10.86	0.003
(2) 2003-2007 vs. 2008-2012	200.21	49.30	11.45	6.44	0.000
Based on aggregate variation in unemployment					
(3) 2003-2012 national mean	142.24	73.87	32.50	12.00	0.071
(4) 2003-2012 state mean	193.67	48.04	24.54	10.68	0.000
(5) 2003-2012 state $\times$ county type mean	153.87	35.26	18.44	11.38	0.000
(6) Residuals wrt. county and state $\times$ year FEs	59.97	15.48	45.32	14.85	0.167
Based on cross-sectional variation in unemployment					
(7) Yearly national mean	131.08	53.86	44.45	14.32	0.057
(8) **Yearly state mean**	180.11	37.73	30.52	12.01	0.000
(9) Yearly state $\times$ county type mean	98.36	24.63	33.41	13.37	0.006
(10) State mean in 2002	120.52	31.05	31.98	12.58	0.002
Based on aggregate variation in wage growth					
(11) National CPI growth	59.86	13.28	30.38	12.90	0.005

Note: This table presents the employment effects of PV installations in slack and tight labor markets for various alternative sample splits. Except for rows (2), (6), (10) and (11), the name of each row specifies a different unemployment benchmark (e.g., in row (3) the benchmark is the 2003-2012 national unemployment mean). The labor market is said to be slack if unemployment in $t-1$ is above the benchmark, and said to be tight otherwise. In row (2), labor markets are defined as being slack in 2007 and earlier, and tight in 2008 and later. In row (6), a labor market is slack if the residual of regressing unemployment on county and state-year fixed effects is positive in $t-1$. In row (10), a labor market is slack if its unemployment rate in 2002 was higher than the state mean in 2002, and tight otherwise. In row (11), a labor market is slack if its year-over-year growth of the median wage in construction is lower than national inflation. Panel A reports the OLS results, and Panel B reports the IV results with the same model specification as in Tables 3.1 and 3.2, respectively. Columns entitled "Coeff" report the OLS/IV coefficient estimate of *installed capacity p.c.* for the subsample with slack and tight labor market, respectively. Columns entitled "SE" report the corresponding standard errors, clustered at the level of 94 spatial planning regions. The column entitled "P-Value" report the p-values of the test of the null hypothesis that the employment effect of PV installations is smaller in a slack labor market than in a tight labor market.

respectively.[28] In row (6), we regress unemployment on county and state-year fixed effects and split the residuals from this regression at zero to determine whether a county's labor market is classified as slack or tight. Positive residuals indicate a slack market and negative residuals indicate a tight market. Rows (7) to (10) provide cross-sectional splits. In row (7), a labor market is defined as slack if its unemployment rate in $t-1$ is above the national mean in $t-1$. Slack and tight labor markets in rows (8) and (9) are defined similarly, but with respect to the state and state $\times$ county-type average in $t-1$, respectively. In row (10), a labor market is said to be slack if in 2002, the last year before the sample period, its unemployment rate was above the 2002 state mean, and said to be tight otherwise. Finally, in row (11) we use wage growth in the construction sector as a measure of slack. If wage growth lags behind national inflation, the labor market is defined as slack and tight otherwise.[29]

The estimated employment effects of investments in slack labor markets are above the corresponding estimates for tight labor markets in all specifications of Table 3.3. Except in the OLS specifications in row (9) and the IV specification in row (6), these differences are economically meaningful with the employment gains in slack market conditions being at least 50 percent larger than the ones in tight conditions. The coefficients are statistically different from each other on conventional levels in all but four of the 22 specifications.

Taken together, we consistently find that more jobs are created in slack than in tight markets. This empirical pattern is present irrespective of the exact definition of labor market slack and robust to using an IV strategy instead of OLS.[30]

[28]This means that in row (4), a county is said to have a slack labor market if its unemployment rate in $t-1$ is above the average unemployment rate between 2003 and 2012 in the same state. In row (5), the unemployment rate in $t-1$ is compared to the average unemployment rate between 2003 and 2012 within all counties of the same county type (either urban or rural) and the same state.

[29]For this definition of labor market tightness the sample size is smaller than for the other specifications as the wage data is only available for counties with more than 1,000 employees in construction due to data confidentiality requirements.

[30]Appendix C.3 provides further robustness checks regarding data and covariate choices.

3.5 Discussion of Mechanisms

There are several potential explanations for the empirical pattern of fewer jobs being created locally when the labor market is tight. First, there may be a larger incentive to substitute labor with capital. Second, installation firms might meet their labor demand with hiring workers from outside the local labor market. Third, investment in tight labor markets might lead to crowding out, as in Michaillat (2014). Workers installing PV systems might be effectively drawn from other jobs if the labor market is tight—putting pressure on wages in the process, while they might have been unemployed if the labor market is slack. Finally, there may be direct effects of labor market tightness on the costs of investment (e.g., due to higher wages for installation workers or higher markups) implying a lower effective labor demand for each Euro invested in a tight labor market. However, the latter effect cannot play a role here, as we use the physical investment volume as explanatory variable.[31]

Effects across Industries

The first channel, a substitution of labor with capital in tight labor markets, is unlikely to drive the results in the context of PV investment. The production function of installing PV systems has been stable over the entire time period. The production process consists of bringing the PV systems to the customer and workers carrying solar panels onto rooftops and mounting them there, with little scope for different installation techniques. The only substitute technology available for installation are telehandlers that can lift the panels onto roofs, which still

[31] Another potential explanation for differential effects in tight and slack markets are nonlinearities in the response of employment to PV installations. To allow for non-linear effects, Appendix C.4.1 computes the marginal effects from a variant of model (3.1) that adds the installed capacity squared as a covariate. This analysis shows that nonlinearities cannot account for the differential employment creation in slack and tight markets.

require extensive manual labor to mount and install the system. Unfortunately, we do not have data on the usage of telehandlers and hence we provide indirect evidence that the usage of telehandlers is unlikely to account for the differences in employment gains in slack and tight labor markets.

To this end, we partition total employment into employment in (i) high exposure industries, (ii) local non-tradables, and (iii) all other industries. The high exposure industries are directly affected by the demand for PV installations, such as electricians.[32] We classify the retail and wholesale sector, the hospitality industry (hotels and restaurants) as well as financial service providers as local services. Companies in these industries might benefit from local demand spillovers. All remaining industries are classified as belonging to "other industries".[33] If it were the case that the difference in slack versus tight labor markets was driven by changes in the installation technology, we would expect that the differential response is entirely driven by differences in employment gains in the high exposure industries. Employment creation in local industries should not exhibit differential effects, as the labor-saving technology is specific to the installation of PV systems.[34]

Table 3.4 reports the OLS estimates of the employment gains due to PV installations in slack and tight labor markets for each of the three sectors. Panel A presents the results for the time series split, and Panel B presents the results for

[32]For classifying industries as being "high exposure," we take a random sample of firms that are a member of the *Bundesverband Solarwirtschaft* (a trade association of the German solar industry) that install PV systems and consult their Creditreform company profiles to extract their industry classification. Most of the sampled firms are certified electricians; as such, they belong to various industries, including building installation and engineering. The union of the industry codes identified by this procedure constitutes the set of high exposure industries.

[33]"Social services" (industry code 853) is excluded from "other industries" for two reasons. First, the employment data in this industry is non-stationary, as it increases from 4.4 percent to 6.2 percent of the workforce from 2003 to 2012. Second, this sector mostly comprises of the daycare industry (care of elderly and children), and it is unclear whether these are local services or, indeed, "other industries". Table C.1 in Appendix C.1.1 lists the assignment of industry codes to each of the three subsectors.

[34]Note that this additionally assumes that the spillovers from the high exposure to the local sectors do not change with the state of the labor market.

Table 3.4: Sectoral Employment Conditional on Slack: OLS Results

	Industry-specific Employment p.c.					
	High-exposure		Local		Other	
	Slack (1)	Tight (2)	Slack (3)	Tight (4)	Slack (5)	Tight (6)
	Panel A: Time Series Split					
Capacity p.c.	17.96***	2.98**	7.67**	3.94***	20.97	−4.39
	(6.49)	(1.37)	(3.87)	(1.23)	(23.76)	(3.29)
P-val slack < tight	0.012		0.162		0.136	
Jobs per €100,000	0.58	0.10	0.25	0.13	0.67	−0.14
Observations	2044	1956	2044	1956	2044	1956
	Panel B: Cross-Sectional Split					
Capacity p.c.	17.10**	11.45***	15.18***	6.64***	5.94	−5.45
	(8.37)	(4.17)	(3.16)	(2.54)	(15.94)	(8.47)
P-val slack < tight	0.220		0.008		0.211	
Jobs per €100,000	0.55	0.37	0.49	0.21	0.19	−0.17
Observations	1783	2189	1783	2189	1783	2189
Controls	yes	yes	yes	yes	yes	yes
County FE	yes	yes	yes	yes	yes	yes
Year FE	yes	yes	yes	yes	yes	yes

Note: The dependent variable in columns (1) and (2) is employment in the high-exposure sectors (construction and related industries) normalized by the working-age population in 2003 (indicated by "p.c." for "per capita"). The dependent variable in columns (3) and (4) is employment p.c. in local industries (wholesale, retail, hospitality, local services). The dependent variable in columns (5) and (6) is employment p.c. in all remaining industries. Employment by industry is measured annually on June 30th. Table C.1 in Appendix C.1.1 provides details of the industry classifications. *Capacity p.c.* are yearly PV installations measured in megawatt peak (MWp). Except for the dependent variables, the empirical specifications are identical to the ones in Table 3.1. In particular, controls are population growth and new construction per capita. The *year fixed effects* are estimated at the level of the state $\times$ county type (rural or urban county). Panel A reports the results for the time series split, and Panel B reports the results for the cross-sectional split. Standard errors (in parentheses) are clustered at the level of 94 spatial planning regions. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

the cross-sectional definition of labor market slackness. The first finding from Table 3.4 is that in both slack and tight labor markets, PV installations led to a statistically significant increase in employment only in the high-exposure and local sectors. In contrast, the employment gains (or losses) in all other sectors are very imprecisely estimated. Given the nature of the investments we study, this is exactly the pattern of employment gains across industries we expect to find.[35]

A second result from Table 3.4 is that the difference of the employment gains in slack and tight labor markets is driven by differential employment gains both in high exposure and local industries. In the time series as well as the cross-sectional split, the difference in the employment gains between slack and tight markets is sizable, and, in two cases, significantly different from each other.[36] This speaks against adjustments in the production technology of PV installations conditional on the state of the labor market as an explanation for the differential employment gains in slack and tight labor markets.

Spillovers from Neighboring Counties

The second channel, hiring workers from outside the local labor market (a county in our setting), is also unable to fully explain our findings. First, we control for population growth in all of our regressions, which should capture migration as long as workers also change their place of residence. Second, migration and commuting cannot explain the results of our time-series split, given that almost all counties are classified as slack and tight at the same time. Third, we show next that there is no evidence for demand spillovers across regions independent of the state of the labor market, implying that counties in this setting are indeed

[35]This mirrors the result for the full sample of counties in Table C.9 in Appendix C.4.2, where we find that around 60 percent of the entire employment effect originates from the high-exposure sectors, while 40 percent originates from local industries.

[36]The corresponding IV results reported in Table C.11 in Appendix C.4.3 mirror the OLS findings.

self-contained labor markets.

To test for geographic spillovers, we follow the approach of Acconcia et al. (2014) and include investments in neighboring counties as an additional control variable. We consider three possible definitions of neighboring counties: all other counties within the same spatial planning region, the five closest counties based on the distance between both counties' most populous municipalities, and the ten closest counties. For each set of a county's neighbors we calculate the total PV installations in MWp within the set of neighboring counties and normalize the total installations with the working-age population in the county of interest. Given this, we estimate an extended version of the main empirical model (3.1) that includes aggregate PV installations in neighboring counties as an additional covariate. As in the main empirical analyses, we classify counties as having slack or tight labor markets according to their own unemployment rate as described in Section 3.3.2.

Table 3.5 reports the OLS estimates of the demand spillovers conditional on the state of the labor market as defined via the time series split (Panel A) as well as the cross-sectional split (Panel B). In both splits and in all three definitions of a county's set of neighbors, the effect of additional PV installations in geographically proximate regions is at least one order of magnitude smaller than the effect of additional installations within the county. In addition to their small magnitude, all the coefficients are statistically insignificant. The estimated effects of the demand spillovers also do not differ by much between slack and tight labor markets, while the differences of the employment gains due to the within-county investments remain at the same level as in Table 3.1, the main OLS specification.[37] All in all, the employment effects of PV installations are very local in nature, so that demand spillovers are largely unimportant for the interpretation

[37] Appendix C.4.2 shows that there is no evidence for demand spillovers in the full sample either, and Table C.12 in Appendix C.4.3 demonstrates that the IV estimates lead to the same conclusions as the OLS estimates.

Table 3.5: Spillovers from Neighboring Counties: OLS Results

	Employment Rate					
	Planning Region		5 Closest Counties		10 Closest Counties	
	Slack (1)	Tight (2)	Slack (3)	Tight (4)	Slack (5)	Tight (6)
	Panel A: Time Series Split					
Capacity p.c.	32.58**	2.09	34.28**	2.81	32.80**	1.76
	(16.39)	(4.04)	(15.79)	(4.38)	(16.23)	(4.29)
Neighboring capacity p.c.	3.05	0.43	0.72	−0.01	0.91	0.23
	(2.66)	(0.75)	(1.67)	(0.74)	(0.69)	(0.39)
P-val slack < tight	0.024		0.017		0.020	
Jobs per €100,000	1.04	0.07	1.10	0.09	1.05	0.06
Observations	2044	1956	2044	1956	2044	1956
	Panel B: Cross-Sectional Split					
Capacity p.c.	42.25***	13.46**	41.53***	15.36***	37.98***	15.56***
	(11.20)	(6.54)	(12.18)	(5.84)	(10.28)	(5.90)
Neighboring capacity p.c.	−0.86	−0.10	−0.67	−0.86	−0.01	−0.59
	(0.93)	(2.06)	(0.88)	(1.33)	(0.57)	(0.64)
P-val slack < tight	0.010		0.022		0.024	
Jobs per €100,000	1.35	0.43	1.33	0.49	1.22	0.50
Observations	1783	2189	1783	2189	1783	2189
Controls	yes	yes	yes	yes	yes	yes
County FE	yes	yes	yes	yes	yes	yes
Year FE	yes	yes	yes	yes	yes	yes

Note: *Neighboring capacity p.c.* is the sum of PV installations (measured in MWp and normalized by the working-age population) across all other counties in the same spatial planning region (columns (1) and (2)), the 5 closest counties (columns (3) and (4)), or the 10 closest counties (columns (5) and (6)). Closeness is measured by the distance between the counties' most populous municipalities. Controls are population growth and new construction per capita. The *year fixed effects* are estimated at the level of the state × county type (rural or urban county). All other variables are defined as in Table 3.1. Panel A reports the results for the time series split, and Panel B reports the results for the cross-sectional split. Standard errors (in parentheses) are clustered at the level of 94 spatial planning regions. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

of our findings.

Wage Growth

Taken together, this leaves crowding out as the most plausible mechanism. Workers installing PV systems might be drawn from other jobs if the labor market is tight, while they might have been unemployed if the labor market is slack. This is the mechanism identified by Michaillat (2014), who argues that the employment response to additional demand is a general function of the state of the labor market. In his model, diminishing returns to labor lead to a quasi-labor supply curve that is convex in labor market tightness, so that additional labor demand leads to a higher degree of crowding out in a tight than in a slack labor market.

Furthermore, in this model wages respond more strongly to a given demand shock in tight than in slack labor markets. To test this prediction, we estimate how wage growth reacts to PV installations using the following model:

$$\begin{aligned}\frac{[log(wage_{c,t}) - log(wage_{c,t-k})]}{k} &= \beta \cdot \frac{1}{k} \cdot \sum_{\tau=0}^{k-1} PV\ Installations\ p.c._{c,t-\tau} \\ &+ \gamma_1 \cdot \frac{1}{k} \cdot \sum_{\tau=0}^{k-1} Construction\ p.c._{c,t-\tau} + \gamma_2 \cdot \frac{Pop_{c,t} - Pop_{c,t-k}}{k \cdot Pop_{2003}} \\ &+ \delta_c \mathbb{1}[State_c \times CountyType_c] + \varepsilon_c.\end{aligned} \tag{3.3}$$

Here, the dependent variable is the average yearly growth of the median wage in construction in county c over k years. The independent variable of interest is the average amount of yearly PV installations (in per-capita terms) during the same years. We also control for average construction activity, the average yearly change in the working age population, and state $\times$ county-type fixed effects. As baseline, we estimate model (3.3) for wage growth between 2003 and 2012 for the full set of counties. For the cross-sectional split, we consider the same time horizon and classify counties as slack if their unemployment rate in 2002 exceeds the state average unemployment rate in 2002. Along the time series, we split the sample into the periods of 2003 to 2007 (slack) and 2008 to 2012 (tight).[38]

This model builds on Autor et al. (2013) and evaluates the effect of PV installations on wage growth in the medium run. We focus on the medium run to account for the common finding of sticky wages (Taylor, 2016). Sticky wages are particularly relevant in the context of PV installations as almost 90 percent of workers in the construction sector are subject to collective bargaining agreements on the industry and regional level, which on average have a duration of two years.[39] Furthermore, bargaining outcomes may not immediately adjust to demand shocks.[40]

Table 3.6 shows the results of estimating equation (3.3) via OLS. For the full set of counties, there is a statistically significant response of wage growth to additional PV installations (column (1)). The coefficient of 9.21 implies that a one standard deviation increase in capacity leads to a yearly increase in the median construction wage of 0.05 percentage points. The splits in tight and slack labor markets reveal that this additional wage growth is driven almost exclusively by tight markets, where the coefficients are around 10 as well. In contrast, the coefficients for slack markets are small and imprecisely estimated. For this reason, we cannot reject the null hypothesis that the effect of PV installations on wage growth is larger in slack than in tight markets on conventional levels. Overall, these results mirror the findings of the employment gains and point towards crowding out as an explanation.[41]

[38]The cross-sectional split corresponds to row (10) of Table 3.3 and the time series split to row (2). Both splits ensure that the classification of slack and tight markets is constant for the considered horizons.

[39]For durations of collective bargaining agreements, see `https://www.boeckler.de/wsi-tarifarchiv\_4832.htm`, last accessed 25 November 2018.

[40]The survey evidence by Smets and Lamo (2009) shows that the wage elasticity to idiosyncratic shocks is low and that firms are more likely to adjust employment than wages in response to demand shocks. For PV installations, this may be particularly relevant as it was not clear ex ante how permanent these shocks would be.

[41]Table C.13 in Appendix C.4.4 shows the corresponding IV results which have the same qualitative pattern.

Table 3.6: Wage Growth

	Δ Log(Median Wage in Construction)				
		Split along			
		Time series		Cross-section	
	Baseline	Slack	Tight	Slack	Tight
Years	03-12	03-07	08-12	03-12	03-12
	(1)	(2)	(3)	(4)	(5)
Avg. yearly capacity p.c.	9.21^{**}	3.02	11.17^{***}	6.88	10.09^{**}
	(3.53)	(18.44)	(3.15)	(10.27)	(4.05)
Population growth	−0.00	−0.10	0.16^{**}	−0.02	0.02
	(0.05)	(0.08)	(0.07)	(0.12)	(0.08)
Avg. yearly construction p.c.	0.11	0.22	-0.49^{**}	0.42	−0.04
	(0.21)	(0.23)	(0.25)	(0.42)	(0.28)
State × county-type FE	yes	yes	yes	yes	yes
P-value slack > tight		0.206		0.190	
Observations	368	363	370	159	209

Note: The dependent variable is the difference in the log median wage in the construction sector over the years indicated in the row *"Years"* (referred to as "sample years"). *Avg. yearly capacity p.c.* are the average yearly PV installations during the sample years measured in megawatt peak (MWp) normalized by the working-age population in 2003 (indicated by "p.c." for "per capita"). *Population Growth* is the difference in the working-age population over the sample years relative to the working-age population in 2003. *Avg. yearly construction p.c.* is the average yearly number of residential and non-residential buildings completed during the sample years. For the time-series defintion of slack in columns (2) and (3) we split the sample into the years 2003 to 2007 (slack) and 2008 to 2012 (tight). In the cross-section in columns (4) and (5), we split the sample relative to the mean of the average unemployment rate at the state-level in 2002. *P-Val slack > tight* reports the p-value of the test of the null hypothesis that the wage effect of PV installations is larger in slack than in tight labor markets. The number of observations is smaller than the number of counties (400) because the wage data is only available for county-years in which the number of employees in construction exceeds 1000. Standard errors (in parentheses) are clustered at the level of 94 spatial planning regions. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

3.6 Conclusion

This chapter shows that the employment effects of PV investment in Germany are state-dependent. It overcomes all three challenges to the estimation of state dependent multipliers: First, we study an exogenous demand shock, as PV investment was mainly driven by factors independent of regional labor market dynamics. Second, this demand shock provides ample variation in all states of the economy as more than €60 billion were invested across 400 counties over ten years. Third, the composition of investments has been constant as each PV installation of a given size constituted the same shock. This allows us to compare the estimated multipliers between the different states of the economy.

Our results are consistent with crowding-out being responsible for the low impact of PV installations on employment in tight labor markets. Thus, our chapter adds evidence for a particular mechanism underlying state-dependence. Tracing out the mechanism is a plausible way forward in studying state-dependent multipliers as it might help to reconcile the different results in the literature. This chapter takes a first step in this direction.

Finally, our chapter has two policy implications: First, investments during a recession pay a double dividend as they put additional people to work, while they mostly lead to crowding out in booms. Hence, economic downturns are a good time to undertake public investment programs. Second, place-based policies provide a better return in terms of jobs in regions with high unemployment than with low unemployment. This suggests that the design of make-work programs should take economic circumstances of targeted areas into account, even if we ignore all equity concerns.

Appendices

A

Appendix to Chapter 1

A.1 Appendix to Section 1.3: Data

The publication data is from Microsoft Academic.[1] 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.

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

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. Publications in top multidisciplinary journals (Nature, Science, Proceedings of the Academy of Sciences of the United States of America, and Nature Communications) and top three field journals (see Table A.2) are counted separately.

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, conventionality and potential high impact publications 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

6. Following Uzzi et al. (2013), papers are defined to have potential high impact if they score high on novelty and commonness. Specifically, a paper needs to have a novelty score above the 90^{th} percentile of the vintage specific distribution of novelty scores across papers. In addition, it also needs an above median conventionality score relative to the vintage specific distribution of conventionality scores across all papers.

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:

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

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 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.

The same data is collected for the early career prize winners in a similar fashion. The only differences are that we do not collect information on research grants and affiliation due to data availability. We collect all winners from the prizes listed in Table A.1 and check via a search of Microsoft Academic and recipients' CVs whether they continued in academia. Some of our prizes are prizes for Ph.D. theses and some winners leave academia immediately afterward. We exclude these recipients.

Table A.1: List of Early Career Prizes

Prize	Field	No. of Winners
Arnold Eucken Prize	Engineering	11
Emmy Noether Indep. Junior Research Group Leader	All fields	87
Eugen Hartmann Prize	Engineering	4
Heinz-Maier-Leibnitz Prize	All fields	79
Heisenberg Fellowship	All fields	227
Masing Prize	Engineering	9
Max-Planck Research Group	Most fields	46

Note: The Heinz Maier-Leibnitz Prize is awarded to Ph.D. theses in all disciplines by the DFG. The Emmy Noether and Heisenberg program are programs by the DFG for post Ph.D. independent researchers. Similarly, Max-Planck Research Groups are headed by young researchers after their Ph.D. and affiliated with an existing Max-Planck institute. Since the Max-Planck Society spans most, but not all fields (e.g. not engineering) and engineers are less likely to participate in programs such the Heisenberg Fellowship and the Emmy Noether program, we include several engineering specific early career prizes (See e.g. `http://www.dfg.de/dfg_magazin/wissenschaftliche_karriere/heisenberg/was_das_programm_auszeichnet/index.html`, accessed on 10 August 2018). The Arnold Eucken, Eugen Hartmann and Masing Prize are all prizes by different engineering associations for young researchers, e.g. the Masing Prize is awarded by the German Association for Materials Science and Engineering. All of these prizes are awarded early on in a researcher's career and we approximate when she "should have" received a Leibniz Prize based on her age. Hence, we assign a placebo Leibniz Prize year by adding the average years between the early career prize and age 45. The time lags are for each prize: Arnold Eucken (11 years), Emmy Noether (10), Eugen Hartmann (11), Heinz-Maier-Leibnitz (11), Heisenberg Fellowship (9), Masing Prize (11), Max-Planck (10).

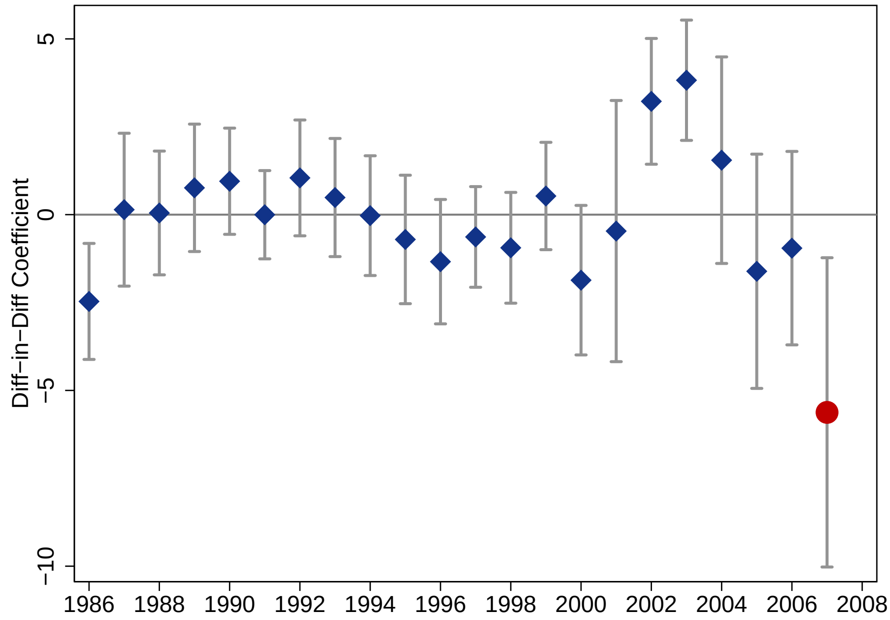
Table A.2: List of Top 3 Journals by Field

Journal	Discipline	Rank
Plant Cell	Agricultural and Biological Sciences	1
Trends in Ecology and Evolution	Agricultural and Biological Sciences	2
Current Biology	Agricultural and Biological Sciences	3
Science	Arts and Humanities	1
Psychological Bulletin	Arts and Humanities	2
Annals of the New York Academy of Sciences	Arts and Humanities	3
Cell	Biochemistry, Genetics and Molecular Biology	1
Journal of the American Chemical Society	Biochemistry, Genetics and Molecular Biology	2
Nature Genetics	Biochemistry, Genetics and Molecular Biology	3
Academy of Management Journal	Business, Management and Accounting	1
Journal of Finance	Business, Management and Accounting	2
Strategic Management Journal	Business, Management and Accounting	3
Journal of the American Chemical Society	Chemical Engineering	1
Angewandte Chemie - International Edition	Chemical Engineering	2
Nano Letters	Chemical Engineering	3
Chemical Reviews	Chemistry	1
Journal of the American Chemical Society	Chemistry	2
Angewandte Chemie - International Edition	Chemistry	3
Bioinformatics	Computer Science	1
IEEE Transactions on Pattern Analysis and Machine Intelligence	Computer Science	2
Lecture Notes in Computer Science	Computer Science	3
European Journal of Operational Research	Decision Sciences	1
Management Science	Decision Sciences	2
Research Policy	Decision Sciences	3
Astrophysical Journal	Earth and Planetary Sciences	1
Journal of Geophysical Research	Earth and Planetary Sciences	2
Monthly Notices of the Royal Astronomical Society	Earth and Planetary Sciences	3
Journal of Finance	Economics, Econometrics and Finance	1
American Economic Review	Economics, Econometrics and Finance	2
Quarterly Journal of Economics	Economics, Econometrics and Finance	3
Journal of Power Sources	Energy	1
Journal of the Electrochemical Society	Energy	2
Bioresource Technology	Energy	3
Advanced Materials	Engineering	1
Nano Letters	Engineering	2
Nature Materials	Engineering	3
Environmental Science & Technology	Environmental Science	1
Applied and Environmental Microbiology	Environmental Science	2
Journal of Geophysical Research	Environmental Science	3
Blood	Immunology and Microbiology	1
Journal of Experimental Medicine	Immunology and Microbiology	2
Nature Biotechnology	Immunology and Microbiology	3
Advanced Materials	Materials Research	1
Nano Letters	Materials Research	2
Nature Materials	Materials Research	3
Bioinformatics	Mathematics	1
IEEE Transactions on Pattern Analysis and Machine Intelligence	Mathematics	2
Lecture Notes in Computer Science	Mathematics	3
New England Journal of Medicine	Medicine	1
The Lancet	Medicine	2
JAMA - Journal of the American Medical Association	Medicine	3
Nature	Multidisciplinary	1
Science	Multidisciplinary	2
Proceedings of the National Academy of Sciences of the United States of America	Multidisciplinary	3

Neuron	Neuroscience	1
Journal of Neuroscience	Neuroscience	2
Nature Neuroscience	Neuroscience	3
Diabetes Care	Nursing	1
American Journal of Clinical Nutrition	Nursing	2
Stroke	Nursing	3
Physical Review Letters	Physics and Astronomy	1
Nano Letters	Physics and Astronomy	2
Applied Physics Letters	Physics and Astronomy	3
Journal of Personality and Social Psychology	Psychology	1
Psychological Bulletin	Psychology	2
Trends in Cognitive Sciences	Psychology	3
Journal of Personality and Social Psychology	Social Sciences	1
IEEE Transactions on Information Theory	Social Sciences	2
Child Development	Social Sciences	3

Note: The ranking is taken from `https://www.scimagojr.com/journalrank.php` and we use the ranking for 2006 based on the Scimago Journal Ranking. According to Scimago, it expresses the average number of weighted citations received in 2006 by the documents in the journal published in the three preceding years. We exclude some fields where no Leibniz Prize recipients are active in, namely dentistry, health professions, and veterinary medicine. We then count the publications in all of these journals for each individual Leibniz Prize recipient. This also circumvents the issue of mapping Leibniz Prize recipients into narrow fields.

Figure A.1: Placebo Treatment Exercise: Number of Publications (all types)



Note: This figure shows the diff-in-diff coefficient of equation (1.1). For each coefficient, the treatment indicator has been reassigned to the (prize) year on the x-axis and the following three prize cohorts. For example, for the case of the 1986 coefficient, the "treated" prize recipients are those who received their prize between 1986 and 1989. All other prize cohorts are then considered untreated. The coefficient depicted by the red circle is the baseline coefficient for the actual treatment group of prize recipients from 2007 to 2010. The dependent variable is the number of publications (all types). 95 percent confidence intervals are based on standard errors clustered on the year of prize reception. I use the weights of Iacus et al. (2012) to arrive at the average treatment effect on the treated.

A.2 Appendix to Section 1.4: Further Robustness

In this section, we provide additional evidence for the plausibility of the identifying assumption using a falsification exercise. Furthermore, our results are not driven by the choice of the length of the pre-period, 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 individual treatment effects for all prize recipients.

Falsification Exercise

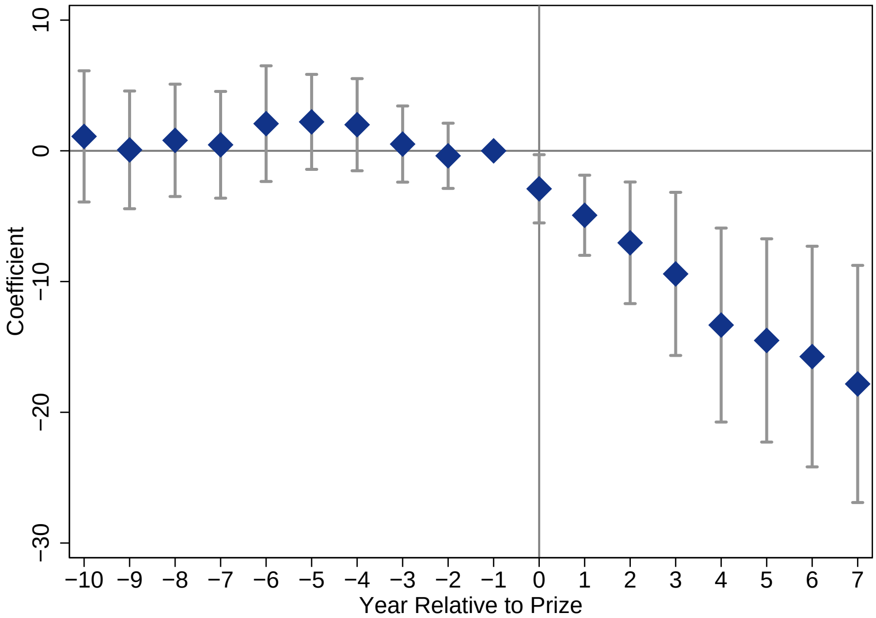
Figure A.1 shows the results from a simple falsification exercise. We re-assign the treatment indicator to each four-year span between 1986 and 2010. Specifically, we first assign treatment to the Leibniz Prize winners from 1986 to 1989 and re-estimate equation (1.1). Next, we assign treatment to the winners between 1987 and 1990 and use all other winners as controls, etc. The diff-in-diff coefficient of interest for each of these placebo treatments is shown in Figure A.1. The last coefficient in red is the actual baseline treatment coefficient. Except when assigning the treatment indicator to the cohorts 2002 to 2005 or 2003 to 2006, all other coefficients are insignificant and centered around zero. Given that we estimate 21 placebo regressions, one significant coefficient would be expected by chance alone. Since we define overlapping placebo treatments, it is not surprising that there are two significant placebo coefficients. Reassuringly, the baseline coefficient is larger in absolute magnitude than all other coefficients.

Additional Robustness

We first assess robustness to extending the pre-period to ten years prior to prize reception. Figure A.2 depicts time-varying coefficients for the overall number of publications. Reassuringly, prior to prize reception, all coefficients are clustered around zero and post prize become significantly negative. Table A.3 presents results for the same specification and dependent variables as in the main Table 1.2. The only difference is that since the pre-period is extended to ten years before prize reception, the measure of abstract similarity relative to the early stock of publications is not defined for this period. The results are quantitatively similar to the baseline results and tend to be larger in magnitude (especially for the top field publications). The publication counts of all types, top ranked multidisciplinary and top field journals are statistically significant. Hence, extending the pre-period does not alter our results.

All publication count variables are by definition count variables. Especially for

Figure A.2: Time-Varying Treatment Effect on the Number of Publications (Longer Pre-Period)



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. 95 percent confidence intervals are based on standard errors clustered on the year of prize reception. I use the weights of Iacus et al. (2012) to arrive at the average treatment effect on the treated.

Table A.3: Effect of the Leibniz Prize Reform on Scientific Productivity (Diff-in-Diff Estimates, Longer Pre-Period)

	(1) All Pubs	(2) Top Multidisc. Pubs	(3) Top Field Pubs	(4) Abstract Sim. I	(5) Novelty	(6) Convent.	(7) High Impact Pubs
Post Prize	1.11^{**}	-0.02	-0.09	0.01	0.12^{**}	-0.01	0.02
	(0.49)	(0.04)	(0.09)	(0.02)	(0.05)	(0.03)	(0.01)
Post Prize × Post 2007	-6.04^{**}	0.08^{***}	0.44^{**}	-0.07	-0.06	0.02	0.09
	(2.62)	(0.03)	(0.18)	(0.06)	(0.09)	(0.10)	(0.07)
Mean	9.22	0.14	0.48	0.25	1.81	-0.64	0.04
Winners	257	257	257	252	256	256	257
R^2	0.21	0.03	0.04	0.14	0.06	0.06	0.02
Observations	4626	4626	4626	3536	3973	3973	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 (1.1). The dependent variable in column (1) is the count of all types of publication per year. In column (2) only publications in top general interest journals (Science, Nature, PNAS, Nature Communications) are counted. Column (3) uses publications in field specific top 3 journals. In column (4), the dependent variable is the text similarity of abstracts to each other within a given year. In columns (5) and (6), novelty and conventionality are defined as in Lee et al. (2015). In column (7), the dependent variable is the count of publications with potential high impact, i.e. having high conventionality and novelty. In all regressions, we use the weights suggested by Iacus et al. (2012) to identify the average treatment effect on the treated. 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 A.4: Robustness I: Estimation Using Count Data Models

	Poisson				Negative Binomial			
	(1) All Pubs	(2) Top Multidisc. Pubs	(3) Top Field Pubs	(4) Impact Pubs	(5) All Pubs	(6) Top Multidisc. Pubs	(7) Top Field Pubs	(8) Impact Pubs
Post Prize × Post 2007	-0.28^{***}	0.43	0.42^{*}	0.44	-0.15^{**}	0.43	0.44^{**}	0.42
	(0.10)	(0.31)	(0.24)	(0.80)	(0.06)	(0.33)	(0.18)	(0.54)
Rel. to mean	−0.24	0.53	0.53	0.56	−0.14	0.53	0.55	0.53
Winners	257	114	163	98	257	114	163	98
Observations	3341	1482	2119	1274	3341	1482	2119	1274

Note: This table shows the results from a difference-in-differences estimation with five years before receiving the Leibniz Prize as pre-period and seven years after as post-period. The estimation equation is as in equation (1.1). The dependent variable in columns (1) and (5) is the count of all types of publication per year. In columns (2) and (6) only publications in top general interest journals (Science, Nature, PNAS, Nature Communications) are counted. Columns (3) and (7) uses publications in field specific top 3 journals. Lastly, in columns (4) and (8), the dependent variable is the count of publications with potential high impact, i.e. having high conventionality and novelty. The first four columns are estimated using a fixed effects poisson model, the latter four columns are estimated using a fixed effects 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. Robust standard errors in parentheses. *, ** and *** denote significance on the 10 percent, 5 percent and 1 percent level, respectively.

the top multidisciplinary 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 Table A.4. In both cases, the coefficient for the number of all publications and top field publications is statistically significant. In terms of magnitude, the poisson regression estimate implies a reduction of 24 percent relative to the mean number of publications (all types) and the negative binomial regression a reduction of 14 percent. This is smaller than the OLS estimate of 53 percent, but qualitatively similar. The increase in the number of publications in top ranked journals is of similar magnitude as the OLS results, underlining that the results do not depend on the estimation method.

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.

Normalizing publication counts with the square root of the number of authors in columns (1) to (4) of Table A.5 yields results that are qualitatively and quantitatively similar to our baseline results. In column (1), the coefficient implies a reduction of 2.49 publications per author, or 43 percent relative to the mean. For top multidisciplinary journals there is a significant increase of 44 percent relative to the mean. In contrast to the raw counts, the coefficient for top field publications is not significant when normalized with the number of authors (p-value =

Table A.5: Robustness II: Weighting the Dependent Variables

	(1) All Pubs (fract)	(2) Top Multidis. Pubs (fract)	(3) Top Field Pubs (fract)	(4) High Imp. Pubs (fract)	(5) All Pubs (cit)	(6) Top Multidis. Pubs (cit)	(7) Top Field Pubs (cit)	(8) High Impact Pubs (cit)
Post Prize	0.79**	0.00	−0.03	0.01	−3.61	−7.13	−16.90**	0.13
	(0.32)	(0.02)	(0.04)	(0.01)	(8.22)	(4.59)	(6.49)	(0.09)
Post Prize ×	−2.49***	0.04**	0.08	0.07*	−35.64	3.03	22.27***	1.81
Post 2007	(0.84)	(0.02)	(0.06)	(0.04)	(34.95)	(4.72)	(7.37)	(1.06)
Mean	5.84	0.09	0.28	0.03	105.94	8.59	22.39	0.43
Winners	257	257	257	257	257	257	257	257
R^2	0.12	0.03	0.03	0.02	0.08	0.02	0.02	0.02
Observations	3341	3341	3341	3341	3285	3285	3285	3285

Note: This table shows the results from a difference-in-differences estimation with five years before receiving the Leibniz Prize as pre-period and seven years after as post-period. The estimation equation is as in equation (1.1). In columns (1) to (4), all counts are normalized with the square root of the number of authors per publication. In columns (5) to (8), all publication counts are weighted with the number of citations received in the three years following publication. The number of observations drops as the most recent prize cohorts (2008 to 2010) have not had enough time to accrue three year forward citations for the publications in the last years under study. The dependent variable in columns (1) and (5) is the count of all types of publications per year. In columns (2) and (6) only publications in top general interest journals (Science, Nature, PNAS, Nature Communications) are counted. Columns (3) and (7) use publications in field specific top 3 journals. In columns (4) and (8), the dependent variable is the count of publications with potential high impact, i.e. having high conventionality and novelty. In all regressions, we use the weights suggested by Iacus et al. (2012) to identify the average treatment effect on the treated. 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.

0.23). In turn, though, the coefficient on potential high impact publications is statistically significant on the 10 percent level.

Columns (5) to (8) focus on the same publication counts, but these are now weighted with the number of citations received in the three years following publication. There is no significant decrease in the overall number of citation-weighted publications and positive point estimates for publications in top multidisciplinary journals (not significant), top field journals (significant on 1 percent level) and potential high impact publications (almost significant, p-value = 0.10). The fact that there is no significant decrease in the overall number of publications when we weight these with their follow-on citations is in line with the proposed mechanism of Leibniz Prize recipients focusing on fewer high quality publications in response to the reform of 2007. 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 1.2 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 re-estimated in a leave-one-out fashion. The weights suggested by Iacus et al. (2012) are re-calculated in every iteration. Figures A.3 and A.4 show the results of this exercise. Each of the eight panels shows the estimated coefficient for the Post 2007 $\times$ Post Prize interaction and accompanying 90 percent confidence intervals for each of the eight dependent variables. As comparison, the red line in each figure shows the baseline coefficient for the full sample. Although there is some movement in the coefficients, by and large the confidence intervals overlap and the coefficients are close to each other and the baseline estimate, indicating that individual cohorts do not drive the results.

Table A.6: Robustness of Inference I: Scientific Productivity Effects in Diff-in-Diff Specification (Wild Cluster Bootstrap)

	(1) All Pubs	(2) Top Multidisc. Pubs	(3) Top Field Pubs	(4) Abstract Simil. I	(5) Abstract Simil. II	(6) Novelty	(7) Convent.	(8) High Impact Pubs
Panel A: Unrestricted Wild Cluster Bootstrap								
Post Prize	1.57**	−0.03	−0.08	0.02	−0.01*	0.15***	−0.08*	0.01
	(0.02)	(0.48)	(0.43)	(0.21)	(0.09)	(0.00)	(0.09)	(0.41)
Post Prize × Post 2007	−5.62**	0.08*	0.29	−0.06	−0.02***	−0.13	0.15	0.13
	(0.02)	(0.06)	(0.23)	(0.29)	(0.00)	(0.29)	(0.21)	(0.28)
Panel B: Restricted Wild Cluster Bootstrap								
Post Prize	1.57***	−0.03	−0.08	0.02**	−0.01	0.15***	−0.08**	0.01
	(0.00)	(0.30)	(0.17)	(0.03)	(0.66)	(0.00)	(0.02)	(1.00)
Post Prize × Post 2007	−5.62**	0.08*	0.29	−0.06	−0.02***	−0.13	0.15	0.13
	(0.02)	(0.06)	(0.28)	(0.24)	(0.00)	(0.27)	(0.17)	(0.18)
Mean Dep.	10.64	0.15	0.54	0.27	0.13	1.88	−0.69	0.05
Observations	3341	3341	3341	2719	2536	2964	2964	3341

Note: This table shows the results from a difference-in-differences estimation with five years before receiving the Leibniz Prize as pre-period and seven years after as post-period. The estimation equation is as in equation (1.1). The dependent variable in column (1) is the count of all types of publication per year. In column (2) only publications in top general interest journals (Science, Nature, PNAS, Nature Communications) are counted. Column (3) uses publications in field specific top 3 journals (See Table A.2 for the full list of journals). In column (4), the dependent variable is the text similarity of abstracts to each other within a given year. In column (5), 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 (6) and (7), novelty and conventionality are defined as in Lee et al. (2015). Lastly, in column (8), the dependent variable is the count of publications with potential high impact, i.e. having high conventionality and novelty. In all regressions, we use the weights suggested by Iacus et al. (2012) to identify the average treatment effect on the treated. P-values in parentheses are clustered on the level of the year of the prize and are calculated using the unrestricted (restricted) wild cluster bootstrap in Panel A (Panel B) as proposed by Cameron et al. (2008) with 1000 repetitions. *, ** and *** denote significance on the 10 percent, 5 percent and 1 percent level, respectively.

Table A.7: Robustness of Inference II: Triple Difference Comparison Relative to Early Career Prize Winners (Wild Cluster Bootstrap)

	(1) All Pubs	(2) Top Multi. Pubs	(3) Top Field Pubs	(4) Abstract Simil. I	(5) Abstract Simil. II	(6) Nov.	(7) Conv.	(8) High Imp. Pubs
Panel A: Unrestricted Wild Cluster Bootstrap								
Post Prize	-0.88^{**}	0.00	-0.01	-0.00	0.03^{*}	-0.06	0.05	-0.02
	(0.02)	(0.76)	(0.82)	(0.65)	(0.05)	(0.61)	(0.35)	(0.43)
Post Prize × Post 2007	1.68^{***}	0.00	0.06	0.01	-0.04^{***}	0.08	-0.01	0.02
	(0.00)	(0.70)	(0.36)	(0.12)	(0.00)	(0.50)	(0.94)	(0.35)
Post Prize × Leibniz	3.84^{***}	-0.01	-0.08^{***}	0.03	-0.00	0.21^{*}	-0.09	0.04
	(0.01)	(0.81)	(0.00)	(0.17)	(0.40)	(0.08)	(0.21)	(0.14)
Post Prize × Leibniz × Post 2007	-4.66^{**}	0.06	0.11	-0.05^{*}	-0.00	-0.23	0.10	0.02
	(0.02)	(0.14)	(0.49)	(0.10)	(0.80)	(0.31)	(0.41)	(0.75)
Constant	-1.12^{*}	-0.03^{***}	-0.06	-0.04^{**}	0.27^{***}	-1.52^{***}	1.98^{***}	-0.03^{**}
	(0.06)	(0.00)	(0.20)	(0.01)	(0.00)	(0.00)	(0.00)	(0.02)
Panel B: Restricted Wild Cluster Bootstrap								
Post Prize	-0.88^{***}	0.00	-0.01	-0.00	0.03^{**}	-0.06	0.05	-0.02
	(0.00)	(0.81)	(0.81)	(0.63)	(0.01)	(0.50)	(0.30)	(0.26)
Post Prize × Post 2007	1.68^{***}	0.00	0.06	0.01	-0.04^{***}	0.08	-0.01	0.02
	(0.00)	(0.71)	(0.33)	(0.14)	(0.00)	(0.45)	(0.94)	(0.19)
Post Prize × Leibniz	3.84^{***}	-0.01	-0.08^{***}	0.03	-0.00	0.21^{**}	-0.09	0.04^{*}
	(0.00)	(0.83)	(0.01)	(0.17)	(0.43)	(0.03)	(0.18)	(0.08)
Post Prize × Leibniz × Post 2007	-4.66^{***}	0.06^{*}	0.11	-0.05^{*}	-0.00	-0.23	0.10	0.02
	(0.00)	(0.08)	(0.36)	(0.08)	(0.76)	(0.20)	(0.36)	(0.68)
Constant	-1.12^{**}	-0.03^{***}	-0.06	-0.04^{**}	0.27^{***}	-1.52^{***}	1.98^{***}	-0.03^{**}
	(0.03)	(0.00)	(0.15)	(0.01)	(0.00)	(0.00)	(0.00)	(0.01)
Mean Dep.	8.50	0.06	0.23	0.22	0.13	2.20	-0.90	0.03
Observations	7345	7345	7345	5620	5794	6222	6222	7345

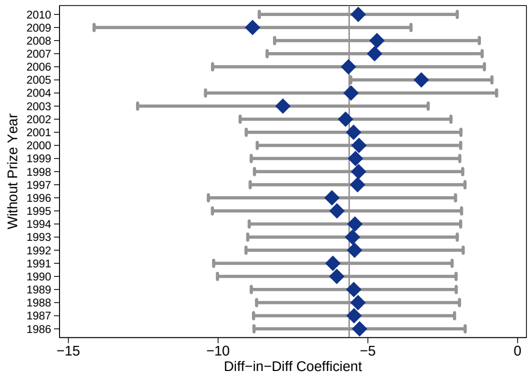
Note: This table shows the results from a difference-in-difference-in-differences estimation with five years before receiving the Leibniz Prize as pre-period and seven years after as post-period. Additional differences are between Leibniz Prize recipients and early career prize winners and between prize cohorts before and after 2007. The year of prize reception for Early Career Prize Winners is assigned by the average difference between receiving an early career prize and receiving the Leibniz Prize. 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 given in equation (1.2). The dependent variable in column (1) is the count of all types of publication per year. In column (2) only publications in top general interest journals (Science, Nature, PNAS, Nature Communications) are counted. Column (3) uses publications in field specific top 3 journals (See Table A.2 for the full list of journals). In column (4), the dependent variable is the text similarity of abstracts to each other within a given year. In column (5), 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 (6) and (7), novelty and conventionality are defined as in Lee et al. (2015). Lastly, in column (8), the dependent variable is the count of publications with potential high impact, i.e. having high conventionality and novelty. In all regressions, we use the weights suggested by Iacus et al. (2012) to identify the average treatment effect on the treated. P-values in parentheses are clustered on the level of the year of the prize and are calculated using the unrestricted (restricted) wild cluster bootstrap in Panel A (Panel B) as proposed by Cameron et al. (2008) with 1000 repetitions. *, ** and *** denote significance on the 10 percent, 5 percent and 1 percent level, respectively.

Table A.8: Robustness of Inference III: Funding Amount and Funding Duration (Wild Cluster Bootstrap)

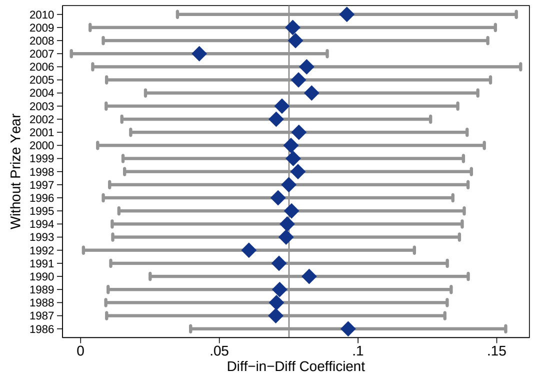
	(1) All Pubs	(2) Top Multi. Pubs	(3) Top Field Pubs	(4) Abstract Simil. I	(5) Abstract Simil. II	(6) Nov.	(7) Conv.	(8) High Imp. Pubs
Panel A: 2000 to 2006 Prize Cohorts vs. 1986 to 1992 Prize Cohorts – Varying Funding Amount								
Unrestricted Wild Cluster Bootstrap								
Post Prize	0.57	0.01	−0.14	−0.03	−0.00	0.06	0.03	−0.01
	(0.61)	(0.85)	(0.47)	(0.38)	(0.56)	(0.49)	(0.50)	(0.79)
Post Prize ×	−0.40	−0.05	0.13	0.06*	0.00	0.01	−0.10	−0.01
Pre 1992	(0.78)	(0.48)	(0.59)	(0.06)	(0.73)	(0.89)	(0.29)	(0.85)
Restricted Wild Cluster Bootstrap								
Post Prize	0.57	0.01	−0.14	−0.03*	−0.00	0.06	0.03*	−0.01
	(0.45)	(0.20)	(0.26)	(0.07)	(0.45)	(0.53)	(0.09)	(0.97)
Post Prize ×	−0.40	−0.05	0.13	0.06*	0.00	0.01	−0.10	−0.01
Pre 1992	(0.76)	(0.44)	(0.61)	(0.08)	(0.74)	(0.90)	(0.29)	(0.84)
Mean Dep.	10.77	0.16	0.55	0.27	0.13	1.81	−0.65	0.06
Observations	1911	1911	1911	1554	1349	1684	1684	1911
Panel B: 1986 to 1992 Prize Cohorts vs. 2007 to 2010 Prize Cohorts – Varying Funding Duration								
Unrestricted Wild Cluster Bootstrap								
Post Prize	0.38	−0.05	0.04	0.02	−0.01	0.06	−0.10	−0.03
	(1.00)	(1.00)	(1.00)	(0.42)	(0.16)	(0.46)	(0.43)	(1.00)
Post Prize ×	0.14	0.08	−0.10	−0.01	−0.00	−0.01	0.14	0.16
Post 2007	(1.00)	(1.00)	(1.00)	(0.84)	(0.95)	(0.98)	(0.40)	(1.00)
Restricted Wild Cluster Bootstrap								
Post Prize	0.38	−0.05	0.04	0.02	−0.01	0.06	−0.10	−0.03
	(1.00)	(1.00)	(1.00)	(0.40)	(0.17)	(0.44)	(0.21)	(1.00)
Post Prize ×	0.14	0.08	−0.10	−0.01	−0.00	−0.01	0.14	0.16
Post 2007	(1.00)	(1.00)	(1.00)	(0.83)	(0.95)	(0.98)	(0.46)	(1.00)
Mean Dep.	10.92	0.18	0.58	0.28	0.14	1.83	−0.70	0.05
Observations	1482	1482	1482	1165	946	1268	1268	1482
Panel C: 2000 to 2006 Prize Cohorts vs. 2007 to 2010 Prize Cohorts – Varying Funding Amount and Duration								
Unrestricted Wild Cluster Bootstrap								
Post Prize	2.64***	0.01	−0.22	0.01	0.01	0.14	−0.02	0.02
	(0.00)	(0.67)	(0.30)	(0.55)	(0.62)	(0.10)	(0.59)	(0.65)
Post Prize ×	−4.83***	0.05	0.38*	−0.04	−0.03***	−0.11	0.10	0.12
Post 2007	(0.00)	(0.27)	(0.08)	(0.41)	(0.00)	(0.38)	(0.35)	(0.31)
Restricted Wild Cluster Bootstrap								
Post Prize	2.64***	0.01	−0.22*	0.01	0.01	0.14**	−0.02	0.02
	(0.00)	(1.00)	(0.06)	(0.15)	(0.62)	(0.04)	(0.16)	(1.00)
Post Prize ×	−4.83**	0.05	0.38	−0.04	−0.03***	−0.11	0.10	0.12
Post 2007	(0.03)	(0.36)	(0.13)	(0.41)	(0.00)	(0.37)	(0.37)	(0.23)
Mean Dep.	12.86	0.18	0.64	0.30		2.09	−0.77	0.08
Observations	1339	1339	1339	1223	1227	1286	1286	1339

Note: This table shows the results from a difference-in-differences estimation with five years before receiving the Leibniz Prize as pre-period and seven years after as post-period. The estimation equation is as in equation (1.1). The dependent variable in column (1) is the count of all types of publication per year. In column (2) only publications in top general interest journals (Science, Nature, PNAS, Nature Communications) are counted. Column (3) uses publications in field specific top 3 journals (See Table A.2 for the full list of journals). In column (4), the dependent variable is the text similarity of abstracts to each other within a given year. In column (5), 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 (6) and (7), novelty and conventionality are defined as in Lee et al. (2015). Lastly, in column (8), the dependent variable is the count of publications with potential high impact, i.e. having high conventionality and novelty. 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. P-values in parentheses are clustered on the level of the year of the prize and are calculated using the unrestricted (restricted) wild cluster bootstrap as proposed by Cameron et al. (2008) with 1000 repetitions. *, ** and *** denote significance on the 10 percent, 5 percent and 1 percent level, respectively.

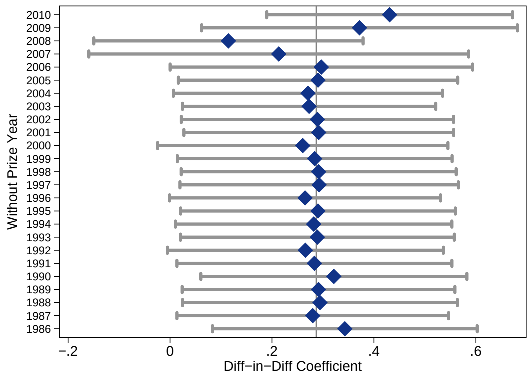
Figure A.3: Leave-One-Out: Dropping Prize Cohorts I



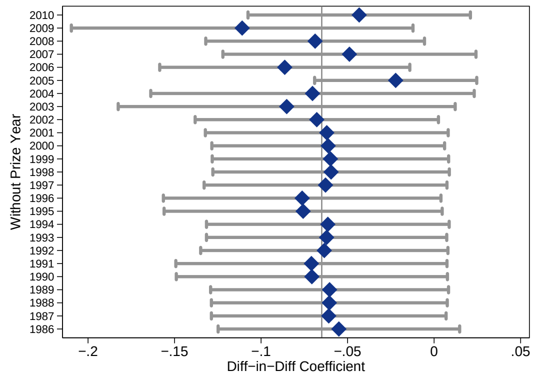
(a) No. of Publications (all types)



(b) No. of Hit Publications



(c) No. of Top 3 Publications



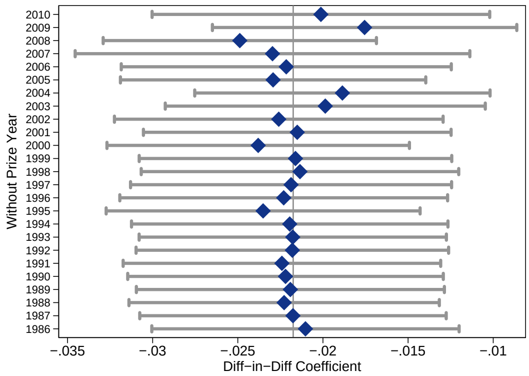
(d) Text Similarity I

Note: This figure shows the average treatment effect on the treated of receiving the Leibniz Prize in 2007 or later, dropping each prize cohort one by one. In each iteration, we re-calculate the weights of Iacus et al. (2012) to arrive at the average treatment effect on the treated. The dependent variable in Panel (a) is the count of all types of publication per year. In Panel (b) only publications in top general interest journals (Science, Nature, PNAS, Nature Communications) are counted. Panel (c) uses publications in field specific top 3 journals. In Panel (d), the dependent variable is the text similarity of abstracts to each other within a given year. 90 percent confidence intervals are based on standard errors clustered on the year of prize reception.

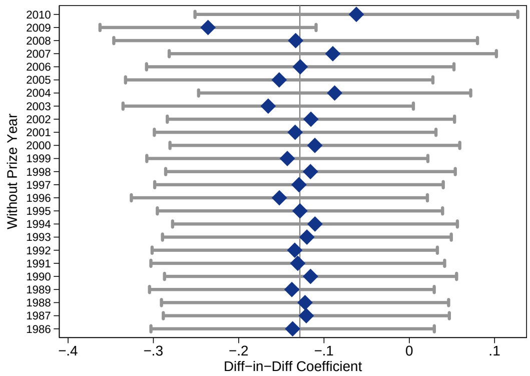
Adjusting for Small Number of Clusters

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 triple diff analysis and 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

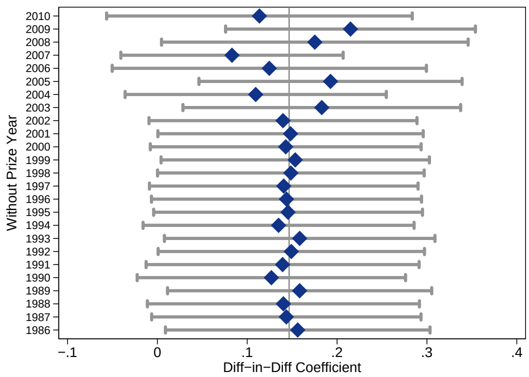
Figure A.4: Leave-One-Out: Dropping Prize Cohorts II



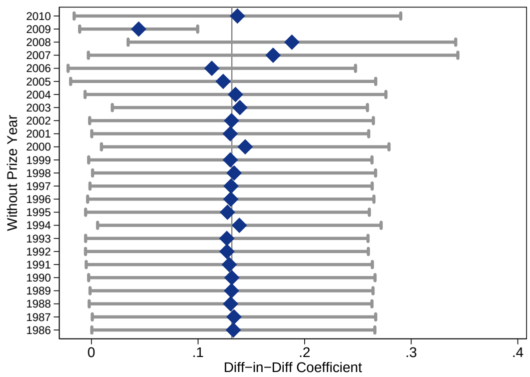
(a) Text Similarity II



(b) Novelty



(c) Conventionality



(d) No. of High Impact Publications

Note: This figure shows the average treatment effect on the treated of receiving the Leibniz Prize in 2007 or later, dropping each prize cohort one by one. In each iteration, we re-calculate the weights of Iacus et al. (2012) to arrive at the average treatment effect on the treated. The dependent variable in Panel (a) is the text similarity of abstracts relative to the early stock of publications of an author. Panels (b) and (c) use novelty and conventionality as defined by Lee et al. (2015) as a dependent variable, respectively. Lastly, in Panel (d), the dependent variable is the count of publications with potential high impact, i.e. having high conventionality and novelty. 90 percent confidence intervals are based on standard errors clustered on the year of prize reception.

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 Tables A.6 to A.8. Below each coefficient we show the associated p-value in parentheses. In Table A.6 we repeat the baseline analysis of Table 1.2. Only the coefficient on the publications in top field journals loses significance (p-value of 0.23 to 0.28), compared to the baseline results. In addition, the results are close to each other 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.[2] Similarly, in Tables A.7 and A.8, inference in the triple-differences exercise and the differentiation by the increase in real resources is also unchanged. The only differences are the following: In the triple-differences exercise, for publications in top multidisciplinary journals, the coefficient turns significant on the 10 percent level using the restricted wild cluster bootstrap. For publications in top field journals, the coefficient loses significance (p-value = 0.13) in the differentiation according to an increase in real funding amount and period when using the restricted wild cluster bootstrap. Overall, however, there is little evidence for a bias in our standard errors due to the comparably small number of clusters.

Additional Heterogeneity Analysis

To show the heterogeneity of the estimated treatment effect across individual prize recipients, we estimate individual treatment effects for each post 2007 Leibniz Prize recipient. This is done in two ways: first, a simple mean comparison is undertaken for each treated prize recipient relative to the control prize winners in the same field by university/research institute stratum. For example, in the smallest stratum (social sciences by research institute), Ulman Lindenberger would be compared to Wolfgang Klein and Wolfgang Prinz, before and after receiving the Leibniz Prize. Second, we interact the treatment dummy with each prize recipient. Figures A.5 to A.8 show the results of this exercise. The blue circles denote positive coefficients and the red diamonds negative ones. For brevity,

[2] In addition, in unreported regressions we use randomization based inference as proposed by MacKinnon and Webb (2018) and find that results tend to be more significant using this approach. The results are available from the author upon request.

we focus on the results for all publications and publications in top ranked journals, as these are significant in the main specification. 78 percent (86 percent) of the estimated treatment effects on the overall number of publications are negative if we do the simple mean comparison (the interaction analysis). Similarly, a majority of coefficients is positive for the number of publications in top ranked multidisciplinary and field specific journals (the proportion ranges from 56 percent to 78 percent). There is a sizable tail of individuals with very large treatment effects, but it is re-assuring that the majority of treatment effects are in line with the estimated baseline effect.

A.3 Appendix to Section 1.6: 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 for the treatment prize cohorts.

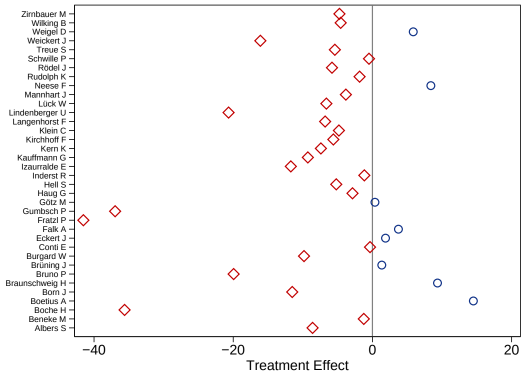
The DFG is the main source of research funding in Germany, spending around €3 billion in 2017.[3] We scrape information on traditional individual research grants (*Sachbeihilfen*) from the DFG's GEPRIS database. It contains information on all grants of the DFG since the early 2000s.[4] We focus on individual 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.[5] The data encompasses the title of the project, the applicant's

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

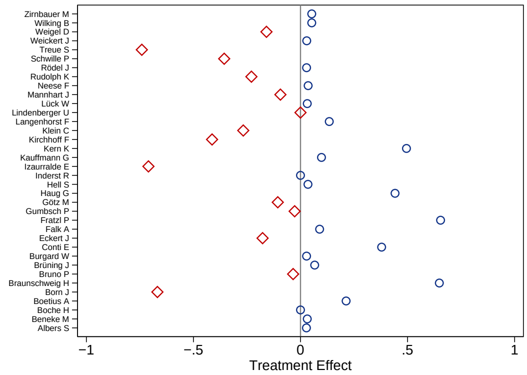
[4] The data can be found at `http://gepris.dfg.de/gepris/OCTOPUS`, last accessed on 07 August 2018.

[5] 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.

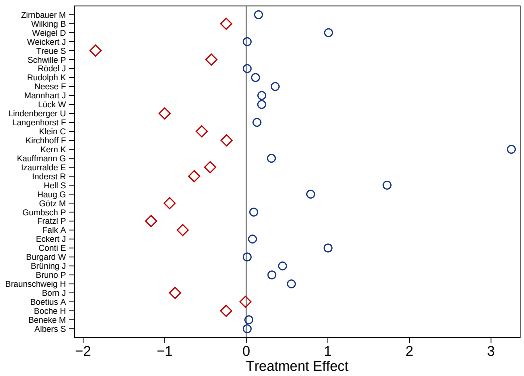
Figure A.5: Individual Treatment Effects – Means I



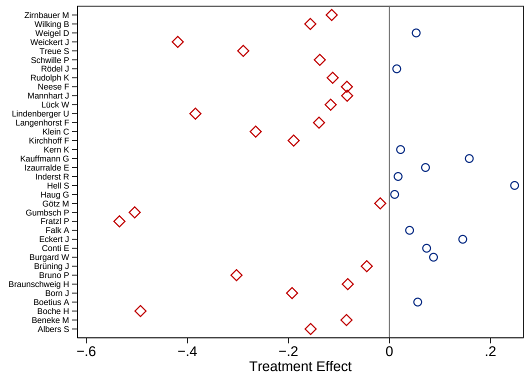
(a) No. of Publications (all types)



(b) No. of Hit Publications



(c) No. of Top 3 Publications



(d) Text Similarity I

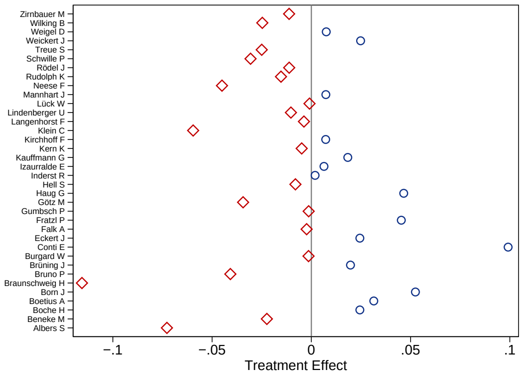
Note: This figure shows the estimated treatment effects for each post 2007 Leibniz Prize recipient. The treatment effect is estimated via a regression for each treatment recipient, using all control recipients in the same university/research institute by field stratum. The dependent variable in Panel (a) is the count of all types of publication per year. In Panel (b) only publications in top general interest journals (Science, Nature, PNAS, Nature Communications) are counted. Panel (c) uses publications in field specific top 3 journals. In Panel (d), the dependent variable is the text similarity of abstracts to each other within a given year. Hollow diamonds depict negative coefficients, hollow circles positive ones.

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.[6]

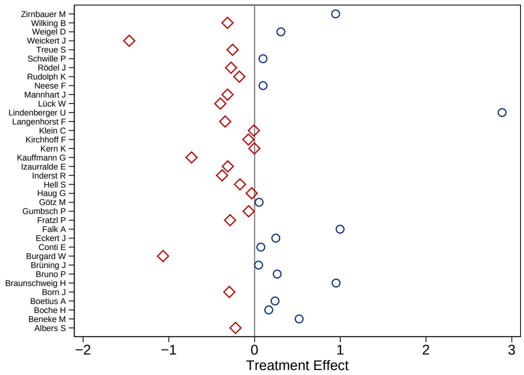
We scrape all available grants (around 60,000) and merge them to the Leibniz Prize recipients via their names. Due to the limited coverage of the data on

[6]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 in the same fashion.

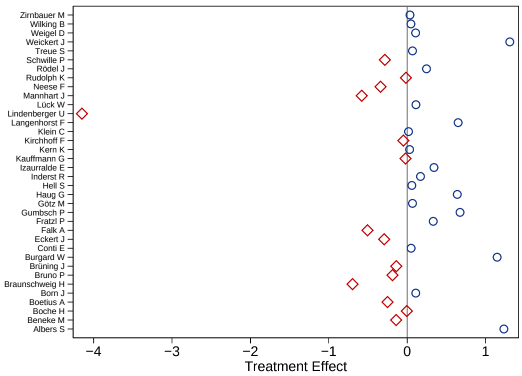
Figure A.6: Individual Treatment Effects – Means II



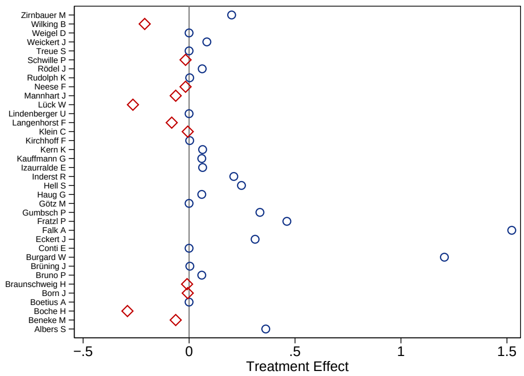
(a) Text Similarity II



(b) Novelty



(c) Conventionality

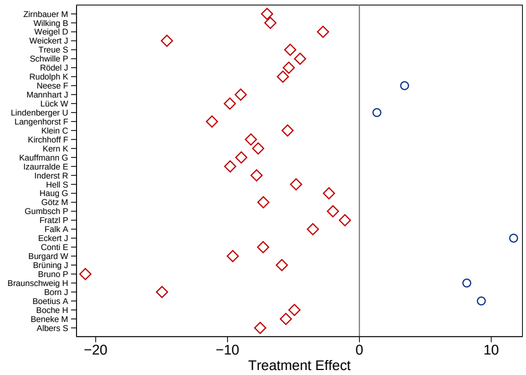


(d) No. of High Impact Publications

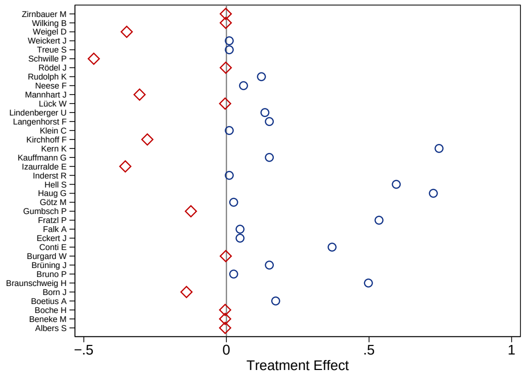
Note: This figure shows the estimated treatment effects for each post 2007 Leibniz Prize recipient. The treatment effect is estimated via a regression for each treatment recipient, using all control recipients in the same university/research institute by field stratum. The dependent variable in Panel (a) is the text similarity of abstracts relative to the early stock of publications of an author. Panels (b) and (c) use novelty and conventionality as defined by Lee et al. (2015) as a dependent variable, respectively. Lastly, in Panel (d), the dependent variable is the count of publications with potential high impact, i.e. having high conventionality and novelty. Hollow diamonds depict negative coefficients, hollow circles positive ones.

the post 2000 period, we can only study the prize cohorts from 2004 to 2010 to observe some years prior to prize reception. Figure A.11 shows the three year moving average of the number of 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

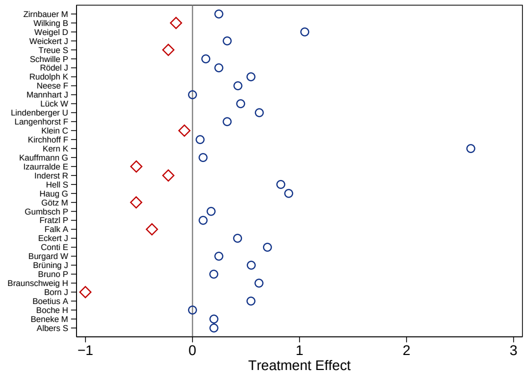
Figure A.7: Individual Treatment Effects – Interactions I



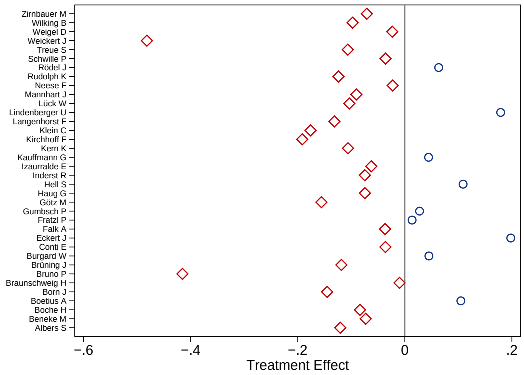
(a) No. of Publications (all types)



(b) No. of Hit Publications



(c) No. of Top 3 Publications



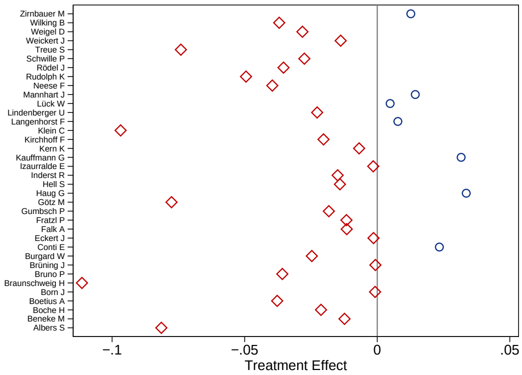
(d) Text Similarity I

Note: This figure shows the estimated treatment effects for each post 2007 Leibniz Prize recipient. The treatment effect is estimated by including an interaction for each post 2007 Leibniz Prize recipient. The dependent variable in Panel (a) is the count of all types of publication per year. In Panel (b) only publications in top general interest journals (Science, Nature, PNAS, Nature Communications) are counted. Panel (c) uses publications in field specific top 3 journals. In Panel (d), the dependent variable is the text similarity of abstracts to each other within a given year. Hollow diamonds depict negative coefficients, hollow circles positive ones.

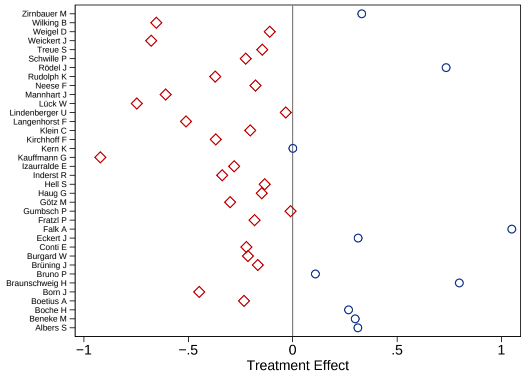
around the time when the Leibniz funding ends for this group. This suggests that the prize recipients do substitute regular grants with the Leibniz funding and that this effect is stronger for the group with a larger funding amount.[7]

[7] The (unreported) regression coefficient implies a reduction of 15% relative to the mean, but this is not statistically significant.

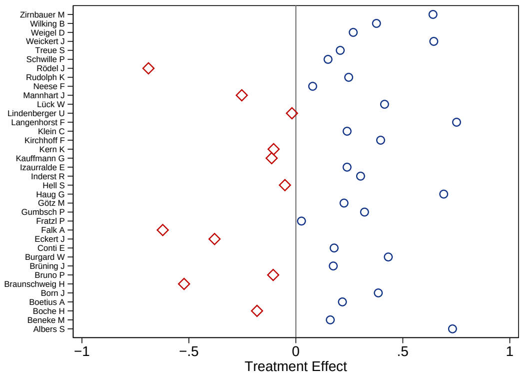
Figure A.8: Individual Treatment Effects – Interactions II



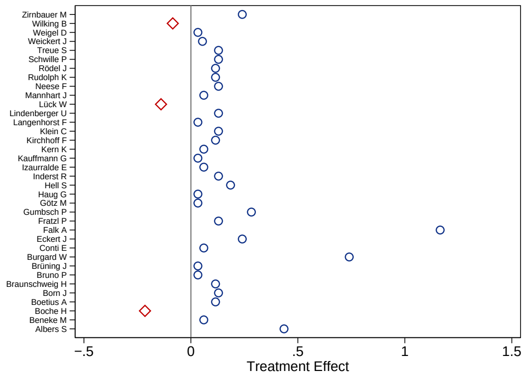
(a) Text Similarity II



(b) Novelty



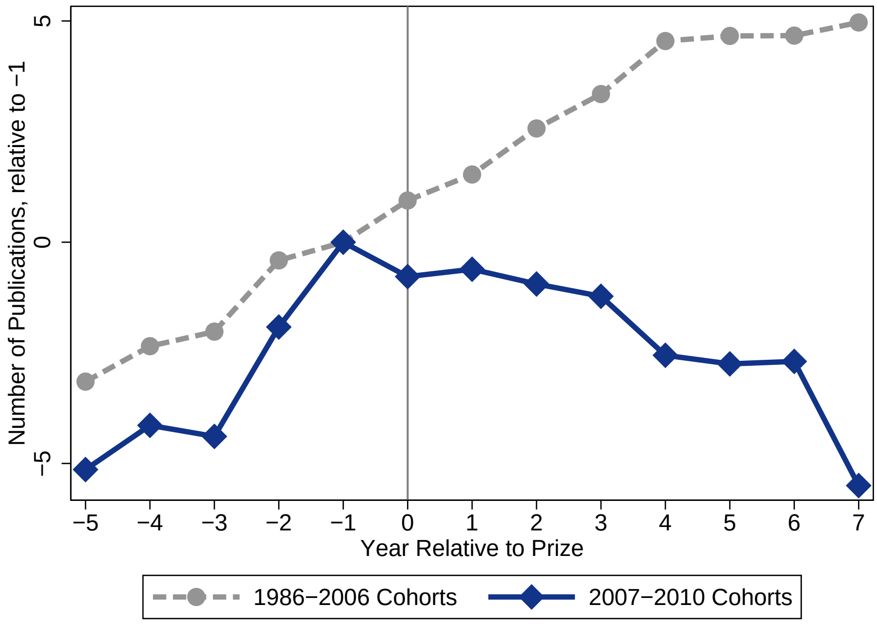
(c) Conventionality



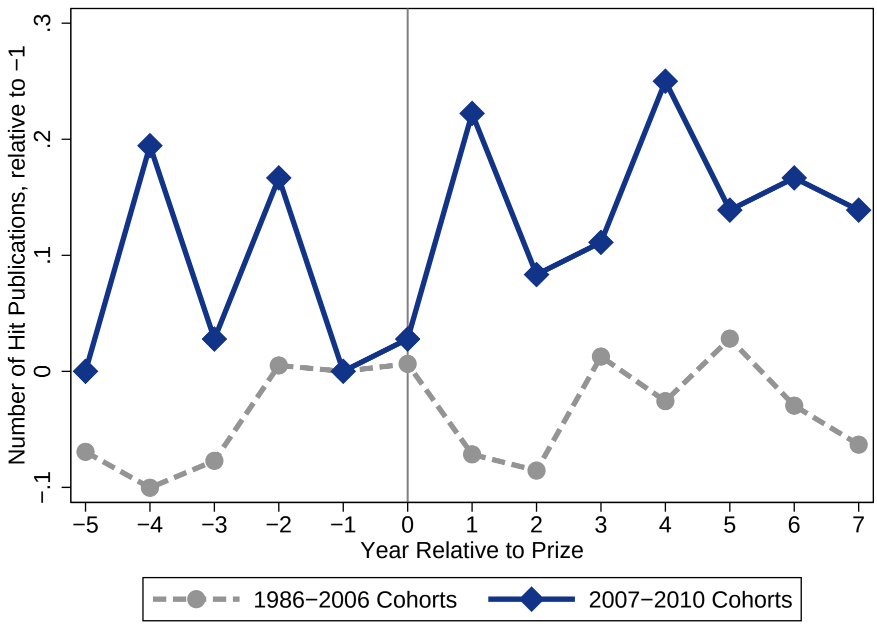
(d) No. of High Impact Publications

Note: This figure shows the estimated treatment effects for each post 2007 Leibniz Prize recipient. The treatment effect is estimated by including an interaction for each post 2007 Leibniz Prize recipient. The dependent variable in Panel (a) is the text similarity of abstracts relative to the early stock of publications of an author. Panels (b) and (c) use novelty and conventionality as defined by Lee et al. (2015) as a dependent variable, respectively. Lastly, in Panel (d), the dependent variable is the count of publications with potential high impact, i.e. having high conventionality and novelty. Hollow diamonds depict negative coefficients, hollow circles positive ones.

Figure A.9: Effect on Scientific Productivity (Non-Parametric Evidence, not averaged)



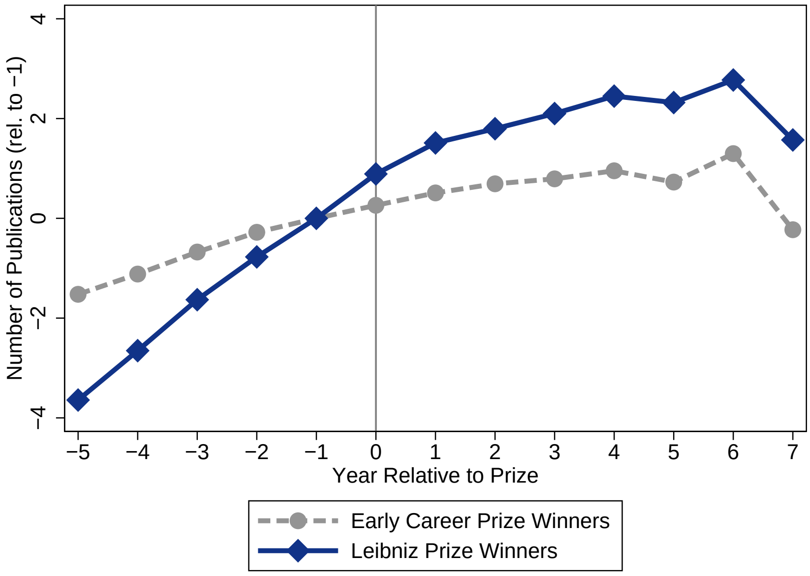
(a) No. of Publications (all types)



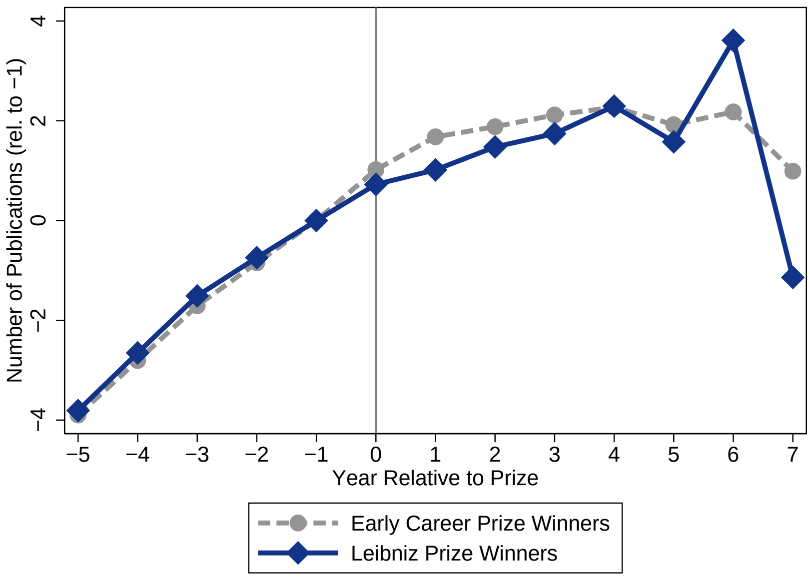
(b) No. of Hit Publications

Note: Mean number of publications (all types in Panel (a), hit publications in Panel (b)) per year, relative to year of Leibniz Prize reception. Means within groups are calculated using the weights of Iacus et al. (2012). Values are normalized with respect to relative year -1.

Figure A.10: Number of Publications (all types): Leibniz Prize Winners vs. Early Career Prize Winners



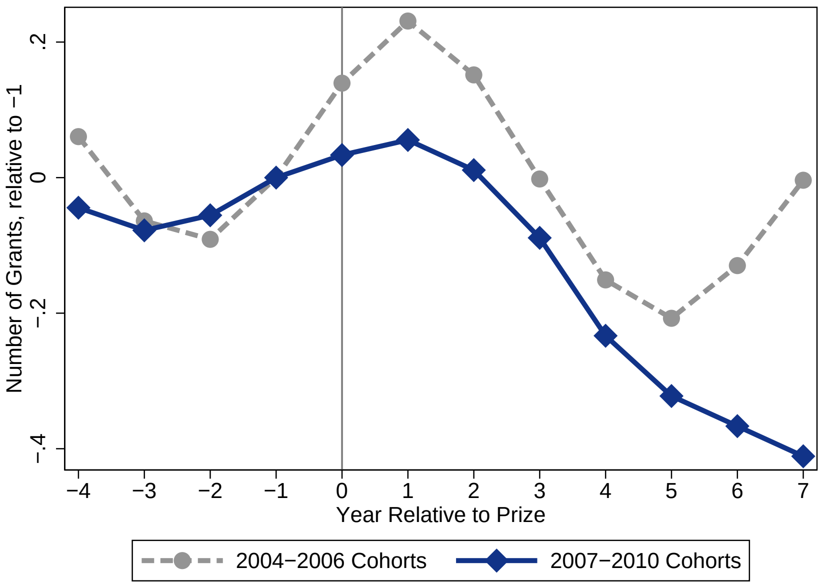
(a) Before Matching



(b) After Matching

Note: Five year moving average number of publications (all types) per year, relative to year of (assigned) Leibniz Prize reception. Panel (a) shows the means without prior coarsened exact matching on scientific field and number of publications prior to Prize reception. Panel (b) shows the means after matching. Only prize winners from 2000 to 2010 are used due to data availability. Prior to matching, 565 (early career and Leibniz Prize) winners are used, after matching 224 are retained. Means within groups are calculated using the weights of Iacus et al. (2012). Values are normalized with respect to relative year -1. Data for relative years 6 and 7 is not averaged.

Figure A.11: Effect of the Leibniz Prize Reform on the Number of Other Grants (Non-Parametric Evidence)



Note: Three-year moving average of the number of grants per researcher for the Leibniz Prize winners from 2004 to 2006, relative to the year before prize reception. Only traditional individual research grants (*Sachbeihilfen*) are counted. The data on grants is taken from the DFG's GEPRIS online database. I use the weights of Iacus et al. (2012) to arrive at the average treatment effect on the treated. Relative years -4 and 7 are not averaged.

B

Appendix to Chapter 2

B.1 Appendix to Section 2.3: Defining Product Entry and Exit

Product entry is defined as the producing of a product code that a company did not produce in the previous year. Analogously, product exit is when a company does not produce a product anymore that it produced in the previous year. Our production data on the 9-digit level is originally on the plant level, which we aggregate to the firm level using the firm identifiers in our data. The German product classification is based on the European Prodcom classification. Although the German classification does not change each year as the Prodcom data does, there is one large change in the classification during our period of study. In 2009, the new classification system GP 2009 replaced the old system GP 2002. This is an issue for our measurement of product entry and exit, as they would be grossly inflated if we did not adjust for the change in classification.

The change in the classification comes in two forms. Simple changes are when a single code from the GP 2002 matches into a single code from the GP 2009. Complex changes are when one or more codes from the old classification are mapped into one or more codes from the new classification. One approach in the literature to generate a concordance between two classifications has been developed by Van Beveren et al. (2012). They map all codes into synthetic codes and then analyze product entry and exit using these time-consistent synthetic codes. This is especially advantageous when the unit of observation is the product code itself. Since we are interested when a specific firm introduces a new product or drops an old one, we develop a different procedure. The key advantage is that it does not rely on synthetic codes. Synthetic codes can get quite large and encompass many original product codes. This makes it difficult to interpret these codes still as narrowly defined products if they span groups of multiple products.

We exploit the fact that we study a balanced panel and observe firms both before and after the change in the product classification. Hence, in the year of the classification change, a product code not produced by the firm in the previous year is only counted as a product entry if none of the following applies:

1. There is a one-to-one mapping between the old and new classification. The company produces the old code in the previous year and the new code after the classification change.

2. There is a mapping of many old codes to one new code. The company produces the new code after the classification change and one of the old codes.

3. There is a mapping of one old code to many new codes. The company produces the old code in the previous year and one of the new codes in the year after the classification change.

4. There is a mapping of multiple old codes to multiple new codes. The company produces one of the old codes in the previous year and one of the new

codes in the year after the classification change.

The same procedure is applied to define product dropping, except now the company has to cease production of the code. We thus get a measure of product adding and dropping on the firm level.

B.2 Appendix to Section 2.4: Summary Statistics

In Table B.1 we present some summary statistics for the firms in our sample for the first and last year of study. Given that our control firms have higher employment, they unsurprisingly also have higher R&D expenditures of around €6.7m and sales of €382m. They are somewhat less likely to be in East Germany, mirroring our results in Figure 2.1. Reassuringly, the propensity to have positive R&D expenditures is very similar across the two groups, with around 70 percent of firms engaging in R&D. We do not present results on industry affiliation, as our CEM procedure ensures balancing on this dimension.

Table B.1: Summary Statistics

	Control firms		Treatment Firms	
	2007	2011	2007	2011
R&D spending	6696.16	6485.11	3211.71	3438.63
R&D employees	63.04	61.81	28.19	30.13
R&D indicator	0.70	0.72	0.68	0.70
Sales	382.16	382.94	211.50	217.62
Add product	0.04	0.01	0.03	0.01
Drop product	0.02	0.01	0.02	0.02
East Germany	0.04	0.04	0.08	0.08
Number of firms	225		764	

Note: Source: RDC of the Federal Statistical Office and Statistical Offices of the Federal States, *AFiD Panel Industrieunternehmen [2008-2011]*, own calculations.

C

Appendix to Chapter 3

C.1 Appendix to Section 3.2: Data

C.1.1 Data Sources and Definitions

Table C.1: Data Sources and Definitions

Variable	Description	Source
Dependent Variables		
Employment rate	Employees subject to social security contributions in the county of residence normalized by the working-age population.	Federal Employment Agency (*Bundesagentur für Arbeit*)
Employment p.c. in high-exposure sectors	Employees subject to social security contributions in the county of residence in photovoltaic-related industries (industry codes 31, 321, 332, 401, 453, 454, 518, 519, 524, 731, 742, 743) of the German Classification of Economic Activity, Version 2003, normalized by the working-age population. From 2008 onwards, the original data is classified following the revised German Classification of Economic Activity, Version 2008. The set of industry codes is the union of industry codes of a random sample of	Employment data at the three-digit industry level purchased from the Federal Employment Agency.

Variable	Description	Source
	firms that are members of the German Solar Association (*Bundesverband Solarwirtschaft*) We cross-walk the data from the industry classification in 2008 into the industry classification of 2003 following the official correspondence table.	
Employment p.c. in local services	Employees subject to social security contributions in the county of residence in local, non-tradable industries (wholesale and retail–industry codes G except 518, 519, 524; hospitality–industry code H; and financial services–industry codes 651, 652) of the German Classification of Economic Activity, Version 2003, normalized by the working-age population. The data from 2008 onwards follows a revised industry classification. Converted into the industry classification as of 2003 as described above.	Employment data at the three-digit industry level purchased from the Federal Employment Agency.
Employment p.c. in other sectors	Employees subject to social security contributions in the county of residence in all the industries not included in "high-exposure sector," "local services," and except social services (industry code 853), normalized by the working-age population. The data from 2008 onwards follows a revised industry classification. Converted into the industry classification as of 2003 as described above.	Employment data at the three-digit industry level purchased from the Federal Employment Agency.
Working-age population	The population of working age (between 15 and 65 years of age) in 2003. In our analysis, most variables are normalized by the working-age population (indicated by "p.c." in the variable name).	German Statistical Office, population statistics (code 173-21-4)
Median wage in construction	The monthly gross median wage in construction, averaged over employees and year. The data is only accessible for county-years in which construction employment exceeded 1000.	Wage data purchased from the Federal Employment Agency.
Photovoltaic Investments, Instruments, and Classification of Tight / Slack Labor Markets		
Photovoltaic installations (in MWp)	Capacity and location of each PV system in Germany measured in MWp and day of connection to the energy grid. We aggregate capacity from the project lists using county and municipality identifiers.	Deutsche Gesellschaft für Sonnenenergie; project lists here: `http://www.energymap.info`.

Variable	Description	Source
Rooftop potential	Estimates of rooftop space based on the aerial maps of 4500 dwellings; see Appendix C.1.2 for details.	Lödl et al. (2010)
Solar radiation	Yearly average global irradiance on the optimally inclined surface.	PVGIS project of the European Union
Feed-in tariff	Guaranteed price per kWh of produced electricity for installations with an output capacity of less than 30 kWp.	Renewable Energy Act
Costs of solar installation	Industry survey on total installation costs per kWp.	Janzing (2010); Bundesverband Solarwirtschaft e.V. (2012, 2014)
Interest rate	Average interest rate on mortgage loan (prior to 2003), effective interest rates of commercial banks for housing loans.	Bundesbank (series BBK01.SU0010 and BBK01.SUS131)
Ownership structure	Buildings by type of ownership: multiple ownership, private person, housing cooperative, region or state, municipal housing companies, private housing company, other private companies and non-profits.	Housing questionnaire of the Census 2011
Unemployment rate	Individuals receiving unemployment benefits in the county of residence normalized by the working-age population. At the state and national level, we compute the unemployment rate as the sum of unemployed individuals divided by the sum of the working-age population.	Federal Employment Agency
Control Variables		
County type	Counties comprise either of a single municipality (so-called city counties or *Kreisfreie Städte*) or multiple municipalities (so-called rural counties or *Landkreise*).	Federal Office for Building and Regional Planning (*Bundesamt für Bauwesen und Raumordnung*)
Population growth	The ratio of the working-age population in any given year and the working-age population in 2003.	German Statistical Office, population statistics (code 173-21-4)
Construction	Number of residential and non-residential buildings completed in a given year.	German Statistical Office, construction statistics of completed buildings (code 311-21)
Total area, settlement and dwelling area	The total area of a county in km^2 as of 2008. Includes data on the usage of data for settlement, dwellings and in eleven other categories.	German Statistical Office, area statistics (code 331-11)

Variable	Description	Source
Square meters (living area)	Floor space in residential buildings. Data is measured on 31.12.2008.	German Statistical Office, housing statistics (code 035-21-5)
Apartments/ building	Number of apartments per residential building. Raw data gives number of buildings with 1, 2 or more apartments. For the last category an average of six apartments per building is assumed. Data is measured on 31.12.2008.	German Statistical Office, housing statistics (code 035-21-5)
Population / population density	Total population, measured on 31.12.2008. Population density is population per area in km^2.	German Statistical Office, population statistics (code 173-01-5)
Education shares	Employment Shares by Education. The ratio of employees with a university degree to the total number of employees and the ratio of employees with vocational training to the total number of employees as of Q2 2003. The baseline is the share of employees with less education than vocational training.	Federal Employment Agency
Industry shares	A vector of three variables, all as of Q2 2003: the share of employees in agriculture (industry codes 01–03), the share of employees in manufacturing (industry codes 05–39), and the share of employees in construction (industry codes 41–43). The omitted category is the share of employees in services (industry codes 45–95).	Employment data at the three-digit industry level purchased from the Federal Employment Agency
School & university students p.c.	The official statistics provide the numbers of school students for ten different school types. We use the sum across all school types. Both the number of school students and the number of university students are measured in 2003 and normalized by the working-age population.	German Statistical Office, school statistics (code 192-32-4) University statistics of the German Rectors' Conference (*Hochschulrektorenkonferenz)*
Solar panel manufacturer	Locations of the establishments of German solar panel and components manufacturers.	EEM Energy & Environment Media GmbH

Redistricting

The administrative boundaries of counties changed in three East German states (Saxony-Anhalt in 2007, Saxony in 2008, Mecklenburg-West Pomerania in 2011) during the sample period. These reforms took place in response to declining rural population in East Germany and mainly merged several former counties into a single one in order to save administrative costs. We recalculate all the variables from before the administrative reforms to the level of the county boundaries after the reform. All but three former counties are completely merged into new counties, so that the aggregation of these data is straightforward. For the three counties, whose municipalities are assigned to two or three new counties (*Demmin*, county code 13052, in Mecklenburg-West Pomerania, and *Zerbst/Anhalt*, county code 15151, as well as *Aschersleben-Staßfurt*, county code 15352 in Saxony-Anhalt), we disaggregate each statistic based on the relative population shares before the county merger. That is, if the old county A is split to merge into the new counties B and C and if 2/3 of the pre-reform population of county A will be assigned to county B (leaving 1/3 for county C), we construct (virtual) counties B and C before the reform by assigning 2/3 of the value of each statistic (e.g., employment in manufacturing) from county A to the (virtual) county B and 1/3 of the value of each statistic to the (virtual) county C.

C.1.2 Estimation of Rooftop Potential

Following the approach of Lödl et al. (2010) we estimate the rooftop potential for each county in Germany with the following three steps:

1. We classify each German municipality according to the five criteria in Table C.2 into four types: very rural, rural, suburban or urban.

2. In a next step, we multiply the settlement area of each municipality with the estimated rooftop potential per km^2 of settlement area by municipality type. Lödl et al. (2010) calculate the average rooftop potential for each municipality type shown in Table C.2 based on aerial maps of Bavaria and assumptions on roof angles and exposition.

Table C.2: Estimating Rooftop Potential Following Lödl et al. (2010)

Category	Very Rural	Rural	Suburban	Urban
Thresholds for Classification				
Population	≤ 2000	≤ 5000	≤ 20000	> 20000
Population density (per km^2)	≤ 100	≤ 200	≤ 300	> 300
Settlement area (in km^2)	≤ 0.4	≤ 0.8	≤ 1.5	> 1.5
Living area p.c. (in m^2)	> 48	> 45	> 42	≤ 42
Number of apartments	≤ 1.4	≤ 1.6	≤ 1.8	> 1.8
Rooftop Potential Estimates from Lödl et al. (2010)				
Settlement area per dwelling (in m^2)	3734	1793	795	795
Rooftop potential per dwelling (in kWp)	25.8	13.9	5.7	0.25×5.7
Number of Municipalities				
N	4413	3997	1854	946

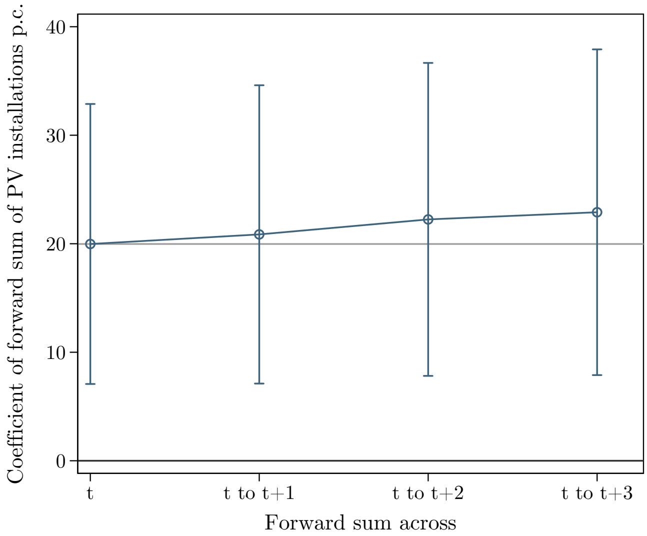
3. In a last step, we aggregate the estimates of rooftop potential of all municipalities to the county level.

C.2 Appendix to Section 3.3: Empirical Strategy

C.2.1 Persistence

In this section, we follow the approach by Shoag (2015) to provide suggestive evidence that PV installations affect employment predominantly in the same year they are installed, without significant long-lasting effects. To this end, we sum the variables on both sides of equation (3.1) over a number of adjacent years and estimate the effect of the sum of PV installations on the sum of employment for the relevant time period. More precisely, we estimate models of the following

Figure C.1: Persistence of Employment Effects



Note: This figure plots the coefficients of the sum of PV installations across adjacent years estimated via model (C.1) along with their 95 percent confidence interval. The coefficient for the sum across t and $t+1$ is obtained by regressing the sum of *employment p.c.* across t and $t+1$ on the sum of *installed capacity p.c.* across t and $t+1$ (along with the sums of all other covariates). The coefficients for the sums across t to $t+2$ and t to $t+3$ are obtained from equivalent regressions. For the sum containing only the current year t, the coefficient is equal to the baseline coefficient of *installed capacity p.c.* in column (1) of Table 3.1.

form for different horizons j:

$$\begin{aligned}\sum_{\tau=t}^{t+j} Employment\ p.c._{c,\tau} = \beta_j \sum_{\tau=t}^{t+j} PV\ Installations\ p.c._{c,\tau} + CountyFE_c \\ +\delta_{c,t}\,\mathbb{1}[Year_t \cdot State_c \cdot CountyType_c] + \sum_{\tau=t}^{t+j} Controls_{c,\tau} + \varepsilon_{c,t}.\end{aligned} \tag{C.1}$$

If the effects of PV investment on employment were long-lasting, the coefficient β_j should rise with longer horizons j as past and current effects accumulate. If, on the other hand, the effects were predominantly contemporary, the coefficients β_j should be close to the baseline effect estimated from model (3.1) regardless of the length of the horizon j.

Figure C.1 shows the coefficients β_j for horizons up to $j = 3$ together with a line at 19.98, which is the baseline effect reported in column (1) of Table 3.1. The coefficients rise at best slightly above the baseline value when adding additional years, suggesting that there are no substantial dynamic effects of PV investment on employment.

C.2.2 Discussion of IV Assumptions

Section 3.3 points out that *remuneration potential* as described in Section 3.3.1 can serve as an instrument for investments in rooftop PV systems. Section 3.4.2 presents the main results when estimating the empirical model (3.1) using *remuneration potential p.c.* as an instrument for *installed capacity p.c.* of rooftop systems. The IV strategy serves as a check for whether unobserved factors drive our results. In this section, we discuss the two main IV assumptions, relevance and exogeneity.

Relevance and First Stage

Table C.3 shows that the time variation in remuneration potential at the county level is a strong predictor of annual PV installations. For the pooled sample of all German counties in column (1), an increase in the remuneration potential of €1m led to additional PV installations of 0.023 MWp (or 23 kWp) on average, similar to the coefficients for the years 2004 and 2010 in Figure 3.3. Given the (weighted) average price of installations of €3,121 per kWp, this implies additional investments of about €72,000. Comparing counties with and without slack labor markets, the average change in investments in response to changes in the remuneration potential tends to be smaller if the labor market is classified as being slack, both according to the time series and the cross-sectional definition. Nevertheless, even then a €1m increase in remuneration potential leads to additional PV installations of at least 8 kWp, corresponding to investments of around €25,000. Moreover, these effects are precisely estimated, so that the remuneration potential is a strong instrument with Kleibergen-Paap F-statistics of 22 and

Table C.3: First Stage

	Installed Capacity p.c. (in MWp)				
Split along		Time series		Cross-section	
	Baseline	Slack	Tight	Slack	Tight
	(1)	(2)	(3)	(4)	(5)
Remuneration p.c.	0.0232***	0.0077***	0.0259***	0.0090***	0.0304***
	(0.0025)	(0.0016)	(0.0034)	(0.0017)	(0.0034)
Population growth	0.0003***	0.0001	0.0001	0.0001	0.0005***
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Construction p.c.	0.0026*	0.0007	0.0025	−0.0016	0.0048**
	(0.0015)	(0.0005)	(0.0025)	(0.0015)	(0.0020)
County FE	yes	yes	yes	yes	yes
Year FE	yes	yes	yes	yes	yes
F-stat. instrument	88.22	22.11	59.15	28.43	79.67
Observations	4000	2044	1956	1783	2189

Note: The dependent variable *installed capacity p.c.* are PV installations measured in megawatt peak (MWp) normalized by the working-age population in 2003 (as indicated by "p.c." for "per capita"). *Remuneration p.c.* is the remuneration potential for PV systems of the size of the county's rooftop potential, given local solar radiation, the current installation costs, and the applicable feed-in tariff. *Population Growth* is the ratio of the working-age population in a given year to the working-age population in 2003. *Construction p.c.* is the number of residential and non-residential buildings completed in a given year. The *year fixed effects* are estimated at the level of the state $\times$ county type (rural or urban county). *F-statistic instrument* is the Kleibergen-Paap F-statistic of the instrument (*remuneration potential p.c.*). Standard errors (in parentheses) are clustered at the level of 94 spatial planning regions. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

higher, well above the critical value of 10.

Exclusion Restriction

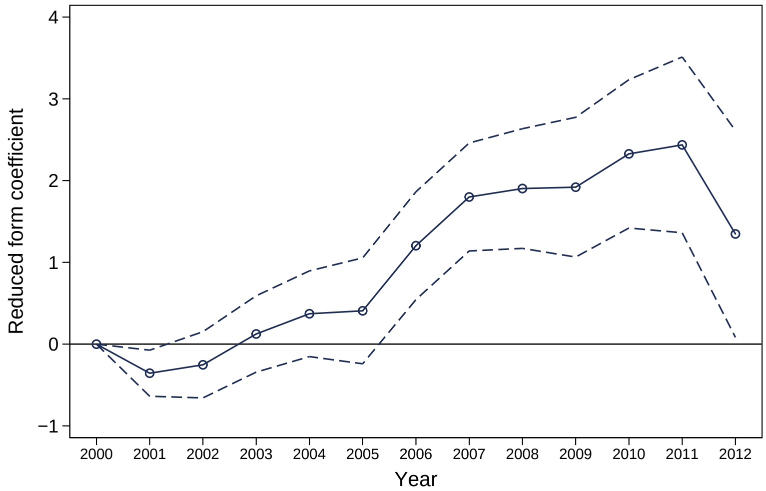
The exclusion restriction requires that conditional on covariates, the instrument does not directly influence employment outcomes. In particular, the instrument is not allowed to influence local employment over and above the common employment trends that is filtered out by the time fixed effects at the state $\times$ county-type level. While this assumption is untestable, it is unlikely that any of the factors that drives the variation in the remuneration potential directly affects the county-specific employment outcomes.

For one, there is no indication for the existence of direct feedback effects between the time-varying components of the remuneration potential (feed-in tariff, installation costs, interest rates) and the heterogeneous local employment outcomes net of the time fixed effects. The feed-in tariff is chosen at the national level in order to ensure a certain volume of renewable energy production, in line with the aim of the Renewable Energy Act. Accordingly, the feed-in tariff has been adjusted in response to total installed capacity, either by amendments of the law (in 2009) or directly linked to the level of past installations by law (from 2009 onwards). The feed-in tariff has never been altered in response to labor market conditions in particular counties or states. Similarly, the changes in the national average costs of PV installations are mainly driven by conditions on the world market for solar panels, which is dominated by Asian manufacturers.[1] In Appendix C.3 we also show that the results are unaffected by excluding those counties that host establishments of German solar panel manufacturers. Finally, the mortgage rate tracks the ECB refinancing rate, which is set for the Euro zone as a whole irrespective of the idiosyncratic labor market conditions in specific German counties.

The exclusion restriction also fails if the cross-sectional variation in *rooftop potential* $\times$ *radiation* is correlated with labor market dynamics. Yet, it is unlikely that the stock of housing foreshadows local labor market dynamics. According to the German Census of 2011, 87.9 percent of private houses were built before 2000 and 95 percent before 2004. Rooftop potential was hence largely fixed before the photovoltaic investment boom between 2004 and 2012. To alleviate the potential concern that our estimates pick up effects from construction nevertheless, we directly control for construction activity in all regressions. Solar radiation, in turn, is unlikely to have a direct impact on labor markets. The climate is temperate across all German regions so that potential effects of heat on labor productivity (e.g., Dell et al., 2012) are irrelevant.

[1] Between 2000 and 2013, the share of German manufacturers in world solar cell production has never exceeded 20 percent, while Asian manufacturers produced at least 48 percent (`http://www.earth-policy.org/data\_center/C23`, last accessed on April 5th 2018).

Figure C.2: Reduced Form Coefficient of Rooftop Potential × Radiation over Time



Note: The connected circles show, for each year between 2001 and 2012, the average differences in employment relative to 2000 for each 1,000 MW of potential peak solar energy production (as measured by the product of rooftop potential and radiation). The 90 percent confidence intervals are plotted as dashed lines. The estimates are obtained from the model described in Footnote 2.

Figure C.2 provides a plausibility check for whether the cross-sectional variation in *rooftop potential* × *radiation* is correlated with employment via channels other than PV installations. It displays the reduced form effect of *rooftop potential* × *radiation* on employment, that is, the average increase in employment relative to 2000 for each 1,000 MW of potential peak solar energy production.[2] The exclusion restriction implies that *rooftop potential* × *radiation* affects employment only via PV installations. The magnitude of the estimated employment effects are hence expected to track the overall time path of PV investments displayed in Panel (b) of Figure 3.1. The results in Figure C.2 support this hypothesis. The

[2]Formally, Figure C.2 displays the estimates $\hat{\gamma}_t$ of the following regression:

$$\begin{aligned} \mathit{Employment\ p.c.}_{c,t} = & \sum_{t=2001}^{2012} \gamma_t \, (\mathit{Rooftop\,Potential}_c \cdot \mathit{Radiation}_c \ \mathit{p.c.}_c) \cdot \mathbb{1}[\mathit{Year}_t] \\ & + \mathit{CountyFE}_c + \tilde{\delta}_{c,t} \, \mathbb{1}[\mathit{Year}_t \cdot \mathit{State}_c \cdot \mathit{CountyType}_c] + \mathit{Controls}_{c,t} + \epsilon_{c,t}. \end{aligned}$$

Rooftop Potential$_c$ · Radiation$_c$ $p.c._c$ is the product of a county's potential for rooftop PV installations (measured in 1,000 MWp) and the county's average solar radiation, normalized by the county's working-age population. As such, it measures the potential yearly energy production in 1,000 MWp under optimal conditions, if the entire available suitable roof space was covered by PV systems. All other variables are as defined in equation (3.1).

employment gains predicted by *rooftop potential* $\times$ *radiation* become statistically and economically different from zero only after the start of the photovoltaic investment boom in 2004, peak at the height of the boom in 2010 and 2011, and drop in 2012 mirroring the drop in investments in this year.

C.3 Appendix to Section 3.4: Main Results

This section evaluates the robustness of the main finding that the employment gains of investments are larger when the labor market is slack compared to when it is tight, both for our OLS and our IV specification.

C.3.1 Robustness of the OLS Results

For brevity, each row in Table C.4 documents the result for a different specification and reports the OLS estimates for the observations with slack and tight labor markets according to both the time series and cross-sectional classifications of the state of the labor market. Columns entitled as "Coeff" report the coefficient estimates for the relevant sample, and columns entitled "SE" present the corresponding standard errors clustered at the level of 94 German spatial planning regions (as in the main specification). The columns "P-Val" contain the p-values of the test of the null hypothesis that the effect of PV installations on employment is smaller in slack than in tight labor markets. For comparison, row (0) reports these statistics for the corresponding empirical specifications from Table 3.1.

Determinants of Rooftop Potential

The first set of robustness checks adds those variables as additional covariates that are components of the highly non-linear estimate of rooftop potential used for constructing the instrument (see Appendix C.1.2 for details on the estimation

Table C.4: Robustness: OLS

	Time series split					Cross-sectional split				
	Slack		Tight		P-Val	Slack		Tight		P-Val
	Coeff	SE	Coeff	SE		Coeff	SE	Coeff	SE	
(0) Baseline	36.61	17.09	2.78	3.86	0.017	37.91	13.83	13.34	6.47	0.036
Determinants of Rooftop Potential										
(1) Area × year	33.62	20.04	−0.35	4.22	0.036	37.04	14.32	16.34	7.47	0.073
(2) Total settlement area × year	36.37	17.25	3.15	3.92	0.019	40.49	14.70	13.40	6.47	0.029
(3) Total dwelling area × year	35.47	17.20	4.04	3.92	0.024	39.88	14.87	12.00	6.72	0.028
(4) Population density × year	30.96	16.55	0.05	3.72	0.023	33.45	12.88	10.50	6.28	0.033
(5) Square meters p.c. × year	29.79	16.48	−0.88	3.88	0.024	35.17	13.67	10.46	6.71	0.036
(6) Apartments/building × year	16.58	15.24	−2.17	3.71	0.092	21.28	9.62	7.80	6.63	0.087
Structural Characteristics										
(7) Education share × year	28.55	16.88	4.23	4.61	0.057	39.79	13.65	18.53	7.13	0.048
(8) School & uni students × year	29.31	16.14	1.25	3.97	0.030	36.28	13.58	10.13	6.70	0.023
(9) Industry share × year	31.09	15.57	−2.09	4.95	0.010	31.21	12.84	14.11	7.81	0.084
(10) Industry & education share × year	28.32	15.22	−0.45	4.87	0.017	30.86	12.70	14.39	7.70	0.087
Model Specification										
(11) Investments uncleaned	14.60	10.76	−1.24	1.03	0.070	4.61	2.06	4.35	3.03	0.469
(12) Without solar panel manufacturers	31.23	18.05	1.43	3.88	0.039	38.30	14.28	14.08	7.13	0.043
(13) Only city counties	71.25	43.62	−3.62	22.21	0.067	24.73	21.63	36.90	33.31	0.632
(14) Only rural counties	30.66	14.96	1.85	3.76	0.018	22.13	7.67	13.60	6.67	0.124

Note: This table presents the results of various modifications of the baseline specification of Table 3.1. It provides OLS estimates of the employment effect of PV installations conditional on the state of the labor market according to both the time series and the cross-sectional split of the sample. Each row of the table represents the result of a different modification of the baseline specification; see the text for details. Columns entitled "Coeff" report the OLS coefficient estimate of *installed capacity p.c.* for the subsamples with slack and tight labor market, respectively. Columns entitled "SE" report the corresponding standard errors, clustered at the level of 94 spatial planning regions. Columns entitled "P-Val" report the p-value of the test of the null hypothesis that the employment effect of PV installations is smaller in a slack labor market than in a tight labor market.

of rooftop potential following Lödl et al., 2010). These variables are geographic characteristics of counties and their municipalities that exhibit little time variation, but that are strongly correlated with total PV installations. We test whether these characteristics are correlated with employment dynamics as well by holding them constant as measured in 2008 (exactly as in the estimation of rooftop potential) and interacting them with year dummies.

Different measures of a county's area constitute the first type of variable used for the estimation of rooftop potential. In addition to the standard covariates, row (1) controls for total county area, row (2) controls for total settlement area (area designated to buildings and transport), and row (3) controls for total dwelling area (area designated to buildings). The second determining factor for the measure of rooftop potential is population density, for which we control in row (4). The final set of variables concerns housing, with the square meters of housing per individual of working age being an additional covariate included in row (5), and the average number of apartments per residential building being added in row (6). Overall, none of the additional covariates changes the results substantially, suggesting that there is no single cross-sectional determinant of PV installations that explains away the main findings. That being said, the addition of the number of apartments per building has the largest effect on the results, leading to a drop in the coefficients by at least one third. A potential reason for this is that the number of apartments per building are strongly correlated with the ownership structure of housing.[3] As installing PV systems requires unanimous consent of all owners of a building, a diverse ownership structure increases the transaction costs of the investment decision, leading to lower investments. As a consequence, the number of apartments per building, via their strong correlation with the ownership structure, is a strong predictor of PV installations, and hence absorbs parts

[3] The correlation coefficient between the number of apartments per building and the number of individually owned buildings per capita is -0.92.

of their employment effects.[4] Nevertheless, even with the number of apartments per building as an additional covariate, the p-value of the null hypothesis that the effect of investments on employment is larger in tight than in slack labor markets remains below 0.1.

Structural Characteristics

The second set of robustness checks explores whether controlling for structural characteristics alters the results. As for the determinants of rooftop potential, these characteristics hardly vary over time, so that we allow for flexible, year-specific effects of the structural characteristics as measured in 2003. We first consider structural features with respect to education, as individuals with different levels of education may face different employment prospects over time. In row (7), we add employment shares by education (with a college degree, with completed vocational training both interacted with year dummies) to the standard set of covariates, and in row (8) we add the number of school and university students (as share of the working-age population and interacted with year dummies) as an additional regressor. Next, we investigate whether the results are driven by industry-specific shocks that may be, for some reason, correlated with PV installations. Row (9) allows for flexible, year specific shocks to the main sectors of the economy—agriculture, manufacturing, and construction (services serve as a baseline)—by including the employment share in each of these industries (interacted with year dummies) as control variables. Finally, row (10) allows for both industry and education specific shocks by adding the employment shares by both industry and education to the empirical model. Neither of these alternative specifications substantially reduces the difference between the employment effects of PV installations across slack or tight labor markets.

[4]We use ownership structure as an alternative instrument in row (17) of Table C.5 in Appendix C.3.2. Table C.8 shows that single ownership is a strong predictor for PV installations.

Model Specification

The last set of robustness checks alters the specification of the empirical model. In the main analyses of the chapter, the variable measuring investments is the sum of a county's installed output capacity of PV systems smaller than 500 kWp. In row (11), we estimate the employment effects of total investments, i.e., the sum of a county's installed capacity regardless of the size of the systems. This results in a few PV systems of large size, most likely greenfield systems, driving a significant amount of the variation in (uncleaned) PV investments. As a result, the OLS coefficients drop significantly, and are equal for slack and tight labor markets in the cross-sectional split. Note, however, that the IV estimates in Table C.5 remain at their baseline level, presumably because the variation of PV installations explained by the instrument—the remuneration potential for rooftop systems—primarily predicts variation in rooftop installations.[5]

Finally, we check whether the composition of the sample has an effect on the results. In row (12), we exclude the 52 counties from the sample that include establishments of solar panel manufacturers.[6] The concern here is that we pollute the estimates of the employment effects of PV installations with employment effects of the solar panel manufacturers, for which the German Renewable Energy Act constituted a significant demand shock, but which also faced increasing competition from abroad. The results with the restricted sample are very close to

[5]One explanation for this finding is that the planning and installation of large greenfield PV systems is undertaken by more specialized firms than the installation of mostly small rooftop systems, so that local variation in demand for greenfield installations does not translate into local employment gains, in contrast to the variation in demand for rooftop systems. While, to the best of our knowledge, there is no hard data on the relative number of firms installing rooftop and greenfield systems, one indication for firms installing greenfield systems being more specialized is that the newest amendment of the Renewable Energy Act prescribes a procurement process for systems larger than 750 kWp. A cursory search for firms installing rooftop and greenfield systems, respectively, also suggests that the latter serve geographically much larger markets.

[6]The web portal "solarserver.de" lists the major German solar panel manufacturers and their locations: `https://www.solarserver.de/service-tools/statistik-und-marktforschung/photovoltaik/unternehmen.html`, last accessed on May 3rd, 2018. About half of the establishments of solar manufacturers are located in former East Germany.

the baseline results, however, so that the main findings are unlikely to be driven by employment in solar panel manufacturing. This also corroborates the findings regarding the employment effects by industry in Section 3.5. Finally, rows (13) and (14) ask whether our findings are driven by city counties (*Kreisfreie Städte*) or rural counties (*Landkreise*). Given that buildings in rural counties are much more suitable for rooftop PV systems due to the availability of larger rooftops that are not shaded by neighboring buildings, it comes with little surprise that our effects are mostly driven by rural counties.

C.3.2 Robustness of the IV Results

Table C.5 performs the same robustness checks using IV as the ones performed via OLS in Table C.4. In addition, Table C.5 also shows that the IV results are robust to alternative definitions of the instrument. Apart from the already well-known differences in the magnitudes of the coefficients, the robustness checks as estimated via IV lead, by and large, to the same conclusions as the ones estimated via OLS. For this reason, we abstain from describing each of the rows in Table C.5, but focus instead on those rows in which the OLS and the IV results differ.

Determinants of Rooftop Potential

The first set of robustness checks adds those variables as covariates that are predictors for the estimate of rooftop potential.[7] Given equation (3.2), the functional form of the instrument *remuneration potential*, these are particularly demanding for the IV strategy. This is due to the fact that controlling more strongly for the cross-sectional determinants of rooftop potential results in identification relying more strongly on the interaction of rooftop potential and radiation. Nevertheless, the main findings are robust to adding these determinants. The most noticeable

[7] Appendix C.1.2 provides the details of this estimation.

Table C.5: Robustness: IV

	Time series split					Cross-sectional split				
	Slack		Tight		P-Val	Slack		Tight		P-Val
	Coeff	SE	Coeff	SE		Coeff	SE	Coeff	SE	
(0) Baseline	148.43	45.38	22.60	10.86	0.003	180.11	37.73	30.52	12.01	0.000
Determinants of Rooftop Potential										
(1) Area × year	317.95	111.89	0.00	14.99	0.002	369.11	116.01	64.54	20.12	0.005
(2) Total settlement area × year	152.88	47.24	21.71	10.74	0.002	216.09	41.65	31.29	11.95	0.000
(3) Total dwelling area × year	147.62	45.20	23.83	10.44	0.002	200.52	40.01	28.87	12.30	0.000
(4) Population density × year	138.95	49.92	14.25	10.94	0.006	178.22	42.07	25.07	13.56	0.000
(5) Square meters p.c. × year	120.86	49.37	6.71	13.22	0.010	200.36	62.49	29.54	14.37	0.003
(6) Apartments/building × year	75.29	50.85	9.50	11.82	0.107	144.03	49.46	16.12	13.88	0.005
Structural Characteristics										
(7) Education share × year	169.61	58.98	29.44	14.46	0.008	240.28	49.87	50.84	16.97	0.000
(8) School & uni students × year	141.48	48.35	25.80	12.52	0.007	200.17	46.28	31.57	15.82	0.000
(9) Industry share × year	172.92	58.21	19.76	16.23	0.004	360.86	208.59	56.36	22.31	0.075
(10) Industry & education share × year	184.38	62.45	24.18	17.11	0.005	395.08	218.71	61.09	22.74	0.066
Model Specification										
(11) Investments uncleaned	111.23	39.84	19.73	9.86	0.010	293.11	163.36	21.08	8.27	0.048
(12) Without solar panel manufacturers	165.87	54.19	23.02	10.93	0.004	196.74	36.69	37.04	12.98	0.000
(13) Only city counties	31.14	434.81	99.41	167.96	0.564	266.39	172.82	265.03	158.37	0.498
(14) Only rural counties	141.37	43.21	20.58	10.35	0.002	97.04	23.95	26.76	13.04	0.003
(15) Instr: costs & income	179.80	55.67	21.48	12.05	0.002	168.24	45.95	30.43	13.98	0.002
(16) Instr: rooftop-p × radiation × year	58.65	24.13	21.43	6.18	0.054	119.72	29.72	25.58	9.72	0.001
(17) Instr: ownership × radiation × year	80.77	31.09	32.87	8.15	0.055	158.23	46.89	22.77	11.87	0.002

Note: This table presents the results of various modifications of the baseline specification of Table 3.2. It provides IV estimates of the employment effect of PV installations conditional on the state of the labor market according to both the time series and the cross-sectional split of the sample. Each row of the table represents the result of a different modification of the baseline specification; see the text for details. Columns entitled "Coeff" report the IV coefficient estimate of *installed capacity p.c.* for the subsample with slack and tight labor market, respectively. Columns entitled "SE" report the corresponding standard errors, clustered at the level of 94 spatial planning regions. Columns entitled "P-Val" report the p-value for the test of the null hypothesis that the employment effect of PV installations is smaller in a slack labor market than in a tight labor market.

difference to the OLS results is that adding the interaction of a county's area with year dummies in row (1) leads to larger coefficient estimates in almost all subsamples. These estimates are, however much less precise than the baseline results. Similar to the OLS estimates, adding the number of apartments per building reduces the estimates, presumably (and as discussed in Appendix C.3.1) due to their strong predictive power for PV installations.

Structural Characteristics

The second set of robustness checks adds structural characteristics as additional covariates. As for OLS, these additional controls do not alter the results substantially. In comparison to the OLS results, the IV coefficients and standard errors are inflated in the cross-sectional split when we control for industry structure. Most likely, this is a result of the instrument becoming weaker due to the addition of a large number of regressors that vary in the cross-section.

Model Specification

The third set of robustness checks modifies the model specification. The biggest difference of the IV estimates to the OLS estimates of Table C.4 in rows (11) to (14) is that the coefficients of the IV results do not drop significantly when we consider all the PV installations in the data (instead of only the smaller systems with a capacity of less than 500 kWp) as our measure of physical investments in row (11). The drop in the OLS results is most likely due to the small local employment effects of large commercial greenfield systems that are installed by specialized firms. In contrast, the instrument captures the potential profitability of *rooftop* systems so that instrumented investments are much less prone to "measurement error" due to considering all investments in solar energy.

Alternative Instrument Definitions

Finally, in rows (15) to (17) we explore whether alternative definitions of the instrument alter the results. The first stages of all alternative instruments are reported below.

Row (15) splits up the time-varying instrument *remuneration potential* as defined by equation (3.2) into the present value of the net income stream (the product of *rooftop potential* and the discounted sum of the net income flows in the second line of (3.2)) and the current installation costs (the product of *rooftop potential* and *costs*), so that there are two time-varying instruments. The remaining two instruments rely on cross-sectional variation only and are interacted with year dummies to obtain (a large number of) time-varying instruments. Row (16) employs *rooftop potential* $\times$ *radiation* as the instrument, with a similar motivation as before: rooftop potential and radiation jointly determine the return on investment. Row (17), in turn, exploits the alternative idea that the ownership structure of buildings affects the transaction costs of installing a rooftop PV system. As mentioned already in Appendix C.3.1, there has to be unanimous consent of all owners of a building for alterations to the building as a whole, including the installation of solar panels. The implied transaction costs are absent for buildings owned by single individuals or firms. The number of buildings with a single owner (relative to the working-age population) is hence a valid instrument if the ownership structure is independent of labor market developments, arguably a stronger assumption than for the stock of available rooftop space. One potential concern for this idea is that in growing economies (and tight housing markets), individuals may be more inclined to join ownership cooperations, invalidating this potential instrument.

The results for all three of these alternative IV strategies show that the estimated magnitudes of the employment gains in slack and tight labor markets do not differ from the baseline estimates in the cross-sectional split. The same is true for the time series split in the specification with costs and income as the instruments (row (15)). Instrumenting via year-specific effects of *rooftop potential* $\times$ *radiation* and *single ownership* $\times$ *radiation* leads to smaller estimated employment gains in the time-series split (rows (16) and (17)). However, the estimates also become more precise, so that we can reject the null hypothesis that the employment gains are smaller in slack than in tight labor markets at the ten percent level (with p-values at or below 0.055).

Tables C.6, C.7, and C.8 report the associated first stages for the alternative instruments used in rows (15), (16), and (17), respectively. In Table C.6, the present value of the net income stream predicts investment positively and costs predict investment negatively, as one would expect. The first stage F-statistics range from 16.25 to 58.65, indicating strong predictive power. Due to the interaction of *rooftop potential* $\times$ *radiation* with year dummies, we have nine instruments in Table C.7. All interactions predict investment positively with F-statistics of over 14.74 in all specifications. There are missing coefficients in the time series split, as in some years not a single county is classified as having either a slack or tight labor market. Last, Table C.8 shows that *single ownership* $\times$ *radiation* ($\times$ *year*) also positively predicts investment. However, in our cross-sectional splits the predictive power is somewhat lower with a F-statistic of around seven.

C.3.3 Robustness of the Slack Definition

To assess whether the difference between slack and tight labor markets is driven by the tails of labor market slackness and tightness, we split the observations into three instead of two groups.[8] For the time series definition, we split the ten years of observations for each county into terciles according to lagged unemployment.

[8] Splitting even finer into quartiles and quintiles runs into the issue of quickly decreasing sample sizes in the individual groups, making valid inference difficult.

Table C.6: First Stage: Costs and Income

	Installed Capacity p.c. (in MWp)				
Split along		Time series		Cross-section	
	Baseline	Slack	Tight	Slack	Tight
	(1)	(2)	(3)	(4)	(5)
Income p.c.	0.0123***	0.0047***	0.0247***	0.0037**	0.0179***
	(0.0018)	(0.0009)	(0.0035)	(0.0015)	(0.0026)
Cost p.c.	−0.0204***	−0.0093***	−0.0289***	−0.0080***	−0.0281***
	(0.0023)	(0.0021)	(0.0038)	(0.0016)	(0.0034)
Population growth	0.0003***	0.0001	0.0000	0.0001	0.0004***
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Construction p.c.	0.0029**	0.0009	0.0025	−0.0014	0.0049***
	(0.0013)	(0.0006)	(0.0023)	(0.0013)	(0.0018)
County FE	yes	yes	yes	yes	yes
Year FE	yes	yes	yes	yes	yes
F-stat. instrument	58.65	16.25	29.15	17.99	54.10
Observations	4000	2044	1956	1798	2202

Note: The dependent variable *installed capacity p.c.* are PV installations measured in megawatt peak (MWp) normalized by the working-age population in 2003 (as indicated by "p.c." for "per capita"). *Income p.c.* is the net present value of the potential income stream for PV systems of the size of the county's rooftop potential, given the local solar radiation and the applicable feed-in tariff. *Cost p.c.* is the time-varying installation cost of PV systems of the size of the county's rooftop potential. Both income and cost are measured in million Euro.*Population Growth* is the ratio of the working-age population in a given year to the working-age population in 2003. *Construction p.c.* is the number of residential and non-residential buildings completed in a given year. The *year fixed effects* are estimated at the level of the state × county type (rural or urban county). *F-stat. instrument* is the Kleibergen-Paap F-statistic of the instruments (*income p.c.* and *cost p.c.*). Standard errors (in parentheses) are clustered at the level of 94 spatial planning regions. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table C.7: First Stage: Rooftop Potential × Radiation × Year

	Installed Capacity p.c. (in MWp)				
Split along		Time series		Cross-section	
	Baseline	Slack	Tight	Slack	Tight
	(1)	(2)	(3)	(4)	(5)
× 2003	0.0000	0.0000	0.0000	0.0000	0.0000
	(.)	(.)	(.)	(.)	(.)
× 2004	0.0051***	0.0051***	0.0187***	0.0042***	0.0063***
	(0.0009)	(0.0011)	(0.0034)	(0.0009)	(0.0016)
× 2005	0.0069***	0.0068***	0.0090**	0.0057***	0.0074***
	(0.0008)	(0.0010)	(0.0039)	(0.0010)	(0.0014)
× 2006	0.0047***	0.0045***	0.0000	0.0040***	0.0050***
	(0.0006)	(0.0007)	(.)	(0.0011)	(0.0013)
× 2007	0.0073***	0.0071***	0.0000	0.0043***	0.0090***
	(0.0013)	(0.0014)	(.)	(0.0013)	(0.0026)
× 2008	0.0136***	0.0075***	0.0029	0.0060***	0.0183***
	(0.0025)	(0.0013)	(0.0040)	(0.0019)	(0.0044)
× 2009	0.0314***	0.0000	0.0212***	0.0130***	0.0419***
	(0.0041)	(.)	(0.0041)	(0.0024)	(0.0063)
× 2010	0.0490***	0.0558***	0.0389***	0.0239***	0.0607***
	(0.0050)	(0.0070)	(0.0050)	(0.0041)	(0.0067)
× 2011	0.0385***	0.0320***	0.0284***	0.0211***	0.0469***
	(0.0034)	(0.0034)	(0.0042)	(0.0024)	(0.0042)
× 2012	0.0218***	0.0000	0.0121***	0.0115***	0.0276***
	(0.0021)	(.)	(0.0040)	(0.0021)	(0.0032)
Population growth	0.0002***	0.0001	0.0001	0.0001	0.0004***
	(0.0001)	(0.0000)	(0.0001)	(0.0001)	(0.0001)
Construction p.c.	0.0026**	0.0008	0.0006	−0.0012	0.0048***
	(0.0012)	(0.0005)	(0.0019)	(0.0012)	(0.0017)
County FE	yes	yes	yes	yes	yes
Year FE	yes	yes	yes	yes	yes
F-stat. instruments	28.22	26.14	28.04	14.74	34.51
Observations	4000	2044	1956	1798	2202

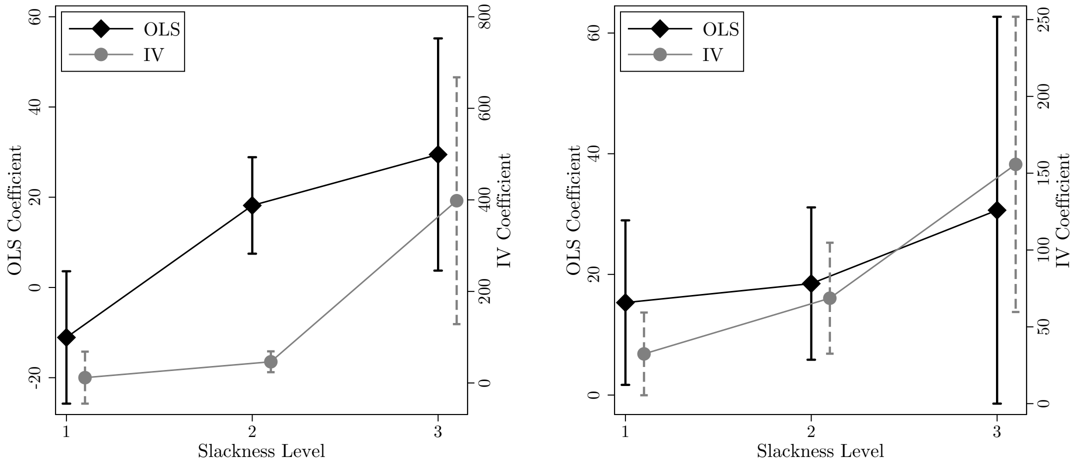
Note: The dependent variable *installed capacity p.c.* are PV installations measured in megawatt peak (MWp) normalized by the working-age population in 2003 (as indicated by "p.c." for "per capita"). *Rooftop potential p.c.* × *radiation* is the interaction of *rooftop potential p.c.* (measured in kWp) and the county's average yearly radiation (measured in kWh). *Rooftop potential p.c.* × *radiation* interacted with year dummies constitutes the set of time-varying instruments. There are missing coefficients in the time series split, as in some years not a single county is classified as having either a slack or a tight labor market. *Population Growth* is the ratio of the working-age population in a given year to the working-age population in 2003. *Construction p.c.* is the number of residential and non-residential buildings completed in a given year. The *year fixed effects* are estimated at the level of the state × county type (rural or urban county). *F-stat. instruments* is the Kleibergen-Paap F-statistic of the instruments. Standard errors (in parentheses) are clustered at the level of 94 spatial planning regions. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table C.8: First Stage: Single Ownership × Radiation × Year

	Installed Capacity p.c. (in MWp)				
Split along		Time series		Cross-section	
	Baseline	Slack	Tight	Slack	Tight
	(1)	(2)	(3)	(4)	(5)
× 2003	0.0000	0.0000	0.0000	0.0000	0.0000
	(.)	(.)	(.)	(.)	(.)
× 2004	0.0781***	0.0809***	0.2390***	0.0432***	0.1280***
	(0.0182)	(0.0250)	(0.0524)	(0.0158)	(0.0391)
× 2005	0.1122***	0.1160***	0.2245***	0.0763***	0.1511***
	(0.0176)	(0.0217)	(0.0457)	(0.0158)	(0.0389)
× 2006	0.0787***	0.0823***	0.0000	0.0626***	0.0740**
	(0.0136)	(0.0172)	(.)	(0.0168)	(0.0360)
× 2007	0.1349***	0.1383***	0.0000	0.0775***	0.1678**
	(0.0272)	(0.0317)	(.)	(0.0164)	(0.0656)
× 2008	0.2555***	0.1330***	−0.0129	0.1237***	0.3678***
	(0.0515)	(0.0250)	(0.0594)	(0.0246)	(0.1169)
× 2009	0.5344***	0.0000	0.3297***	0.2450***	0.8157***
	(0.0942)	(.)	(0.0541)	(0.0451)	(0.1876)
× 2010	0.8016***	1.0217***	0.6010***	0.3939***	1.1540***
	(0.1166)	(0.2240)	(0.0746)	(0.0648)	(0.2223)
× 2011	0.6275***	0.6045***	0.4233***	0.3359***	0.9172***
	(0.0904)	(0.0901)	(0.0578)	(0.0604)	(0.1544)
× 2012	0.3564***	0.0000	0.1630***	0.1877***	0.5172***
	(0.0539)	(.)	(0.0557)	(0.0332)	(0.1122)
Population growth	0.0002**	0.0001	0.0001	0.0001	0.0005***
	(0.0001)	(0.0000)	(0.0001)	(0.0001)	(0.0002)
Construction p.c.	0.0025*	0.0007	0.0011	−0.0007	0.0046**
	(0.0015)	(0.0005)	(0.0023)	(0.0013)	(0.0020)
County FE	yes	yes	yes	yes	yes
Year FE	yes	yes	yes	yes	yes
F-stat. instruments	10.30	13.56	16.92	7.58	6.80
Observations	4000	2044	1956	1798	2202

Note: The dependent variable *installed capacity p.c.* are PV installations measured in megawatt peak (MWp) normalized by the working-age population in 2003 (as indicated by "p.c." for "per capita"). *Single ownership p.c.* × *radiation* is the number of individually owned residential buildings per capita times the county's average yearly radiation (measured in kWh). *Single ownership p.c.* × *radiation* interacted with year dummies constitutes the set of time-varying instruments. There are missing coefficients in the time series split, as in some years not a single county is classified as having either a slack or a tight labor market. *Population Growth* is the ratio of the working-age population in a given year to the working-age population in 2003. *Construction p.c.* is the number of residential and non-residential buildings completed in a given year. The *year fixed effects* are estimated at the level of the state × county type (rural or urban county). *F-statistic instrument* is the Kleibergen-Paap F-statistic of the instruments. Standard errors (in parentheses) are clustered at the level of 94 spatial planning regions. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Figure C.3: Tercile Splits of the Main Slack Definitions



(a) Time Series Split

(b) Cross-Sectional Split

Note: This figure reports the OLS (left axis) and IV (right axis) coefficient and their 95 percent confidence intervals for the main specification from Tables 3.1 and 3.2 with sample splits according to terciles of labor market slackness instead of two groups. Panel (a) reports the result of the tercile split along the time series definition of labor market tightness, and Panel (b) splits the sample according to the cross-sectional definition. The labels on the horizontal axis indicate the tercile of slackness, with "1" indicating the observations with the lowest lagged unemployment rate, and "3" indicating the highest lagged unemployment rate.

For the cross-sectional definition, we split the observations within each state-year cell into terciles according to lagged unemployment.

Figure C.3 depicts the results of this exercise. In Panel (a), we split along the time series dimension and in Panel (b) along the cross-sectional dimension. The horizontal axis shows the different states of the labor market, ranging from 1 (tight) to 3 (slack). The values of the OLS coefficients (black diamonds) are depicted on the left vertical axis and the values of the IV coefficients (grey circles) are shown along the right vertical axis. The capped lines depict 95 percent confidence intervals. For both definitions and for OLS and IV, the effects grow monotonically as the labor market exhibits a higher level of slack. Hence, the difference between slack and tight labor markets does not seem to be solely driven by the tails.

C.4 Appendix to Section 3.5: Discussion of the Mechanism

This section verifies that investment has a stronger effect on employment in slack than in tight markets when allowing for nonlinear effects (Appendix C.4.1). We also perform the analyses outlined in Section 3.5 for the full sample, i.e., without the sample splits (Appendix C.4.2), and use IV instead of OLS for the estimation (Appendix C.4.3).

C.4.1 Nonlinear Employment Effects in Slack and Tight Markets

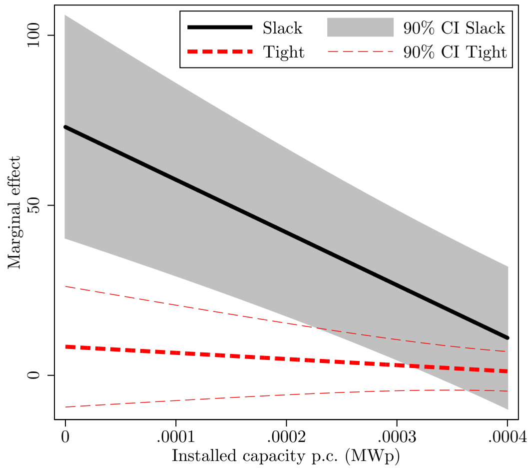
Nonlinearities in the effect of PV installations on employment in tight and slack markets are another potential explanation for the observed differences in employment creation. If the marginal effect of PV installations on employment differs across its domain, and if the mass of *installed capacity* in tight markets is concentrated on values with small marginal effects, and vice-versa for slack markets, this could explain our main findings in Tables 3.1 and 3.2.

To investigate this potential concern, we allow for non-linear effects of PV installations on employment by adding a squared term of $PV\ Installations\ p.c._{c,t}$ to the empirical model (3.1).[9] We estimate this specification for the two main sample splits and compute the marginal effects of PV installations on employment up to the 99th percentile of *installed capacity p.c.*

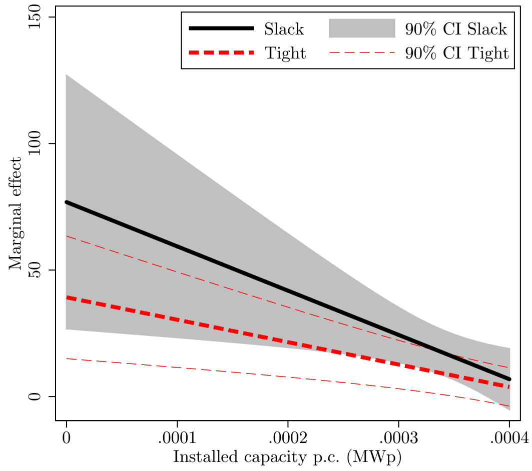
Figure C.4 plots these marginal effects along with their 90 percent confidence intervals. Overall, the marginal effect of PV installations on employment is decreasing for both the time series split in Panel (a) and the cross-sectional split

[9] Adding a cubic term does not alter the results discussed below.

Figure C.4: Marginal Effects Allowing for Nonlinearities



(a) Time Series Split



(b) Cross-Sectional Split

Note: This figure shows the marginal effect of *installed capacity p.c.* on *employment p.c.* estimated via OLS from a variant of model (3.1) that includes $(PV\ Installations\ p.c._{c,t})^2$ as an additional covariate. Panel (a) shows the marginal effects for the time series split, and Panel (b) for the cross-sectional split. The black solid line depicts the marginal effects for the slack labor markets (with shaded 90 percent confidence intervals), and the red dashed line shows the marginal effect for tight labor markets (with thin dashed lines indicating the 90 percent confidence intervals).

in Panel (b). However, the marginal effect is always larger in slack than in tight markets for all levels of installed capacity. The difference is most pronounced for smaller values of *installed capacity p.c.*, which are the observations with the highest mass: The 95th percentile within all subsamples is smaller than 0.00026. As a consequence, these results rule out nonlinearities in the employment response as an explanation for the main findings.

C.4.2 Employment Gains by Sector and Geographic Spillovers: Full Sample

Table C.9 reports the results for employment gains by industry.[10] As should be expected, the employment gains due to PV installations primarily originate from the high-exposure sectors, while local services contribute around 40 percent to the overall employment increase. For the "other" sector the coefficient is eco-

[10] See Section 3.5 for the classification of industries as "high-exposure," "local," and "other".

Table C.9: Sectoral Employment: Baseline

	Industry-specific Employment p.c.					
	OLS			IV		
	High-exp. (1)	Local (2)	Other (3)	High-exp. (4)	Local (5)	Other (6)
Capacity p.c.	15.04***	8.37***	−4.62	23.64***	16.98***	−3.47
	(4.74)	(2.36)	(8.58)	(5.92)	(3.62)	(11.48)
Population growth	0.02*	0.04***	0.30***	0.02	0.04***	0.30***
	(0.01)	(0.01)	(0.02)	(0.01)	(0.01)	(0.02)
Construction p.c.	−0.09	−0.17**	0.11	−0.09	−0.16**	0.11
	(0.12)	(0.08)	(0.24)	(0.12)	(0.08)	(0.24)
County FE	yes	yes	yes	yes	yes	yes
Year FE	yes	yes	yes	yes	yes	yes
Jobs per €100,000	0.48	0.27	−0.15	0.76	0.54	−0.11
F-stat. instr.				88.22	88.22	88.22
Observations	4000	4000	4000	4000	4000	4000

Note: The dependent variable in columns (1) and (4) is employment in the high-exposure sectors (construction and related industries) normalized by the working-age population in 2003 (indicated by "p.c." for "per capita"). The dependent variable in columns (2) and (5) is employment p.c. in local industries (wholesale, retail, hospitality, local services). The dependent variable in columns (3) and (6) is employment p.c. in all remaining industries. Employment by industry is measured annually on June 30th. Table C.1 in Appendix C.1.1 provides details of the industry classifications. All other variables are defined as in Tables 3.1 and 3.2. In columns (4) to (6), installed *capacity p.c.* is instrumented by *remuneration potential p.c.* as defined in Section 3.3.1. Standard errors (in parentheses) are clustered at the level of 94 spatial planning regions. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

nomically close to zero, and we cannot statistically reject the hypothesis that it actually equals zero. The concentration of the effect in the high-exposure sector suggests that we indeed measure the effect of PV investments on employment.

Table C.10 reports the employment gains due to PV installations for the full sample when we account for spatial spillovers.[11] The results show that the spatial spillovers of PV installations are small at best, both when estimated via OLS (Panel A) or IV (Panel B). Compared to the baseline estimates in column (1), the employment effect of within-county PV installations remains unchanged when adding PV installations in neighboring counties as an additional independent variable. Moreover, the estimates of the effect of the neighboring counties' PV installations on employment are at least one order of magnitude smaller than the effect of the within-county PV installations. This holds for all definitions of the set of neighboring counties and for both the OLS and IV estimates. Taken together, the results in Table C.10 suggest that the labor market for PV installations is very local in nature, so that the baseline estimates are a good approximation of the total employment gains due to differential investments across regions.

C.4.3 Employment Gains by Sector and Geographic Spillovers: IV Results

For brevity, Section 3.5 discusses the additional results regarding the employment gains across sectors and geographic spillovers in terms of their OLS estimates. This Appendix reports the IV results of the exact same analyses. The general observation is that the IV results qualitatively mirror the OLS results. The main difference is in the magnitude of the estimates obtained via both strategies, as should be expected given the differences in magnitude in the main OLS and IV results in Tables 3.1 and 3.2, respectively.

[11] See Section 3.5 for the definition of neighboring counties.

Employment Gains by Sector

Table C.11 reports the IV estimates of the employment gains due to PV installations in slack and tight labor markets in each of the subsectors defined by the partition of employees into high-exposure, local, and all other sectors. As for the OLS results in Table 3.4, Panel A presents the results for the time series split, and Panel B presents the results for the cross-sectional split. As before, two results stand out. First, in both slack and tight labor markets, PV installations led to a statistically significant increase in employment only in the high-exposure and local sectors. Second, the difference of the employment gains in slack and tight labor markets is driven by high-exposure and local industries. In both the time series and the cross-sectional split, the difference in employment gains between slack and tight labor markets is sizable. Also, for the IV results we reject the null hypothesis that the employment gains are smaller in slack than in tight labor markets for the high-exposure and local sector in both sample splits. This is a stronger result than the one obtained from OLS, where we reject this hypothesis only for half of the respective coefficients. In sum, the IV results are hence similar if not stronger than the findings from OLS reported in Section 3.5.

Geographic Spillovers

Next, we use the IV strategy to test for geographic spillovers along the line of the corresponding OLS analysis in Section 3.5. As for OLS, we consider three definitions of neighboring counties: all other counties within the same spatial planning region (*Raumordnungsregion*), the five closest counties based on the distance between both counties' most populous municipalities, and the ten closest counties. For the IV specification, we instrument for the neighbors' investments via the sum of the estimated remuneration potential in the neighboring counties, normalized by the working-age population of the county of interest. Given this, we estimate an extended version of the main empirical model (3.1) that includes aggregate PV installations in the neighboring counties as additional covariate and that instruments the county's own as well as the neighboring PV installations via the county's own and the neighbors' aggregate remuneration potential. We classify counties as having slack or tight labor markets according to their own unemployment rate as described in Section 3.3.2, exactly as in the main empirical analyses.

Table C.12 reports the IV estimates. Panel A contains the results of the time series split and Panel B contains the results of the cross-sectional split. In both splits and in all three definitions of a county's set of neighbors, the effect of additional PV installations in geographically proximate regions is at least one order of magnitude smaller than the effect of additional installations within the county. In addition to their small magnitude, these coefficients are mostly statistically insignificant. The estimated effects of the demand spillovers also do not differ between slack and tight labor markets, while the differences of the employment gains due to the within-county investments remain at the same level as in Table 3.2, the main IV specification. As for the OLS results in Table (3.5), we hence conclude that the employment effects of PV installations are very local in nature, so that demand spillovers are unimportant for the interpretation of our main findings.

C.4.4 Wage Response: IV Results

Table C.13 presents the wage response from estimating equation (3.3) using IV instead of OLS. Here, we instrument average yearly PV installations with the average yearly remuneration potential. Similarly to the OLS results in Table 3.6, there is an overall positive effect of PV installations on wage growth in the construction sector. Once again, this is mainly driven by tight labor markets both in the cross-sectional and time series split, though the difference is only statistically significant for the time series split. Thus, this also points towards crowding out as the most plausible mechanism.

Table C.10: Spillovers from Neighboring Counties: Baseline

	Employment Rate			
	Base-line (1)	Planning Region (2)	5 Closest Counties (3)	10 Closest Counties (4)
	Panel A: OLS			
Capacity p.c. (MWp)	19.98***	19.38***	20.20***	18.96***
	(6.35)	(5.99)	(5.63)	(5.87)
Neighboring cap. p.c.		0.32	−0.08	0.20
		(1.21)	(0.85)	(0.49)
Jobs per €100,000	0.64	0.62	0.65	0.61
	Panel B: IV			
Capacity p.c. (MWp)	52.57***	50.20***	52.77***	50.98***
	(13.60)	(13.76)	(13.63)	(13.25)
Neighboring cap. p.c.		1.17	−0.06	0.28
		(2.01)	(1.49)	(0.73)
Jobs per €100,000	1.68	1.61	1.69	1.63
F-stat. instrument(s)	88.22	45.15	53.92	47.72
PopGrowth & constr.	yes	yes	yes	yes
County fixed effects	yes	yes	yes	yes
Year fixed effects	yes	yes	yes	yes
Observations	4000	4000	4000	4000

Note: *Neighboring cap. p.c.* is the sum of PV installations (measured in MWp and normalized by the working-age population) across all other counties in the same spatial planning region (column (2)), the 5 closest counties (column (3)), or the 10 closest counties (column (4)). Closeness is measured by the distance between the counties' most populous municipalities. In Panel B, *capacity p.c.* and *neighboring capacity p.c.* are instrumented by *remuneration potential p.c.* as defined in Section 3.3.1 and the sum of *remuneration potential* in the set of neighboring counties, normalized by the (main) county's working-age population. The *year fixed effects* are estimated at the level of the state × county type (rural or urban county). All other variables are defined as in Tables 3.1 and 3.2. Standard errors (in parentheses) are clustered at the level of 94 spatial planning regions. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table C.11: Sectoral Employment Conditional on Slack: IV Results

	Industry-specific Employment p.c.					
	High-exposure		Local		Other	
	Slack (1)	Tight (2)	Slack (3)	Tight (4)	Slack (5)	Tight (6)
	Panel A: Time Series Split					
Capacity p.c.	36.99**	6.92*	42.81***	10.45***	18.55	4.23
	(15.95)	(3.57)	(12.99)	(3.29)	(34.57)	(9.28)
P-val. slack < tight	0.033		0.006		0.342	
Jobs per €100,000	1.18	0.22	1.37	0.33	0.59	0.14
F-stat. instrument	22.11	59.15	22.11	59.15	22.11	59.15
Observations	2044	1956	2044	1956	2044	1956
	Panel B: Cross-Sectional Split					
Capacity p.c.	52.13**	12.56**	44.63***	11.90***	42.88	−3.86
	(24.26)	(5.71)	(16.60)	(3.39)	(32.22)	(11.85)
P-val. slack < tight	0.057		0.026		0.081	
Jobs per €100,000	1.67	0.40	1.43	0.38	1.37	−0.12
F-stat. instrument	28.43	79.67	28.43	79.67	28.43	79.67
Observations	1783	2189	1783	2189	1783	2189
Controls	yes	yes	yes	yes	yes	yes
County FE	yes	yes	yes	yes	yes	yes
Year FE	yes	yes	yes	yes	yes	yes

Note: The dependent variable in columns (1) and (2) is employment in the high-exposure sectors (construction and related industries) normalized by the working-age population in 2003 (indicated by "p.c." for "per capita"). The dependent variable in columns (3) and (4) is employment p.c. in local industries (wholesale, retail, hospitality, local services). The dependent variable in columns (5) and (6) is employment p.c. in all remaining industries. Employment by industry is measured annually on June 30th. Table C.1 in Appendix C.1.1 provides details of the industry classifications. *Capacity p.c.* are yearly PV installations measured in megawatt peak (MWp), which are instrumented by *remuneration potential p.c.* as defined in Section 3.3.1. Except for the dependent variables, the empirical specifications are identical to the one in Table 3.2. In particular, controls are population growth and new construction. The *year fixed effects* are estimated at the level of the state × county type (rural or urban county). Panel A reports the results for the time series split and Panel B reports the results for the cross-sectional split. Standard errors (in parentheses) are clustered at the level of 94 spatial planning regions. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table C.12: Spillovers from Neighboring Counties: IV Results

	Employment Rate					
	Planning Region		5 Closest Counties		10 Closest Counties	
	Slack (1)	Tight (2)	Slack (3)	Tight (4)	Slack (5)	Tight (6)
	Panel A: Time Series Split					
Capacity p.c.	144.40***	18.35	147.19***	18.90	144.50***	14.84
	(48.55)	(11.57)	(46.72)	(12.03)	(47.33)	(11.86)
Neighboring capacity p.c.	1.92	1.87	0.33	1.04	0.68	1.24
	(5.61)	(1.39)	(4.58)	(1.43)	(2.29)	(0.83)
P-val slack < tight	0.005		0.004		0.003	
Jobs per €100,000	4.63	0.59	4.72	0.61	4.63	0.48
F-stat. instr.	10.95	28.74	15.95	32.42	10.96	31.66
Observations	2044	1956	2044	1956	2044	1956
	Panel B: Cross-Sectional Split					
Capacity p.c.	193.21***	33.49**	200.80***	40.22***	201.16***	39.14***
	(41.08)	(13.01)	(39.31)	(13.85)	(39.51)	(13.12)
Neighboring capacity p.c.	−2.37	−2.21	−3.16*	−3.62	−1.67	−1.94*
	(2.36)	(2.72)	(1.78)	(2.31)	(1.03)	(1.11)
P-val slack < tight	0.000		0.000		0.000	
Jobs per €100,000	6.19	1.07	6.43	1.29	6.44	1.25
F-stat. instr	11.72	40.21	10.43	40.70	10.60	39.03
Observations	1783	2189	1783	2189	1783	2189
Controls	yes	yes	yes	yes	yes	yes
County FE	yes	yes	yes	yes	yes	yes
Year FE	yes	yes	yes	yes	yes	yes

Note: *Neighboring cap. p.c.* is the sum of PV installations (measured in MWp and normalized by the working-age population) across all other counties in the same spatial planning region (columns (1) and (2)), the 5 closest counties (columns (3) and (4)), or the 10 closest counties (columns (5) and (6)). Closeness is measured by the distance between the counties' most populous municipalities. Installed *capacity p.c.* and *neighboring capacity p.c.* are instrumented by *remuneration potential p.c.* as defined in Section 3.3.1 and the sum of *remuneration potential* in the set of neighboring counties, normalized by the (main) county's working-age population. *F-statistic instruments* reports the Kleibergen-Paap F-statistic of both excluded instruments. Controls are population growth and new construction per capita. The *year fixed effects* are estimated at the level of the state × county type (rural or urban county). All other variables are defined as in Table 3.2. Panel A reports the results for the time series split and Panel B reports the results for the cross-sectional split. Standard errors (in parentheses) are clustered at the level of 94 spatial planning regions. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table C.13: Wage Growth (IV Specification)

	Δ Log(Median Wage in Construction)				
		Split along			
		Time series		Cross-section	
	Baseline	Slack	Tight	Slack	Tight
Years	03-12	03-07	08-12	03-12	03-12
	(1)	(2)	(3)	(4)	(5)
Avg. yearly capacity p.c.	7.30^{*}	-15.61	20.42^{***}	-3.36	10.60^{*}
	(4.35)	(24.25)	(5.10)	(11.06)	(5.41)
Population growth	-0.01	-0.11	0.18^{***}	-0.04	0.03
	(0.05)	(0.07)	(0.07)	(0.11)	(0.07)
Avg. yearly construction p.c.	0.16	0.33	-0.88^{***}	0.60	-0.06
	(0.20)	(0.26)	(0.30)	(0.44)	(0.26)
State × county-type FE	yes	yes	yes	yes	yes
P-value slack > tight		0.080		0.131	
F-statistic instrument	125.29	68.68	127.53	56.47	87.81
Observations	368	363	370	159	209

Note: The dependent variable is the difference in the log median wage in the construction sector over the years indicated in the row *"Years"* (referred to as "sample years"). *Avg. yearly capacity p.c.* are the average yearly PV installations during the sample years measured in megawatt peak (MWp) normalized by the working-age population in 2003 (indicated by "p.c." for "per capita"). This is instrumented with the average remuneration potential per capita over the sample years. *Population Growth* is the difference in the working-age population over the sample years relative to the working-age population in 2003. *Avg. yearly construction p.c.* is the average yearly number of residential and non-residential buildings completed during the sample years. For the time-series defintion of slack in columns (2) and (3) we split the sample into the years 2003 to 2007 (slack) and 2008 to 2012 (tight). In the cross-section in columns (4) and (5), we split the sample relative to the mean of the average unemployment rate at the state-level in 2002. *P-Val slack > tight* reports the p-value of the test of the null hypothesis that the wage effect of PV installations is larger in slack than in tight labor markets. The number of observations is smaller than the number of counties (400) because the wage data is only available for county-years in which the number of employees in construction exceeds 1000. Standard errors (in parentheses) are clustered at the level of 94 spatial planning regions. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Bibliography

Acconcia, Antonio, Giancarlo Corsetti, and Saverio Simonelli (2014): "Mafia and Public Spending: Evidence on the Fiscal Multiplier from a Quasi-Experiment," *The American Economic Review*, 104(7), 2185–2209.

Adelino, Manuel, Igor Cunha, and Miguel A. Ferreira (2017): "The Economic Effects of Public Financing: Evidence from Municipal Bond Ratings Recalibration," *The Review of Financial Studies*, 30(9), 3223–3268.

Aerts, Kris and Tobias Schmidt (2008): "Two for the price of one?: Additionality effects of R&D subsidies: A comparison between Flanders and Germany," *Research Policy*, 37(5), 806–822.

Akcigit, Ufuk, Douglas Hanley, and Nicolas Serrano-Velarde (2013): "Back to Basics: Basic Research Spillovers, Innovation Policy and Growth," Working Paper 19473, National Bureau of Economic Research.

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

Almeida, Heitor and Murillo Campello (2007): "Financial constraints, asset tangibility, and corporate investment," *The Review of Financial Studies*, 20(5), 1429–1460.

Almus, Matthias and Dirk Czarnitzki (2003): "The effects of public R&D subsidies on firms' innovation activities: the case of Eastern Germany," *Journal of Business & Economic Statistics*, 21(2), 226–236.

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

——— (2010): "The credibility revolution in empirical economics: How better research design is taking the con out of econometrics," *Journal of Economic Perspectives*, 24(2), 3–30.

Aschhoff, Birgit (2009): "The effect of subsidies on R&D investment and success: do subsidy history and size matter?" ZEW Discussion Papers 09-032, ZEW – Leibniz Centre for European Economic Research.

Auerbach, Alan J. and Yuriy Gorodnichenko (2012): "Measuring the Output Responses to Fiscal Policy," *American Economic Journal: Economic Policy*, 4(2), 1–27.

——— (2013): "Fiscal Multipliers in Recession and Expansion," in *Fiscal Policy after the Financial Crisis*, University of Chicago Press, 63–98.

Autor, David, David Dorn, and Gordon H. Hanson (2013): "The China Syndrome: Local Labor Market Effects of Import Competition in the United States," *American Economic Review*, 103(6), 2121–68.

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.

Bachmann, Rüdiger and Eric R. Sims (2012): "Confidence and the Transmission of Government Spending Shocks," *Journal of Monetary Economics*, 59(3), 235–249.

Becker, Bettina (2015): "Public R&D policies and private R&D investment: A survey of the empirical evidence," *Journal of Economic Surveys*, 29(5), 917–942.

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.

Benedictus, Rinze, Frank Miedema, and Mark W.J. Ferguson (2016): "Fewer numbers, better science," *Nature News*, 538(7626), 453.

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

Bickel, Peter, Tobias Kelm, Jochen Mayer, Frithjof Staiss, Ole Langniss, and Dietmar Edler (various years): "Evaluierung der KfW-Förderung für Erneuerbare Energien im Inland," *Gutachten für die KfW Bankengruppe*.

Biolsi, Christopher (2017): "Nonlinear Effects of Fiscal Policy over the Business Cycle," *Journal of Economic Dynamics and Control*, 78, 54 – 87.

Bloom, Nicholas, Mark Schankerman, and John Van Reenen (2013): "Identifying technology spillovers and product market rivalry," *Econometrica*, 81(4), 1347–1393.

Boockmann, Bernhard, Claudia M. Buch, and Monika Schnitzer (2014): "Evidenzbasierte Wirtschaftspolitik in Deutschland: Defizite und Potentiale," *Perspektiven der Wirtschaftspolitik*, 15(4), 307–323.

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.

Brohm, Rainer (2010): "Marktentwicklung und Perspektiven der Photovoltaik in Deutschland," Tech. rep., BSW-Solar.

Bronzini, Raffaello and Eleonora Iachini (2014): "Are incentives for R&D effective? Evidence from a regression discontinuity approach," *American Economic Journal: Economic Policy*, 6(4), 100–134.

Bronzini, Raffaello and Paolo Piselli (2016): "The impact of R&D subsidies on firm innovation," *Research Policy*, 45(2), 442–457.

Brückner, Markus and Anita Tuladhar (2014): "Local Government Spending Multipliers and Financial Distress: Evidence from Japanese Prefectures," *Economic Journal*, 124(581), 1279–1316.

Bryan, Kevin A. and Jorge Lemus (2017): "The direction of innovation," *Journal of Economic Theory*, 172, 247–272.

Buch, Claudia M., Katja Patzwald, Regina T. Riphahn, and Edgar Vogel (2019): "Verstehen - Entwickeln - Testen - Verbessern: Rahmenbedingungen für evidenzbasierte Politik," *Wirtschaftsdienst*, 99(2), 106–112.

Buchheim, Lukas and Martin Watzinger (2017): "The Employment Effects of Countercyclical Infrastructure Investments," Mimeo.

Buchheim, Lukas, Martin Watzinger, and Matthias Wilhelm (forthcoming): "Job Creation in Tight and Slack Labor Markets," *Journal of Monetary Economics*.

Bundesministerium für Finanzen (2009): "Informationen zum Konjunkturpaket vom Januar 2009," Monatsbericht des BMF, Berlin.

Bundesverband Solarwirtschaft e.V. (2012): "Statistische Zahlen der deutschen Solarstrombranche (Photovoltaik)," Berlin.

——— (2014): "Statistische Zahlen der deutschen Solarstrombranche (Photovoltaik)," Berlin.

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

Caggiano, Giovanni, Efrem Castelnuovo, Valentina Colombo, and Gabriela Nodari (2015): "Estimating Fiscal Multipliers: News from a Nonlinear World," *Economic Journal*, 125(584), 746–776.

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.

Chodorow-Reich, Gabriel (forthcoming): "Geographic Cross-Sectional Fiscal Spending Multipliers: What Have We Learned?" *American Economic Journal: Economic Policy*.

Chodorow-Reich, Gabriel, Laura Feiveson, Zachary Liscow, and William Gui Woolston (2012): "Does State Fiscal Relief During Recessions Increase Employment? Evidence from the American Recovery and Reinvestment Act," *American Economic Journal: Economic Policy*, 4(3), 118–145.

Cohen, Lauren, Joshua Coval, and Christopher Malloy (2011): "Do Powerful Politicians Cause Corporate Downsizing?" *Journal of Political Economy*, 119(6), 1015–1060.

Congressional Budget Office (2009): "Estimated Macroeconomic Impacts of the American Recovery and Reinvestment Act of 2009," Washington D.C.

——— (2014): "Estimated Impact of the American Recovery and Reinvestment Act on Employment and Economic Output in 2014," Washington D.C.

Conley, Timothy G. and Bill Dupor (2013): "The American Recovery and Reinvestment Act: Solely a Government Jobs Program?" *Journal of Monetary Economics*, 60(5), 535–549.

Czarnitzki, Dirk and Cindy Lopes-Bento (2014): "Innovation subsidies: Does the funding source matter for innovation intensity and performance? Empirical evidence from Germany," *Industry and Innovation*, 21(5), 380–409.

David, Paul A., Bronwyn H. Hall, and Andrew A. Toole (2000): "Is public R&D a complement or substitute for private R&D? A review of the econometric evidence," *Research Policy*, 29(4), 497–529.

De Blasio, Guido, Davide Fantino, and Guido Pellegrini (2014): "Evaluating the impact of innovation incentives: evidence from an unexpected shortage of funds," *Industrial and Corporate Change*, 24(6), 1285–1314.

Dechezleprêtre, Antoine, Elias Einiö, Ralf Martin, Kieu-Trang Nguyen, and John Van Reenen (2016): "Do Tax Incentives for Research Increase Firm Innovation? An RD Design for R&D," Working Paper 22405, National Bureau of Economic Research.

Dell, Melissa, Benjamin F. Jones, and Benjamin A. Olken (2012): "Temperature Shocks and Economic Growth: Evidence from the Last Half Century," *American Economic Journal: Macroeconomics*, 4(3), 66–95.

Depner, Heiner, Natalia Gorynia-Pfeffer, Carsten Lohmann, Wolfgang Möller, and Ingrid Voigt (2011): "Wirksamkeit der aus dem Konjunkturpaket II geförderten FuE Projekte des Zentralen Innovationsprogramms Mittelstand (ZIM)," Expertise im Auftrag des Bundesministeriums für Wirtschaft und Energie.

Depner, Heiner, Tim Vollborth, Julia Wolff von der Sahl, and Natalia Gorynia-Pfeffer (2018): "Wirksamkeit der geförderten FuE Projekte und Kooperationsnetzwerke des Zentralen Innovationsprogramms Mittelstand (ZIM)," Expertise im Auftrag des Bundesministeriums für Wirtschaft und Energie.

Dube, Arindrajit, Thomas Hegland, Ethan Kaplan, and Ben Zipperer (2018): "Excess Capacity and Heterogeneity in the Fiscal Multiplier: Evidence from the Recovery Act," Mimeo.

Dupor, Bill and Rodrigo Guerrero (2017): "Local and Aggregate Fiscal Policy Multipliers," *Journal of Monetary Economics*, 92, 16–30.

EuPD Research (2013): "German PV Module Price Monitor 2013," Bonn.

EY (2018): "Worldwide R&D Incentives Reference Guide," `https://www.ey.com/gl/en/services/tax/worldwide-r-d-incentives-reference-guide---country-list`, last accessed 13 March 2019.

Farhi, Emmanuel and Iván Werning (2016): "Fiscal Multipliers: Liquidity Traps and Currency Unions," in *Handbook of Macroeconomics*, ed. by John B. Taylor and Harald Uhlig, Elsevier, vol. 2, 2417–2492.

Fazzari, Steven M., James Morley, and Irina Panovska (2014): "State-dependent Effects of Fiscal Policy," *Studies in Nonlinear Dynamics & Econometrics*, 19(3), 285–315.

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

Fuchs-Schündeln, Nicola and Tarek A. Hassan (2016): "Natural Experiments in Macroeconomics," in *Handbook of Macroeconomics*, ed. by John B. Taylor and Harald Uhlig, Elsevier, vol. 2, 923–1012.

González, Xulia and Consuelo Pazó (2008): "Do public subsidies stimulate private R&D spending?" *Research Policy*, 37(3), 371–389.

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

Görg, Holger and Eric Strobl (2007): "The effect of R&D subsidies on private R&D," *Economica*, 74(294), 215–234.

Grembi, Veronica, Tommaso Nannicini, and Ugo Troiano (2016): "Do Fiscal Rules Matter?" *American Economic Journal: Applied Economics*, 8(3), 1–30.

Guceri, Irem and Li Liu (2019): "Effectiveness of Fiscal Incentives for R&D: Quasi-experimental Evidence," *American Economic Journal: Economic Policy*, 11(1), 266–91.

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.

Haltiwanger, John, Ron S. Jarmin, and Javier Miranda (2013): "Who creates jobs? Small versus large versus young," *Review of Economics and Statistics*, 95(2), 347–361.

Hopenhayn, Hugo and Francesco Squintani (2016): "The Direction of Innovation," Mimeo.

Howell, Sabrina T. (2017): "Financing innovation: evidence from R&D grants," *The American Economic Review*, 107(4), 1136–1164.

Huld, Thomas, Richard Müller, and Attilio Gambardella (2012): "A New Solar Radiation Database for Estimating PV Performance in Europe and Africa," *Solar Energy*, 86(6), 1803–1815.

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.

Jaffe, Adam B. and Trinh Le (2015): "The Impact of R&D Subsidy on Innovation: a Study of New Zealand Firms," Working Paper 21479, National Bureau of Economic Research.

Janzing, Bernward (2010): "Innovationsentwicklung der Erneuerbaren Energien," *Renews Spezial*, 37, 4–13.

Kolata, Gina (2009): "Grant System Leads Cancer Researchers to Play it Safe," *The New York Times*, A1.

Kunkel, André (2010): "Kreditklemme: Gefahr erkannt, Gefahr gebannt?" *ifo Schnelldienst*, 63(09), 32–36.

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.

Lödl, Martin, Georg Kerber, Rolf Witzmann, Clemens Hoffmann, and Michael Metzger (2010): "Abschätzung des Photovoltaik-Potentials auf Dachflächen in Deutschland," in *11. Symposium Energieinnovation Alte Ziele - Neue Wege*, ed. by Technische Universität Graz, Graz: Verlag der Technischen Universität Graz.

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

——— (2018): "Randomization Inference for Differences-in-differences with Few Treated Clusters," Tech. rep., Carleton University, Department of Economics.

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

Michaillat, Pascal (2014): "A Theory of Countercyclical Government Multiplier," *American Economic Journal: Macroeconomics*, 6(1), 190–217.

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

Nakamura, Emi and Jón Steinsson (2014): "Fiscal Stimulus in a Monetary Union: Evidence from US Regions," *American Economic Review*, 104(3), 753–92.

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

Owyang, Michael T., Valerie A. Ramey, and Sarah Zubairy (2013): "Are Government Spending Multipliers Greater during Periods of Slack? Evidence from Twentieth-Century Historical Data," *American Economic Review*, 103(3), 129–134.

Ramey, Valerie A. (2011): "Identifying Government Spending Shocks: It's all in the Timing," *The Quarterly Journal of Economics*, 126(1), 1–50.

Ramey, Valerie A. and Sarah Zubairy (2018): "Government Spending Multipliers in Good Times and in Bad: Evidence from US Historical Data," *Journal of Political Economy*, 126(2), 850–901.

Rendahl, Pontus (2016): "Fiscal Policy in an Unemployment Crisis," *Review of Economic Studies*, 83(3), 1189–1224.

Roulleau-Pasdeloup, Jordan (2017): "The Government Spending Multiplier in a Deep Recession," Mimeo.

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

Schnitzer, Monika and Martin Watzinger (2017): "Measuring the spillovers of venture capital," Mimeo.

Shoag, Daniel (2015): "The Impact of Government Spending Shocks: Evidence on the Multiplier from State Pension Plan Returns," Mimeo.

Smets, Frank and Ana Lamo (2009): "Wage Dynamics in Europe: Final Report of the Wage Dynamic Network (WDN)," *European Central Bank*.

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

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

——— (2017): "Kostenstrukturerhebung im Verarbeitenden Gewerbe sowie des Bergbaus und der Gewinnung von Steinen und Erden – Qualitätsbericht," Wiesbaden.

——— (2018): "Ausgaben, Einnahmen und Personal der öffentlichen und öffentlich geförderten Einrichtungen für Wissenschaft, Forschung und Entwicklung," 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.

Stifterverband für die deutsche Wissenschaft (2018): "Forschung und Entwicklung in der Wirtschaft 2016," .

Suárez Serrato, Juan Carlos and Philippe Wingender (2016): "Estimating Local Fiscal Multipliers," Working Paper 22425, National Bureau of Economic Research.

Taylor, John B. (2016): "The Staying Power of Staggered Wage and Price Setting Models in Macroeconomics," in *Handbook of Macroeconomics*, ed. by John B. Taylor and Harald Uhlig, Elsevier, vol. 2, 2009–2042.

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.

Van Beveren, Ilke, Andrew B. Bernard, and Hylke Vandenbussche (2012): "Concording EU Trade and Production Data over Time," Working Paper 18604, National Bureau of Economic Research.

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.

Wirth, Harry (2015): "Aktuelle Fakten zur Photovoltaik in Deutschland," Fassung vom 15.05.2015.

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.

Zhao, Bo and Rosemarie H. Ziedonis (2012): "State governments as financiers of technology startups: Implications for firm performance," Mimeo.

Zúñiga-Vicente, José Ángel, César Alonso-Borrego, Francisco J. Forcadell, and José I. Galán (2014): "Assessing the effect of public subsidies on firm R&D investment: a survey," *Journal of Economic Surveys*, 28(1), 36–67.