Essays in Applied Econometrics

Inauguraldissertation

zur Erlangung des akademischen Grades
eines Doktors der Wirtschaftswissenschaften
der Universität Mannheim

vorgelegt von
Maximilian Mähr
aus Reutlingen

Mannheim
2025

Abteilungssprecher	Prof. Dr. Thomas Tröger
Referent	Prof. Dr. Antonio Ciccone
Koreferent	Prof. Dr. Ulf Zölitz
Verteidigung	Mannheim, den 27. Mai 2025

To my parents, Thomas and Sibylle.

Acknowledgements

This thesis marks the end of a long and intense journey – one I could not have completed without the help and support of many people.

First and foremost, I would like to thank my supervisors, Antonio Ciccone and Ulf Zölitz, for their support, encouragement, and time throughout my Ph.D. studies. Antonio, your ability to challenge my thinking while offering thoughtful guidance has been crucial to my development as a researcher. Your insights and high standards consistently pushed me to refine my work, and your influence on how I approach problems has been more profound than you probably realize. Ulf, from the moment I joined you as a pre-doc, your belief in me and my work has been a constant source of motivation. Your openness and encouragement have supported me at every step, and your mentorship has been invaluable in shaping this thesis.

Special thanks go to my co-authors. This thesis would not have been possible without you. Giuseppe, I hope you and your son get to celebrate many Scudettos in the years to come. Jens, you became not only an intellectual sparring partner but also one of my closest friends – even if I sometimes failed to return your calls.

I am grateful to the amazing people I have met during my Ph.D. studies. Marina, Ursula, and Matilde, you made the challenges of the first year both bearable and enjoyable. David and Felix, thank you for tolerating my endless rants about incorrectly clustered standard errors and for the engaging discussions over countless dinners. Society has gained great economists in all of you.

There are many others I have crossed paths with throughout my academic journey who have helped make this thesis possible – far more than I can acknowledge here. Thank you all for your sharp minds and support.

My deepest gratitude goes to my family. To my great-aunt, Bille, thank you for showing me that wisdom and curiosity know no age. To my siblings, Benni and Bina, thank you for filling my life with laughter, support, and unforgettable memories. I am incredibly fortunate to have you both by my side. To my parents, Thomas and Sibylle, I am privileged to have been raised by such wonderful individuals. Thank you for always being there – I love you. To my son, Lovis, who came into my life during the final stages of this journey, your joy has made reaching the finish line so much easier. Thank you for reminding me of what truly matters.

Most of all, I want to thank my best friend and partner, Reka. Thank you for being my safe space, for understanding me without words, and still being able to laugh with me along the way. I will always remember how deeply you believed in my potential to realize my dreams, even when I could not see it myself. None of this would have been possible without you.

It has been a tough journey, but it has been worth the effort.

Contents

List of Tables

List of Figures

Introduction

Choosing a title for this thesis proved unexpectedly challenging. Ideally, a thesis title offers a concise yet meaningful summary of years of research. While a title like *Essays in Economics* might have sufficed, it felt too broad to capture the specificity of my work. Without the inclusion of the third chapter, *Essays in the Science of Sciences* might have been an option, but it would have suggested a depth of expertise in the science system that I do not claim. Ultimately, *Essays in Applied Econometrics* emerged as the most fitting choice, reflecting the thesis's primary focus: developing and employing quantitative methods to causally answer three empirically motivated questions.

Applied econometrics not only provides the methodological foundation for this thesis, but is also the field I have enjoyed and excelled in the most throughout my studies. At its core, applied econometrics refers to the use of econometric methods on real-world data to analyze quantitative models. When applied thoughtfully – and often creatively – it offers a powerful framework for causal inference, the disentanglement of exogenous influences from endogenous confounders. While these methods may sometimes appear artificial or overly meticulous to the outside reader, they are essential for deriving credible and robust conclusions. This thesis demonstrates how these methods can be leveraged to gain insights into three pressing socio-economic questions.

Chapter 1 investigates how appointing a female professor through affirmative action affects universities' hiring decisions and gender attitudes. Despite an increase in the share of women pursuing academic careers, women currently hold only one in four professorships. This gender imbalance has motivated the implementation of affirmative action policies, such as Germany's *Professorinnenprogramm*, which provides financial incentives to universities for appointing women to full professorships. These interventions are subject to considerable debate, with proponents claiming they are essential to address systemic barriers hindering women's advancement, while opponents fear they may undermine merit-based hiring. Employing an instrumental variable design that leverages retirement probabilities among existing professors as an instrument for subsidy uptake, I find that the *Professorinnenprogramm* significantly increases the likelihood of appointing women as full professors. However, these appointments do not appear to influence subsequent female hiring at the professorial level, and have limited trickle-down effects for junior researchers. Notably, the program does significantly increase the share of female Ph.D. students, especially among those who completed their undergraduate studies in the same department, suggesting a role model effect. Despite these changes, there is no measurable impact on research productivity or thematic focus within departments. Finally, I estimate that two-thirds of subsidized appointments would have occurred without the program, implying that departments strategically use subsidies to hire women they would have hired anyway.

Chapter 2 investigates the impact of restricted access to scientific knowledge on the production and dissemination of new research, leveraging the quasi-natural experiment created by the rise of *Sci-Hub*. While the internet has dramatically reduced the marginal cost of distributing scientific articles, access to the majority of peer-reviewed research remains restricted behind paywalls, with only approximately 20% of journals offering open access. This restriction raises questions about whether these financial and legal barriers inhibit the production of knowledge, a critical factor for economic growth. Despite potentially significant impacts, rigorous evidence on this issue is limited due to the endogenous nature of journal access. To address this challenge, this chapter, co-authored with Jens, uses the emergence of *Sci-Hub*, an online platform providing free access to paywalled academic articles, as a natural experiment. Employing an instrumental variable approach based on social connectedness to Almaty, Kazakhstan – *Sci-Hub*'s origin – we analyze global data on platform usage and scientific output to identify causal effects. Our findings indicate that increased access via *Sci-Hub* significantly increases the consumption of paywalled research, as measured by a greater share of references to closed-access journals. Furthermore, we find evidence that researchers in regions with higher *Sci-Hub* usage produce more highly cited work, indicating improvements in research quality, though these effects do not extend to publications in higher-ranking journals or shifts in research topics. The findings highlight the transformative potential of open access in democratizing scientific advancement.

Chapter 3, co-authored with Alida Sangrigoli, Giuseppe Sorrenti, and Gilberto Turati, explores the intersection of public health and governance. We evaluate how media coverage of corruption scandals affects the behavior of public healthcare workers. Focusing on Italy's National Health Service, the study analyzes the reactions of medical staff involved in organ procurement to two prominent corruption scandals – one involving a hospital manager and the other a surgeon. Using a difference-in-differences approach, we compare regions with varying levels of media exposure to these scandals to assess their impact on reported organ donations. Our findings show that media coverage of the surgeon scandal, but not the manager scandal, leads to a significant decline in organ donor reports, indicating that healthcare workers are especially sensitive to corruption within their professional ranks.

Chapter 1

Consequences of Affirmative Action: The Impact of Hiring a Female Professor

1.1 Introduction

Despite an increase in the share of women pursuing academic careers, women currently hold only one in four professorships (European Commission, 2021). In response, policies meant to strengthen the presence of women among professors are becoming increasingly common. These policies include quotas for female recruitment (NRW, 2014; Wallon, Bendiscioli and Garfinkel, 2015), female quotas in funding schemes (National Health & Medical Research Council, 2022), and mandated female representation on academic evaluation panels (Swiss National Science Foundation, 2021).

However, diversity policies are controversial. Proponents argue that intervening in the labor market's matchmaking process is necessary to overcome institutional barriers that impede women's advancement to leadership positions (Mengel, Sauermann and Zölitz, 2019; Card et al., 2020; Dupas et al., 2021; Kleemans and Thornton, 2021; Sarsons et al., 2021; Janys, 2024). Exposure to women can break down negative perceptions by allowing them to demonstrate their capabilities

* I am grateful to Antonio Ciccone, Ulf Zölitz, Jens Oehlen, Giuseppe Sorrenti, Fabian Waldinger, Michèle Tertilt, Philipp Ager and Camille Urvoy, for helpful comments and encouragement. I thank audiences at the University of Mannheim, University of Munich and Stockholm University for comments. The author declares no relevant or material financial interests that relate to the research described in this paper. All errors are my own.

(Dahl, Kotsadam and Rooth, 2021) and create an environment that supports the advancement of other women through role model effects (Jensen and Oster, 2009; Porter and Serra, 2020). Opponents argue that in the absence of highly qualified women, diversity policies may undermine merit-based hiring and deepen the quality gap between male and female candidates. This may reinforce negative stereotypes by displacing competent men with less qualified women and possibly lead to resistance from within targeted organizations (Whelan and Wood, 2012; Besley et al., 2017).

Hence, some studies support diversity policies among professors. Others do not. Surprisingly, we lack empirical evidence on how deliberately increasing the representation of women among professors impacts universities.

In this paper, I provide such evidence by analyzing an affirmative action policy introduced by the German Ministry of Education, the *Professorinnenprogramm*. The program subsidizes the first-time appointment of women to permanent full professorships, offering up to 825,000 Euros per position over five years. Since its inception in 2008, the program has supported the appointment of 845 women, 12% of all female professorship appointments in Germany.

For identification, I exploit the program's subsidy allocation process. Universities that pass an initial application process become eligible for up to three subsidies. Eligible universities then allocate these subsidies across their departments. To address endogeneity in subsidy allocation, I exploit the program's requirement that subsidized appointments must be permanent appointments, which requires permanent financing by the university once the five-year subsidy has expired. This requirement increases the likelihood that subsidized appointments are assigned to departments with a high probability of full professor retirements during or following the subsidy period. Retirement probabilities satisfy the exclusion restriction, as they are determined by historical hiring patterns and are very difficult to adjust given the regulation of retirement in German public universities. Institutional constraints further reinforce this argument: departments cannot independently create new permanent positions – these require negotiations with the federal states and are typically only justified in response to increased teaching demands – nor can they demand or incentivize early retirement. I strengthen this design by also considering retirement probabilities of departments in ineligible universities – those rejected in the initial application stage. This additional cross-sectional variation helps disentangle retirement-driven trends from program effects, allowing identification under considerably weaker assumptions.

I find that my instrument for subsidy uptake is a strong predictor of female hiring. At eligible universities, a 10 percentage-point higher probability of experiencing at least one retirement within the next five years is associated with a 4.7 percentage-point higher probability of appointing a female professor compared to ineligible universities, beyond their pre-existing hiring differences. I validate my identification strategy through multiple robustness checks. Retirement probabilities do not predict female hiring outside subsidy periods or at ineligible universities. Additionally, the observed correlation clearly stands out from a distribution of placebo estimates generated by randomly reassigning university eligibility and departmental retirement probabilities.

Following the appointment of a female professor, the subsequent hiring of full professors remains unchanged. My findings suggest that affirmative action neither facilitates nor impedes the advancement of other women to full professorships.

Among junior researchers, trickle-down effects are limited: there is no statistically significant change in female hiring at the assistant professor or postdoctoral level. However, at the Ph.D. level, the number of women increases by 19%, rising to 29% for doctoral students who completed undergraduate studies in the same department. I provide evidence that this effect is likely to be driven by increased interaction between female students and newly appointed female professors. This mechanism aligns with existing research identifying role models as crucial factors for the advancement of female academics (Porter and Serra, 2020; Blau et al., 2010; Ginther et al., 2020).

Next, I examine how exposure to a female professor influences collaboration patterns. Existing research suggests that exposure to underrepresented groups can reduce stereotypes and increase future engagement with those groups (Carrell, Hoekstra and West, 2015). I hypothesize that if negative stereotypes exist, shifts in gender attitudes might be reflected in the share of female co-authors. Overall, I find no significant increase in female co-authorship. However, when disaggregating effects by gender, I observe a modest rise in male faculty co-authoring with women, particularly two to three years after a female professor joins the department. This effect is primarily driven by junior male faculty – defined as tenured professors with below-median experience – who exhibit a 24% increase in female co-authorship. Further analysis suggests that this pattern emerges mainly from new mixed-gender collaborations originating within the peer network of newly appointed female professors. Taken together, these results indicate that gender attitudes are malleable through increased exposure to women.

I also assess whether the presence of an additional female professor affects a department's research productivity. Neither the quantity nor quality of publications – measured through journal rankings and citations – shows a noticeable shift. Additionally, I investigate whether existing department members engage with new research areas following the arrival of a female professor. Prior studies suggest that women often prioritize different research topics (Dolado, Felgueroso and Almunia, 2012; Chari and Goldsmith-Pinkham, 2017), potentially influencing their colleagues' research trajectories. My analysis of department-specific topic profiles reveals no significant thematic shifts.

Finally, I quantify the program's effectiveness in generating female professorships that would not have occurred in the absence of subsidies. To do so, I compare changes in female hiring across fields with high and low shares of subsidized appointments, relative to the pre-funding period. This analysis relies on substantially stronger identifying assumptions than the previous analysis. Most importantly, it assumes that trends in female hiring would have evolved similarly across fields in the absence of the program, once field-specific linear time trends are accounted for. My estimates suggest that roughly two-thirds of subsidized female hires would have been made in the absence of the program, implying that departments strategically use subsidies to hire women they would have recruited anyway. Based on my estimates, it takes approximately 2.9 subsidized appointments – costing approximately 2.2 million Euros – to generate one additional female professor who would not have been hired without the program.

My study adds to a line of research on how diversity in leadership roles impacts the advancement of women. Most existing studies focus on corporate settings and elections. For example, gender quotas in local governments in India yield mixed results regarding increased women's political participation (Chattopadhyay and Duflo, 2004; Bhavnani, 2017). Beaman et al. (2009) find that female representation reduced gender disparities in aspirations and education through role model effects. However, women running for re-election do not result in increased entry of new women candidates (Bhalotra, Clots-Figueras and Iyer, 2018). Conversely, De Paola, Scoppa and Lombardo (2010) document that a short-term gender quota in local government in Italy boosted women's political participation. In Norway, gender quotas on corporate boards had limited impact beyond immediate changes in board composition (Bertrand et al., 2019). In academia, several studies suggest that female role models can influence the career choices of female students. Porter and Serra (2020) show that exposure to a successful female alumna increases the likelihood of female students choosing an economics major by 89%. Carrell, Page and West (2010) find that top female students at the US Air Force Academy are 26 percentage points more likely to complete a STEM major when taught by female instructors. Bagues et al. (2023), in their analysis of Spain, investigate the appointment of female professors and its impact on future hiring and Ph.D. enrollment. Their identification strategy relies on random assignment of full professorship applicants to peer evaluators. Unlike my findings, they report no effect on the share of female Ph.D. students, though they do not specifically focus on students who completed their undergraduate studies in the same department, where I find the most significant increase. In addition, their analysis includes the hiring of tenured associate professors, whereas my study focuses exclusively on full professorships. This distinction may be important, as newly appointed full professors in Germany typically have greater autonomy and resources, including the capacity to recruit and fund Ph.D. candidates. In contrast, newly hired professors in many Spanish universities often lack comparable institutional support or funding, which may limit their ability to supervise or employ Ph.D. students.

A related set of studies evaluates how diversity affects performance. Ahern and Dittmar (2012), Matsa and Miller (2013), and Nygaard (2011) evaluate the effect of Norway's board composition quota on firm performance and governance, finding no definitive results. Kim and Starks (2016) demonstrate that gender diversity on U.S. corporate boards can enhance firm valuation, driven by the contributions of female directors. In Italy, Flabbi et al. (2019) show that female corporate leadership positively impacts the upper end of the female wage distribution while negatively affecting the lower end, with overall firm performance benefiting from a higher proportion of female workers. Hoogendoorn, Oosterbeek and Van Praag (2013) analyze the impact of gender diversity on business team performance in a field experiment, finding that mixed-gender teams outperform male-dominated teams in terms of profit and sales. I extend the existing literature by not only assessing changes in hiring but also analyzing the broader effects of these appointments on research output and collaboration patterns. Achieving gender parity might lead to more balanced policy recommendations and a broader range of research questions, as women tend to have different policy priorities compared to men. For instance, surveys among economists indicate that women are generally more supportive of government intervention and environmental regulation, whereas men are more inclined to prioritize economic growth and are less concerned about inequality (Chari and Goldsmith-Pinkham, 2017; May, McGarvey and Kucera, 2018). I contribute to this literature by showing that the appointment of a female professor does not lead to a thematic shift in the department's overall research agenda, nor is there a clear tendency to focus more on female-related topics. I am unaware of other studies that causally identify the effect of diversity on the direction of academic research.

Further, my study contributes to a body of research that evaluates policies aimed at increasing the representation of women among full professors. Appendix Figure 1.B.1 indicates that among currently evaluated policies, mentoring programs stand out as the sole measure efficiently increasing the proportion of female full professors. Blau et al. (2010) and Ginther et al. (2020) demonstrate in a randomized control trial that junior female economists in the U.S., when mentored by a senior woman, achieve significantly higher tenure rates (+77%), top-tier publications (+175.6%), and grants (+294.8%). However, the current share of senior female professors is too low to support the large-scale implementation of mentoring programs. In addition, mentoring programs are costly as they burden already stretched senior female researchers (Vernos, 2013; Guarino and Borden, 2017) and their efficiency is likely to decrease with the number of participating junior women.[1] Other policies have proved inefficient in increasing the share of women among professors.[2] Bagues, Sylos-Labini and Zinovyeva (2017) evaluate the random assignment of academics to hiring committees in Italy and Spain, finding that the presence of a female evaluator can reduce female candidates' chances of success by around 5.3% in Italy and 3.3% in Spain. Deschamps (2018) documents a similar effect in sign and magnitude when evaluating gender quotas in academic hiring committees in France. Antecol, Bedard and Stearns (2018) show that 'tenure clock stopping policies' do not significantly affect tenure rates and can even disadvantage female candidates when also men are eligible. Notably, no prior research evaluates the efficiency of affirmative action policies in academia, despite theoretical work highlighting their efficacy (Siniscalchi and Veronesi, 2020). This paper addresses this gap. I show that two-thirds of subsidized female appointments – an implicit affirmative action policy – would have occurred in the absence of the program, suggesting that departments often strategically utilize

[1] Another concern with mentoring is the potential for self-image bias among advisors (Lewicki, 1983). Mentors will likely advise young researchers to become like them and adopt their research characteristics. In the process, young researchers give up some of their characteristics. While this may improve female participation, it shifts the research characteristic distribution toward the mentor, leading to the under-representation of valuable research characteristics in the limit; assuming that all research characteristics are equally valuable in the research process (Siniscalchi and Veronesi, 2020).

[2] Carnes et al. (2015) and Devine et al. (2017) document a significant rise in female hires following a gender bias workshop at the University of Wisconsin-Madison. However, they cannot rule out that their effects are driven by unobserved field- or university-specific factors as the intervention only took place at a single university.

subsidies to hire women they would have hired anyway. This implies that the cost of an additional female professorship is 2.1 million Euros; roughly three times the cost of a subsidized appointment.

The rest of the paper is organized as follows. The next section outlines the institutional setting. Section 1.3 describes the data. Section 1.4 outlines the empirical framework, the results of which are discussed in Section 1.5. Section 1.6 concludes.

1.2 Background

In 2023, approximately 2.9 million students were enrolled in institutions of higher education in Germany. Of these, 50% attended public universities, 37.5% were enrolled at universities of applied sciences, and 8.2% at private universities. The remaining 4.3% studied at specialized institutions such as universities of public administration, art and music colleges, teacher training colleges, and theological colleges (CHE Hochschuldaten, 2024). While all types of institutions of higher education can apply for funds from the *Professorinnenprogramm*, this analysis focuses exclusively on Germany's 83 public universities, as listed in Appendix Table 1.A.1. Other institutions, such as teacher training colleges and universities of applied sciences, are excluded, as they primarily offer practice-oriented, career-focused education.

Public universities are autonomous entities under state oversight, with most state constitutions granting them the right to self-administer within the framework of the respective State Higher Education Act (*Landeshochschulgesetz*). This autonomy leads to substantial variation in legal rules and regulations across institutions. The following overview outlines the most common practices.

1.2.1 German University System

More than two-thirds of the financial resources for universities in Germany are provided by the states, while the federal government contributes 20% (HRK, 2024). Each federal state allocates funding to its universities based on factors such as student enrollment and research performance. University budgets are typically set annually or biennially through negotiations between universities and the federal ministries of research. For example, BW-MF (2022) lists the specific budgets for universities in Baden-Wuerttemberg for the year 2022.

German universities typically organize themselves into specialized departments, each focused on a specific academic field, such as economics. Related fields are grouped into faculties; for instance, economics is usually part of the social sciences faculty. The Federal Statistical Office of Germany recognizes 33 distinct academic fields and eight faculties, as outlined in Appendix Table 1.A.2. On average, each university encompasses sixteen fields, resulting in a total of 1,342 unique departments. Within departments, academic leadership is often divided among chairs, which are organizational units led by professors.

Full Professors While some German universities have introduced tenure-track systems, they are not yet widely adopted. Consequently, most German universities do not follow the U.S. model of categorizing professors into assistant, associate, and full ranks. Instead, the primary distinction is between full professors and assistant professors.[3] Full professors are permanent civil servants, while assistant professors hold temporary positions that may be either tenure-track or non-tenure-track positions. As permanent civil servants, full professors enjoy job security, with dismissal only possible in cases of severe misconduct.[4] The statutory retirement age is gradually rising, starting at 65 for individuals born before 1946 and reaching 67 for those born after 1964. Professors have the option of early retirement from the age of 63 associated with a pension reduction of 3.6% per year. Upon request, they can extend retirement to their 70th birthday (§14 in BMJ (2024)).[5] Salary and pension payments are complemented by costs supporting the professor in fulfilling their duties, like academic support staff and research equipment.

To manage and forecast these costs, professors (as well as other civil servant positions) are assigned designated positions in the federal states' budget, so-

[3] Additionally, there are special cases such as endowed professorships (funded by third-party sources like companies), joint professorships (co-funded with non-university research institutions), honorary professors (part-time university lecturers), and guest professors, which I exclude from the analysis.
[4] The termination of the civil servant status is regulated in §31 BBG, §22 BeamtStG, and §§32 – 36 BBG, §23 BeamtStG.
[5] In particular, retirement can be extended to the age of 68 if there are no conflicting institutional reasons, while postponement until the age of 70 requires a compelling institutional interest.

called 'Planstellen'.[6] While departments are autonomous in the appointment of individuals to a professorship, the creation and renewal of 'Planstellen' can only be authorized in accordance with the federal state. Usually, the creation of new full professorships is tied to predictable long-term increases in teaching demand resulting from higher student demand or the accreditation of new study programs.[7]

Appointing Full Professors A full professorship is filled through a formalized appointment procedure.[8] First, the position is publicly advertised, often internationally.[9] Then, the department selects an appointment committee, which oversees the entire appointment process and is tasked with finding and recruiting the most suitable candidates for the position. The committee consists of department members but can sometimes also include external members. After the application deadline, a hearing is conducted, inviting the most promising candidates. The hearing typically includes a public seminar and interviews with department members. After the hearing, the committee selects the most suitable

[6] Appendix Figure 1.B.2 displays an excerpt from the 2019 budget of the state of Baden-Wuerttemberg, detailing the employment plan for the University of Mannheim.

[7] To predict changes in teaching demand, administrators compare future teaching demand with contemporaneous teaching resources. The stock of teaching resources can be calculated by weighting the department personnel by their position-specific teaching obligations. Teaching obligations differ by positions and state and are normally measured in teaching units. Usually, a teaching unit roughly translates to 15 lectures of 45 minutes per semester (this excludes pre- and post-lecture preparation). For example in the state of North Rhine-Westphalia full professors are assigned nine teaching units per semester, while assistant professors are assigned five teaching units per semester (NRW, 2009). Future teaching demand is calculated by multiplying the anticipated student body by the average course-specific teaching load. Usually, students are expected to participate in 20 lectures of 45 minutes per week. Short-term deviations can be addressed by hiring lecturers on a temporary basis. For instance, in 2013, a significant increase in temporary lecture positions occurred following a school reform, which saw the completion of two secondary school classes, effectively doubling the count of first-year university students.

[8] To formally qualify as a university professor, candidates must show "additional scientific qualifications" beyond their PhD. In Germany, this is often done through a habilitation, an academic exam that demonstrates competence in both research and teaching. Alternatively, candidates can fulfill this requirement through an assistant professorship or by proving "equivalent achievements". What counts as equivalent varies by field and is not standardized. It might include work similar to a habilitation thesis or a set of published articles (cumulative habilitation).

[9] Although a public and, in most cases, international advertisement for a vacant professorship is generally legally required, there are circumstances where this requirement can be waived entirely or the appointment process significantly simplified. The specific state higher education laws outline varying conditions for such cases. For example, no advertisement is necessary if a temporary civil servant or employee position is to be converted into a permanent one or in the case of the availability of an exceptionally qualified individual whose recruitment is of special interest to the university. In some federal states, the Ministry of Science must also approve the advertisement of the professorship, while at some universities, this decision lies with the academic senate.

candidates and requests external, independent evaluations. Following this, the appointment committee ranks and nominates up to three top candidates. After an offer is made, negotiations over the offer occur in a meeting with the dean and the rectorate, covering details like additional compensation and research budgets. Following the negotiation, the university extends a written offer to the candidate. If declined, subsequent candidates are considered until the position is filled or a new advertisement is required.

Assistant Professors Although tenure-track assistant professorships were introduced in 2002, two-thirds of assistant professorships remain non-tenure track. Non-tenure track positions are typically six-year temporary civil service roles. After this period, candidates undergo an evaluation and, if positively assessed, may receive a two-year extension. While candidates are free to apply for permanent positions – such as full professorships – at other institutions at any time[10], they cannot formally request an internal evaluation for a permanent role at their current institution.

In contrast, tenure-track assistant professorships are designed to transition into permanent positions following a successful final evaluation. A key distinction is that candidates in tenure-track roles are entitled to request an internal evaluation for a permanent appointment. Like non-tenure track positions, tenure-track roles are temporary civil service appointments within the W1 salary bracket.

The hiring process for assistant professorships mostly mirrors that of full professors.

Other Researchers The remaining academic personnel within a department includes post-doctoral researchers, doctoral candidates, and research assistants, who are primarily engaged in research activities. They are supported by lecturers and teaching assistants, whose roles are more focused on instruction.

In the German system, doctoral candidates are often hired directly by individual professors – typically through their chairs – rather than through centralized graduate schools, as is more common in the US or UK. This more personalized hiring process may help explain some of the effects observed on Ph.D. recruitment following professorial appointments.

[10] The situation is different for those pursuing a habilitation, who typically remain at the same institution until the process is complete.

1.2.2 Professorinnenprogramm

The *Professorinnenprogramm* is an affirmative action policy initiated by the German Ministry of Education to enhance the representation of women among full professors. The program provides a five-year subsidy of up to 825,000 Euros (165,000 Euros per year) to cover costs associated with the initial appointment of women to full professorships. These expenses include the professor's salary, as well as costs for support staff and research equipment. Universities that successfully complete an initial application procedure can receive subsidies for up to three positions.

Subsidies are contingent on two conditions. First, they are limited to women being appointed to a full professorship in Germany for the first time. Second, the subsidies are available only for permanent positions. This typically requires either the creation of a new budgeted permanent position in coordination with the federal state or the availability of an existing vacant permanent position. The program also supports 'early appointments' of female full professors – defined as appointments to positions that are not yet permanently budgeted – provided there is a guaranteed transition to a regular, budgeted professorship by the end of the subsidy period.

Initiated in 2008 with a budget of 150 million EUR (wave 1), the program was renewed in 2013 (wave 2), 2018 (wave 3), and 2023 (wave 4), each time with subsequently increasing budgets (see Table 1.3). The budgets for each wave are distributed in two application calls, detailed in Table 1.1. Universities that receive positive evaluations in the first call of a wave cannot reapply in the second call of the same wave. My period of analysis covers the first three funding waves. I define each unique combination of funding wave and call as a funding period, sequentially labeled by $g \in G \equiv \{1, 2, 3, 4, 5, 6\}$, with $\tau \equiv \tau(g)$ mapping to the year the evaluation results for funding period g are announced.

Application Process The *Professorinnenprogramm* employs a structured procedure for allocating subsidies.[11] All institutions of higher education are eligible to apply. Participation requires submitting an application to the German Federal Ministry of Education. The application consists of a fifteen-page document detailing statistics and plans related to the gender equality concept. The document comes in two parts. The first part describes the current representation of women

[11] A formal description of the application process is provided in Bundesanzeiger (2018). Appendix Figure 1.B.3 provides a chronological sequence of the application process.

Table 1.1: Application Timeline by Funding Period

Wave	Call	g	Announcement	Application Deadline	Application Results	Appointment Deadline
1	1	1	10/03/2008	16/06/2008	03/09/2008	31/12/2009
1	2	2	10/03/2008	02/03/2009	04/06/2009	31/12/2010
2	1	3	06/12/2012	28/03/2013	11/07/2013	31/12/2014
2	2	4	06/12/2012	28/03/2014	03/07/2014	31/12/2015
3	1	5	10/11/2017	29/05/2018	08/11/2018	31/12/2019
3	2	6	10/11/2017	29/05/2019	05/11/2019	31/12/2020

Note: The table displays how each distinct combination of funding wave and call corresponds to a funding period, denoted sequentially by $g \in G \equiv \{1, 2, 3, 4, 5, 6\}$. The last two columns indicate the respective deadlines and announcement dates associated with each funding period.

at different qualification levels, including statistics on the share of women across departments and ranks over time. The second part outlines existing and planned measures aimed at (1) increasing the proportion of women in top academic positions, (2) promoting career and professional development for young female scientists, and (3) encouraging female student enrollment in underrepresented fields. Universities that received funding in previous calls of the program must provide evidence of successful implementation of their prior equality concept. Importantly, the first-stage application does not specify the positions to be financed, which will only be addressed in the second stage. On average, 82% of universities applied in each of the last three program waves.

Following submission, a twelve-member review panel evaluates all applications. The German Ministry of Education, in consultation with state education ministries, selects the panel members based on disciplinary diversity, representation from major German science organizations, and international expertise. If an application is approved, the ministry commits to funding the initial appointment of up to three female full professors, provided the budget allows.[12] For example, university B may be deemed eligible, while university A is not (Figure 1.1c).

The selection criteria are opaque and not publicly disclosed, nor does the ministry publicly disclose evaluation details or rankings. To gain insight into the selection process, I conduct a text analysis of publicly available application documents. However, because not all universities publish their applications, the analysis may be subject to selection bias. The analysis reveals that neither linguistic

[12] In the third program wave, the ten universities with the highest-ranked applications could receive funding for a fourth appointment.

characteristics nor specific topics within the documents predict eligibility status. A detailed breakdown of this analysis is provided in Appendix Section 1.C.1.

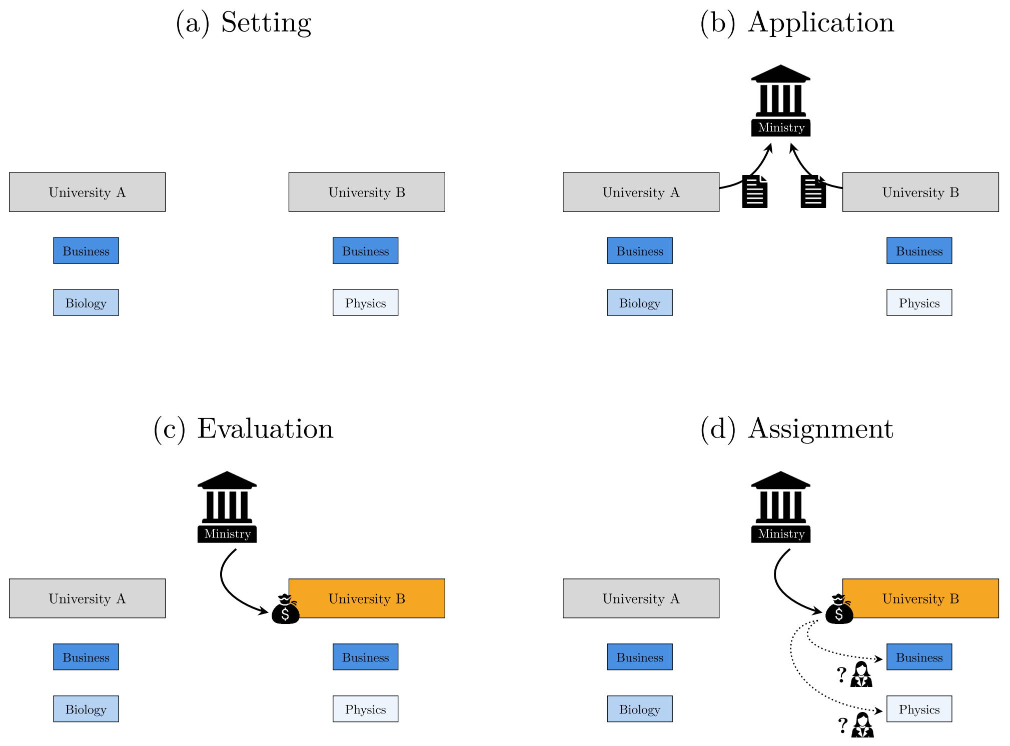
In the second stage, universities that receive a positive evaluation allocate subsidies to departments (Figure 1.1d). The assignment process is entirely within universities and not publicly documented. The only constraint is that the department must submit a funding plan outlining how the position will be permanently financed once the subsidy period has ended.

Once a department is selected to receive funding, it advertises the position with a note indicating that it is supported by the *Professorinnenprogramm*. While the announcement does not explicitly state that the position is to be filled by a woman, the reference to the program – whose purpose is widely understood – implicitly signals this intention. An example job posting is shown in Appendix Figure 1.B.4. This is followed by the job market process and the formal appointment procedure for full professors. Upon the successful appointment of a woman, the university submits a subsidy request to the ministry, specifying the required annual funding and duration. Requests are processed in chronological order until the program's budget is fully allocated.

Descriptives By 2024, the *Professorinnenprogramm* had subsidized 845 female professorships, with 63% originating from public universities. In the sample, I observed 429 subsidized female professorships at public universities. Panel A of Table 1.2 shows that in each wave, approximately 60 out of 83 public universities were eligible for subsidies. Figure 1.B.5c displays the university-specific eligibility status across waves and calls. Panel B of Table 1.2 reveals that each university appointed an average of 2.3 to 2.6 subsidized female professors per university, with annual subsidies ranging from 133,000 Euros in the first wave to 155,000 Euros in the third wave over an average duration of 4.7 years. The reason not every university maximizes the number of subsidized positions is due to the program being oversubscribed. For example, consider the first call of the first wave. Panel A of Table 1.3 shows that if each eligible university were to utilize all three possible subsidized appointments, this would result in 222 appointments. However, as calculated in Panel C of Table 1.3 given an available budget of 105 million Euros, only a maximum of 140 appointments are feasible. Consequently, by design, not all universities are able to subsidize three positions.

The average age of subsidized appointments is 42 years, indicating that the subsidy covers about 7% of all professor-related expenses until retirement, assuming a

Figure 1.1: Application Process



Note: The figure provides a schematic overview of the *Professorinnenprogramm*'s application process. The example assumes two universities, A and B, each with two departments (Figure 1.1a). Universities submit a 15-page equality concept to the German Ministry of Education, detailing current female representation and outlining measures to improve gender equality in academic positions (Figure 1.1b). In the example, both universities submit an application to the ministry. A review panel evaluates the submissions. Successful universities can utilize funding for up to three female full professorships (Figure 1.1c). In the example, university B is deemed eligible, while university A is not. Then, universities internally allocate subsidies to departments (Figure 1.1d).

retirement age of 67. Less than one-third of these appointments are 'early', meaning that a regular budgeted full professorship is not yet available. For these early appointments, a guaranteed transition to a regularly budgeted full professorship must be ensured by the end of the subsidy period.

Appendix Figure 1.B.6 displays the distribution of subsidized appointments across faculties and years. About half of all appointments are in the social sciences and humanities, which already had a relatively high share of female professors before the program. One-third of the appointments are in the natural sciences and engineering, fields with low shares of female professors. Appendix Figure 1.B.7 shows that more than half of call-specific budgets are exhausted within the first six months.

Table 1.2: Appointment Characteristics by Wave and University Type

	Public Universities			Other Universities		
	Wave I	Wave II	Wave III	Wave I	Wave II	Wave III
Panel A: University Characteristics						
Succesful Universities	54	59	57	58	64	70
Total Appointments	141	137	151	133	127	146
Appointments per University	2.52	2.32	2.60	2.38	1.98	2.13
	(0.69)	(0.71)	(0.88)	(0.95)	(0.85)	(0.88)
Panel B: Appointment Characteristics						
Share Regular Appointments	0.67	0.76	0.79	0.53	0.61	0.75
	(0.47)	(0.43)	(0.41)	(0.50)	(0.49)	(0.43)
Subsidy per Year (1,000 EURs)	133.80	136.67	155.52	101.38	109.09	119.66
	(23.58)	(19.39)	(20.14)	(34.24)	(30.52)	(34.03)
Subsidy Period (years)	4.70	4.69	4.88	4.46	4.45	4.91
	(0.72)	(0.73)	(0.45)	(1.09)	(1.05)	(0.45)
Job Search (months)	10.86	13.39	11.60	9.48	12.87	11.17
	(5.68)	(5.96)	(5.70)	(5.72)	(5.71)	(5.92)
Age at Appointment	42.24	42.27	43.13	.	.	.
	(4.86)	(5.96)	(6.08)	(.)	(.)	(.)
Subsidy Period / Work Life	0.07	0.07	0.07	.	.	.
	(0.01)	(0.01)	(0.01)	(.)	(.)	(.)

Note: The distinction between public and other universities is based on the list provided in Appendix Table 1.A.1. All statistics are calculated using data obtained from the Federal Government's funding portal by searching the term '*Professorinnenprogramm*' (Bundesregierung, 2023).

Table 1.3: Wave- and Call-specific Characteristics

	Wave I		Wave II		Wave III		Wave IV		
	Call 1	Call 2	Call 1	Call 2	Call 1	Call 2	Call 1	Call 2	Call 3
Panel A: Instituional Framework									
Budget (Million EURs)	105	45	90	60	130	70	320 Total		
Max. Appointments per University	3	3	3	3	3+1	3+1	3+1	3+1	3+1
Max. Subsidy Duration (years)	5	5	5	5	5	5	5	5	5
Max. Yearly Subsidy (1,000 EURs)	150	150	150	150	165	165	165	165	165
Min. Appointments within Budget	140	60	120	80	158	85	.	.	.
Panel B: Actual Appointments									
Succesful Universities	74	38	82	41	78	49	.	.	.
Appointments	184	90	176	88	176	121	.	.	.
Avg. Subsidy Duration (years)	4.5	4.8	4.6	4.6	4.9	4.9	.	.	.
Avg. Yearly Subsidy (1,000 EURs)	121.6	110.9	125.2	119.8	140.2	134.6	.	.	.
Panel C: World w/o Budget Cap									
Possible Appointments	222	114	246	123	234	147	.	.	.
Max. Required Budget (Million EURs)	166	85	184	92	193	121	.	.	.
Funding Gap (percent)	158	188	204	153	148	172	.	.	.

Note: The last row in Panel A assumes that each subsidized appointment utilizes both the maximum funding period and the maximum funding amount. The calculations in Panel C consider a scenario without budget constraints, where each eligible university subsidizes three female professors, using the maximum funding period and maximum funding amount. All statistics are calculated using data obtained from the Federal Government's funding portal by searching the term '*Professorinnenprogramm*' (Bundesregierung, 2023), along with wave-specific legal regulations (Bundesministerium für Bildung und Forschung, 2008, 2012, 2018, 2022).

1.3 Data

The analysis is based on an annual panel of 1,342 departments from 2003 to 2023, compiled from three primary data sources. First, I utilize administrative records on all academic personnel employed at public German universities. These data serve two main purposes: constructing a departmental panel to track hiring patterns for both junior and senior researchers and estimating departmental retirement probabilities, which are essential for my identification strategy. Second, I incorporate publicly available data on the *Professorinnenprogramm*. Finally, I use data from *OpenAlex* to gather information on scholarly output, which allows for the measurement of both publication quality and research topics at the departmental level.

1.3.1 Measuring Hiring Dynamics

The empirical analysis is primarily based on the *Hochschulpersonalstatistik*, accessed through the Federal Statistical Office of Germany (Destatis, 2018). The repeated cross-sectional data contain anonymized information on all academic personnel employed at public and applied German universities from 2003 to 2023. The *Hochschulpersonalstatistik* provides a wide range of demographic and professional details for each professor, including affiliation, department, pay grade, gender, nationality, year of Ph.D. completion, and whether the individual holds a position such as dean or university president. The list of all included variables, along with averages by gender, is presented in Appendix Table 1.A.3.

To prepare the data for the empirical analysis, I first create individual identifiers based on time-invariant characteristics to enable tracking of professors over time and across different universities. I aggregate the individual-level data into a panel of departments and years, which aligns with the level at which subsidies are assigned. Analogously, I construct department-level panels for post-docs, Ph.D. students, and research assistants, utilizing comparable information available for professors. These additional panels allow for the investigation of potential spillover effects to junior researchers.

1.3.2 Professorinnenprogramm

I corroborate the department rosters with application data on the *Professorinnenprogramm*, which can be retrieved through the Federal Government's funding portal using the search term '*Professorinnenprogramm*' (Bundesregierung, 2023). For each subsidized appointment, I collect the professor's name, institutional affiliation, associated department, type and date of appointment, yearly subsidy endowment, and funding duration. This information enables me to create variables for each department, indicating the start and end of the start-up funding periods.

1.3.3 Measuring Retirement Probabilities

To predict department-specific retirement probabilities, I use a logistic Lasso estimator to identify the most influential predictors from a broad set of potentially relevant characteristics (Friedman, Hastie and Tibshirani, 2010). To prevent overfitting, the data is divided into an estimation sample and a prediction sample (Hansen, 2022). The penalized log-likelihood is estimated using only data from before the first funding period, i.e., for $t < 2008$:

$$\hat{\boldsymbol{\rho}} = \underset{\boldsymbol{\rho}\in\mathbb{R}^k}{\arg\max} \sum_{l} \sum_{t<2008} \left\{ \text{Retire}_{lt}^{(5)} \cdot f(\boldsymbol{X}_{lt}) - g(f(\boldsymbol{X}_{lt})) \right\} - \lambda ||\boldsymbol{\rho}||_1 \qquad \text{(LASSO)}$$
$$\text{where } g(\xi) = \log\left(1 + \exp(\xi)\right).$$

$\text{Retire}_{lt}^{(5)}$ is a binary variable set to one if professor l retires within five years of year t, i.e., $\text{Retire}_{lt}^{(5)} \equiv 1$ if $\text{Retire}_{lt+\tau} = 1$ for any $\tau \in \{0, 1, ..., 4\}$. This modeling choice is intended to reflect the requirement that a permanent position must be guaranteed only at the end of the subsidy period, which spans a maximum of five years. The vector $\boldsymbol{X}_{lt}$ includes individual characteristics of professor l that may predict retirement. In particular, the vector contains variables measuring gender, race, academic field, years since appointment as a full professor (modeled up to a cubic polynomial), state of employment, remuneration bracket, and the number of male colleagues. To account for potential non-linearity, the function $f(\cdot)$ also includes first-order interactions between all these variables. In total, the model contains 214 potential predictors. The function $g(\cdot)$ implements the logistic Lasso, λ describes the penalization parameter. The value of λ is calibrated using ten-fold cross-validation, selecting the value minimizing the mean-squared prediction error (Chen and Lee, 2021).

After training the model on pre-2008 data, it is evaluated on the post-2008 data to forecast individual-level retirement probabilities.[13] The probability of any retirement occurring in department i within the next five years from t is computed as the complement of observing no individual retirements:

$$\text{Retire}_{it}^{(5)} \equiv 1 - \prod_{l \in \mathcal{L}} \left(1 - \widehat{\text{Retire}}_{lt}^{(5)}\right) \quad \forall t \geq 2008.$$

The set $\mathcal{L}$ represents all full professors employed in department i in year t, i.e., $\mathcal{L} \equiv \mathcal{L}(i,t) = \{l : l \in i \text{ in } t\}$.

In Section 1.5.1, I validate my findings using binary retirement indicators based on age thresholds, which produce estimates of comparable magnitude but lower statistical significance. In Appendix Section 1.D.1, I show that this pattern arises because binary measures discard substantial variation in retirement timing – variation my continuous probability metric captures.

1.3.4 Measuring Research Output

To analyze potential changes in research patterns, I collect bibliographical information on all research produced at German public universities through *OpenAlex* (Priem, Piwowar and Orr, 2022). *OpenAlex* serves as a scholarly catalog encompassing the world's academic papers, researchers, journals, and institutions. It succeeded the Microsoft Academic Graph, which was discontinued in May 2021. *OpenAlex* regularly expands its database by aggregating and standardizing data from various sources, including ORCID, Pubmed, arXiv, and Crossref. For each paper, the data include the complete list of authors (including their affiliation at the time of publication), the journal of publication, the references cited by the paper, and the citations it has received. The analysis is limited to research output from researchers who have been affiliated with a public German university at some point, identified through *OpenAlex*'s institution identifier. Research output is measured by the quantity and quality of publications and the number of citations received. I assess the quality of publications using journal impact factors provided in Scopus (2023).

While the research data include institutional information, they lack details on the department and position of researchers. To match research output to the depart-

[13] Due to the focus on forecasting, no inference is made on the model parameters, which eliminates the need for a double selection estimator (Belloni, Chernozhukov and Hansen, 2014).

ment rosters, I utilize complementary information from the *Hochschullehrerverzeichnis*, an annual directory listing German university professors along with their affiliations and descriptions of their disciplines (Hochschulverband, 2002–2022).[14] The matching procedure is described in Appendix Section 1.D.2.

1.4 Empirical Strategy

Identification Strategy To causally identify how affirmative action appointments affect departments, I implement an instrumented difference-in-differences design. My main argument is that the *Professorinnenprogramm*'s requirement to eventually convert subsidized appointments into permanent positions makes departments in eligible universities with high retirement probabilities during the funding period more likely beneficiaries of the program, compared to those with low retirement probabilities. Retirement probabilities are primarily determined by historical hiring decisions, making them predetermined and not subject to manipulation. Additionally, departments cannot create vacancies by dismissing tenured professors, as they hold lifetime civil service positions. Similarly, departments cannot accommodate subsidized appointments through the creation of new permanent positions, as these are contingent on negotiations with the federal state and typically limited to addressing increased teaching demand. These institutional constraints ensure that retirement probabilities serve as a plausible source of exogenous variation.

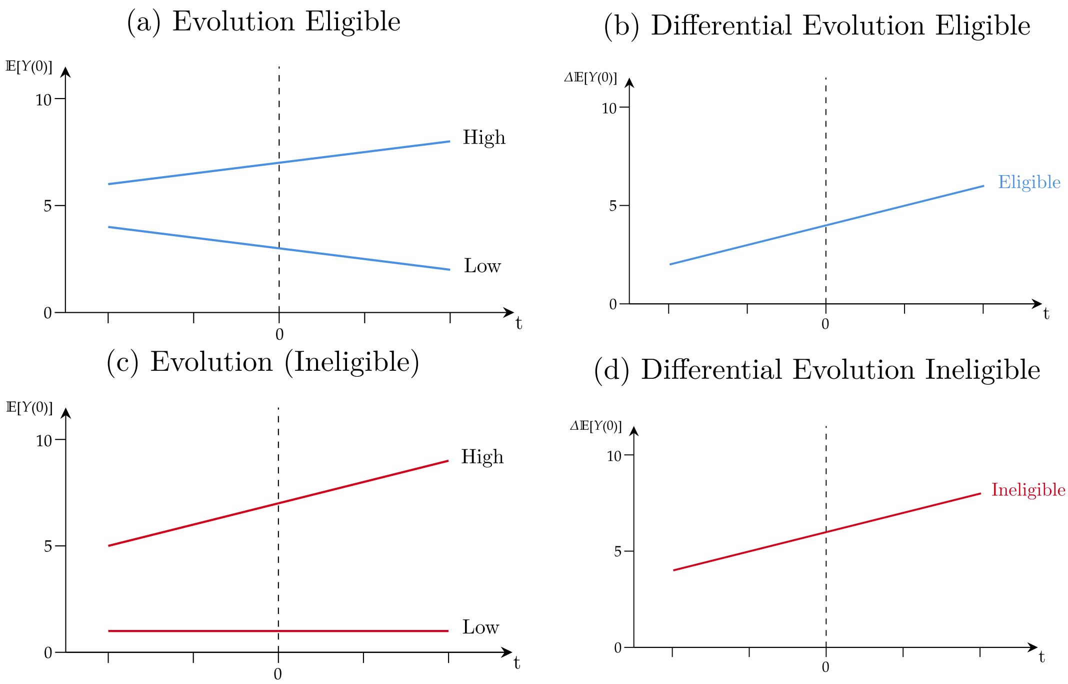
I extend the design by also considering retirement probabilities of departments in ineligible universities – those that did not pass the first-stage application process. Incorporating this additional cross-sectional variation allows identification under considerably weaker assumptions, as it allows for potential retirement-specific trends in potential outcomes.[15]

[14] Another approach would involve inferring the department identifier from the research output of a researcher. However, this method is complicated by the fact that the academic work of researchers does not always align with their department. For instance, an economist might be employed in a business department because the university lacks a dedicated economics department. Moreover, the position of the researcher cannot be identified at all from the *OpenAlex* data.

[15] Retirement-specific trends may arise if retirement probabilities exhibit autocorrelation even after controlling for covariates. In such cases, departments with high retirement probabilities will, on average, be older, while those with low retirement probabilities will tend to be younger. Since the potential outcomes of older and younger departments are unlikely to evolve in the same way, this would violate the parallel trends assumption.

Figure 1.2 illustrates this logic. Figures 1.2a and 1.2c show the trajectories of some arbitrary potential outcome $Y(0)$ in departments with high and low retirement probabilities across eligible and ineligible universities. When excluding ineligible universities, identification would require parallel trends in Figure 1.2a, which effectively rules out that departments with differing retirement probabilities follow different trends in potential outcomes. In contrast, including ineligible universities allows for a more flexible identification strategy, as it suffices to assume that the differential evolution of average untreated outcomes between high- and low-retirement departments in eligible and ineligible universities – illustrated in Figures 1.2b and 1.2d – move in parallel. This assumption seems more plausible.

Figure 1.2: Exclusion Restriction – Parallel Trend Assumption



Note: The figure provides a schematic overview of the parallel trend assumption in the triple difference-in-differences framework described in Section 1.4. Figure 1.2a illustrates an exemplary average residualized potential outcome evolution for departments with high and low retirement probabilities within eligible universities. Figure 1.2b depicts the average differential evolution of potential outcomes between these departments. Analogously, Figures 1.2c and 1.2d show the average evolution of residualized potential outcomes and their differential evolution within ineligible universities. The triple difference estimator assumption requires that the differential evolution in Figures 1.2b and 1.2d follow a parallel trend. In a standard difference-in-differences setting – where the analysis is limited to departments in eligible universities – one would need to assume that the average potential outcomes for departments with high and low retirement probabilities in Figure 1.2a evolve in parallel, which would rule out retirement-specific trends.

Empirical Model I formalize the triple-difference identification strategy using the following regression framework,

$$
\begin{aligned}
\text{Female Hiring}_{it} = {} & \alpha_i + \alpha_{u(i)t} + \alpha_{f(i)t} \\
& + \phi_1 \text{Retire}_{ig}^{(5)} \cdot \text{Post}_{tg} \\
& + \psi_1 \text{Retire}_{ig}^{(5)} \cdot \text{Post}_{tg}^{1st} \cdot \text{Eligible}_{u(i)g} \\
& + \psi_2 \text{Retire}_{ig}^{(5)} \cdot \text{Post}_{tg}^{2nd} \cdot \text{Eligible}_{u(i)g} \\
& + \varepsilon_{it}
\end{aligned}
\tag{IV1}
$$

where I initially restrict the analysis to the initial funding period, $g = 1$, starting in $\tau(g) = 2008$. In Section 1.4, I extend the framework to include all funding periods, addressing issues related to staggered treatment adoption and multiple treatment assignments.

In Equation (IV1), Female Hiring$_{it}$ equals one if department i appoints a female full professor in year t. Retire$_{ig}^{(5)}$ represents the probability that department i experiences any retirement in $t \in [\tau(g), \tau(g) + 5]$. Eligible$_{u(i)g}$ indicates whether university $u(i)$ is eligible for funding in period g.

Post$_{tg}$ is equal to one in all post-funding years, i.e., $t \geq \tau(g) = 2008$.[16] This period can be divided into two distinct phases. The first is the funding period, represented by the indicator Post$_{tg}^{1st}$, which covers the years when funding is available – typically two years.[17] The second is the post-funding phase, indicated by Post$_{tg}^{2nd}$, encompassing all years after the funding has been exhausted. According to the instrument's logic, retirements in eligible universities should predict female hiring only during the funding phase, with no effect in the post-funding phase.

The parameter of interest, ψ_1, captures the differential effect of retirement on female hiring between eligible and ineligible universities in the funding period. The coefficients ψ_2 – the effect of retirement at eligible universities after the funding period – and ϕ_1 – the effect of retirement at ineligible universities – can

[16] Note that, $\tau(g)$ does not always align with the actual year of a subsidized appointment. While this might introduce some noise, it guarantees the temporal alignment of pre- and post-treatment years within a funding period. Otherwise, when expanding the research design to incorporate all funding periods, I would observe multiple pre- and post-treatment years within each of the six funding periods. In the first funding period, I assume $\tau(g) = 2008$, even though some of subsidized appointments were made in 2009 or 2010. In the dynamic version of Equation (IV1), this will lead to a downward bias of the first two lead estimates, because the specification assumes some departments as treated, although the appointment will only happen within the next two years.

[17] Appendix Figure 1.B.7 illustrates the share of funds utilized over time for each funding period.

be seen as a placebo test for the identification strategy, as no effect is expected in either the post-funding period or in ineligible universities.[18].

The specification extensively controls for possible unobserved factors affecting female hiring and the interacted instrument. In particular, α_i captures time-invariant department-specific factors like homophily preference of department i. Further, $\alpha_{u(i)t}$ captures time-varying university-specific factors at university $u(i)$ in t. Partialling out these effects transforms absolute into relative retirement probabilities, which ensures that the instrument has no predictive power in universities with homogeneous retirement probability distributions.[19] Lastly, $\alpha_{f(i)t}$ captures time-varying field-specific factors affecting all fields $f(i)$ in t, such as a large birth cohort of economists retiring in t leading to a surge in the demand for female economists. All other unobserved factors enter the error term ε_{it}, which is clustered at the department level (Abadie et al., 2023).

In a second step, I then use predicted female hiring from Equation (IV1) to estimate the following two-stage least squares regression,

$$Y_{it+h} = \alpha_i + \alpha_{u(i)t} + \alpha_{f(i)t} + \beta\widehat{\text{Female Hiring}}_{it} + u_{it} \tag{IV2}$$

where Y describes some outcome of department i in year $t + h$. For all other variables, the previous explanations apply.

Identifying Assumption Causal identification of β in Equation (IV2) relies on the assumption that in the absence of the *Professorinnenprogramm*, the difference of the average outcome among departments with high and low retirement probabilities in eligible universities evolves in the same way as the difference of the average outcome among units with high and low retirement probabilities in ineligible universities. This implies that conditional on covariates, retirement probabilities in eligible universities are orthogonal to u it in Equation (IV2).

Staggered Adoption and Multiple Treatment In case of a single funding period, Specifications (IV1) and (IV2) allow one to causally identify the effect of hiring a female professor. The presence of multiple funding periods – six in total – necessitates adjusting the estimation procedure. While considering all funding

[18] All other combinations of Retire_{ig}, $\text{Eligible}_{u(i)g}$, and Post_{tg} are omitted due to collinearity with the fixed effects included in the model.

[19] Specifically, a department with a high likelihood of retirement should not be more likely to receive the subsidy if other departments at the same university have similarly high retirement probabilities.

periods allows us to leverage additional temporal variation, as departments receive treatment at different points in time, it also introduces issues of cross-lag contamination and multiple treatment assignments.

When employing a standard fixed-effects estimator with staggered treatment assignment, units that have already been treated serve as comparison units for units that have not yet been treated. This can introduce bias when there is treatment effect heterogeneity (Roth et al., 2023). Additionally, departments may receive subsidies in multiple funding periods, resulting in several treatments. In such cases, fixed-effects estimators are not robust to heterogeneous effects and may be contaminated by the effects of other treatments (de Chaisemartin and D'Haultfœuille, 2023).

To address both issues, I implement a stacked regression design, following Cengiz et al. (2019) and Dube et al. (2023). This approach constructs a separate panel for each funding period g, including all first-time subsidy-receiving departments and *clean* non-subsidy-receiving departments. The stacked regression design aggregates estimates from these funding period-specific panels. Identification holds as long as the parallel trends assumption is valid in each panel.

The stacked panel is constructed as follows: (1) Consider each funding period as a separate event g beginning in year $\tau(g)$. (2) For each event g, fix departmental retirement probabilities and the university eligibility status in $\tau(g)$. (3) For each event g, define an event window $T(g) \equiv [\underline{\tau}(g), \overline{\tau}(g)]$ where $\underline{\tau}(g) \equiv \max\{2003, \tau(g) - c\}$ and $\overline{\tau}(g) \equiv \min\{\tau(g) + c, 2023\}$. The event window is defined by the tuning parameter, c, which is set to $c = 5$.[20] (4) For each funding period g, define an exclusion set containing observations for which $t \notin T(g)$ and departments that previously received funding in some funding period g' for which $\tau(g') < \tau(g)$.

Under these restrictions, each event-specific panel only includes departments receiving subsidies for the first time and *clean* departments that have not received subsidies in any previous subsidy period. Stacking all datasets from different funding periods and interacting all fixed effects with event indicators allows to consistently estimate treatment effects via standard fixed effects estimators, avoiding biases from multiple treatments or staggered adoption (Dube et al., 2023;

[20] Increasing c allows the evaluation of treatment effects over a larger time horizon but also increases the event-specific exclusion sets. In my setting, opting for $c = 5$ allows to balance both effects. In addition, the sample covers the periods from 2005 to 2023; by construction the same boundaries apply to each event-specific panel. Table 1.4 provides an overview of all event-specific panel endpoints.

Table 1.4: Panel Construction by Wave and Call

Wave	Call	g	$\tau(g)$	$\underline{\tau}(g)$	$\overline{\tau}(g)$
1	1	1	2008	2003	2014
1	2	2	2010	2005	2016
2	1	3	2013	2007	2019
2	2	4	2015	2009	2021
3	1	5	2018	2012	2023
3	2	6	2020	2014	2023

Note: The table displays how each distinct combination of funding wave and call corresponds to a funding period, denoted sequentially by $g \in G \equiv \{1, 2, 3, 4, 5, 6\}$. The last two columns define the endpoints of the event windows specific to each funding period, denoted as $T(g) \equiv [\underline{\tau}(g), \overline{\tau}(g)]$.

Wing, Freedman and Hollingsworth, 2024). For inference, standard errors are clustered by department i and event g.

1.5 Results

I present four key results. First, I validate the identification strategy by showing that retirement probabilities are a strong predictor of female hiring in eligible universities. Second, I evaluate the impact of appointing an additional female professor on future hiring at both senior and junior levels. Third, I examine how the presence of a female professor influences collaboration patterns and research output within the department. Finally, I quantify the program's effectiveness by estimating the extent to which subsidized appointments result in the hiring of female professors who would not have been appointed otherwise.

1.5.1 Effects on Hiring

First Stage To establish a causal link between the appointment of female professors and departmental outcomes, I rely on the identification strategy outlined in Section 1.4. I first demonstrate that retirement probabilities in eligible universities during the funding period are strong predictors of female hiring. Table 1.5 presents static estimates of Equation (IV1). The most rigorous specification, shown in Column (4) indicates that among departments at eligible universities, a 10 percentage-point increase in the probability of experiencing any retirement[21] in the funding period is associated with a 4.7 percentage-point higher likelihood of appointing a woman as a full professor compared to departments at ineligible universities, beyond their pre-existing hiring differences. This estimate is highly significant, with an F-statistic exceeding 24. In contrast, estimates for retirement probabilities in ineligible universities, as well as in eligible universities during the post-funding period, are close to zero, suggesting that hiring patterns remain unchanged relative to the pre-funding period.

Complementing these static estimates, Figure 1.3 presents dynamic event study results. The estimates suggest that departments with different levels of retirement are not on diverging outcome trajectories before the funding period. For departments in eligible universities, the pre-trend coefficients remain stable and close to zero. Once the funding period begins, there is a rapid and substantial increase in the share of appointed female professors in departments with high retirement probabilities in eligible universities, relative to comparable departments in ineligible universities and the pre-period. This effect diminishes after the funding period

[21] Within five years of the funding period's start.

Table 1.5: First Stage Estimates

	Dependent Variable: Any Woman Getting Tenure			
	(1)	(2)	(3)	(4)
Retire · Post	0.041 (0.082)	0.054 (0.085)	0.042 (0.082)	0.039 (0.079)
Retire · Eligible · Post^{1st}	0.528*** (0.083)	0.441*** (0.084)	0.495*** (0.089)	0.471*** (0.096)
Retire · Eligible · Post^{2nd}	0.043 (0.086)	0.057 (0.089)	0.044 (0.086)	0.037 (0.097)
Observations	147,591	147,591	147,591	147,591
F-statistic	40.47	27.56	30.94	24.07
Fixed Effects				
Department	-	✓	✓	✓
Field × Year	-	-	✓	✓
University × Year	-	-	-	✓

Note: The table presents regression results from estimating Equation (IV1) across multiple specifications. All regressions are estimated using a combined dataset constructed by stacking funding period-specific panels (as detailed in Section 1.4) and interacting all fixed effects with funding period indicators. The F-statistic reflects the test results for the null hypothesis that the coefficient ψ_1 equals zero. Robust standard errors, clustered by department and funding period, are reported in parentheses. Significance levels are indicated as follows: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

ends – after approximately two years[22] – with estimates reverting to pre-funding levels. This reversion suggests that subsidized appointments do not influence subsequent hiring of female full professors, indicating that the program neither promotes nor hinders the advancement of other women to full professorships.

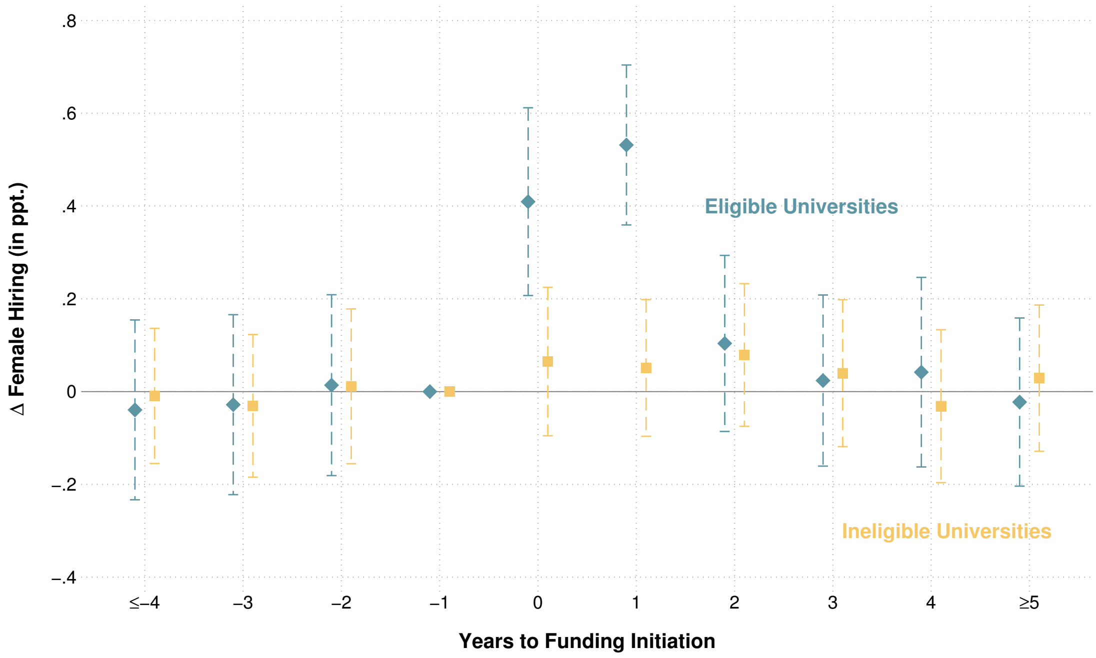
Design Validity The dynamic effects of retirement probabilities in ineligible universities confirm the validity of the identification strategy. In the pre-period, all estimates are tightly clustered around zero, supporting the parallel trends assumption.[23]

I conduct three additional exercises to validate my identification strategy. First, I re-estimate the first-stage equation by randomly reassigning eligibility status among universities and retirement probabilities among departments for each funding period. This exercise allows a comparison of the actual realization of

[22] For a detailed overview of the length of funding periods see Table 1.1. Note that available funds can be exhausted before the end of the funding period. Appendix Figure 1.B.7 shows the share of funds utilized over time for each funding period.

[23] While there is a slight increase in point estimates during the post-period, these remain statistically indistinguishable from zero. This marginal uptick may reflect universities making female appointments in anticipation of a favorable first-round evaluation, which ultimately does not materialize.

Figure 1.3: First Stage – Dynamic Effect by Eligibility



Note: The figure presents first-stage event study estimates based on the regression framework outlined in Equation (IV1). All post-indicators in Equation (IV1) are replaced by a set of indicators that represent the relative time in years from the start of the funding period, as shown on the horizontal axis. Standard errors are clustered by department and funding period. Bars represent 95% confidence intervals.

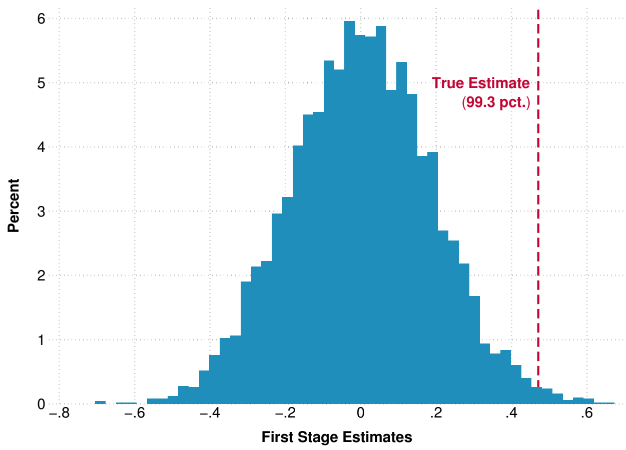
eligibility status and retirement probabilities with hypothetical scenarios that did not realize. As shown in Figure 1.4, the actual correlation is a clear outlier within the nearly normal distribution of placebo estimates.

Second, Columns (1)–(3) of Table 1.6 demonstrate that my estimates are insensitive to how retirement probabilities are incorporated into the instrument in Equation (IV1). Column (1) replicates the main specification, where retirement probabilities are fixed at their initial value within each funding period, aligning with a standard DiD framework where temporal variation is only introduced through a pre-post comparison. Alternatively, retirement probabilities could be recalculated annually and used as a time-varying measure in the interaction term. Column (2) shows that this approach slightly reduces the instrument's statistical power but otherwise does not affect the results. Column (3) further controls for time-varying retirement probabilities as a covariate, yielding similar estimates.[24]

[24] A limitation of incorporating time-varying retirement probabilities, whether as an instrument or a control variable, relates to the issue of "bad controls". If retirement probabilities in the post-period are influenced by those in the funding period – which is mechanically true in the presence of autocorrelation – conditioning on them in any form can introduce meaningful bias (Baker et al., 2025).

Third, Columns (4)–(6) of Table 1.6 show that the first-stage correlation remains stable when measuring retirement with a binary indicator based on different age thresholds. Throughout, the instrument remains sufficiently strong, although estimates from the binary measure are smaller in magnitude[25] and less statistically significant than those from the continuous approach. This pattern reflects that the binary indicators discard variation in retirement timing that the continuous measure is able to capture, as discussed in Appendix Section 1.D.1.

Figure 1.4: First Stage Placebo Estimates



Note: The figure presents the distribution of 5,000 first-stage placebo estimates, which were obtained by randomly reassigning eligibility status among universities and retirement probabilities among departments within each funding period. These estimates represent the regression results for the parameter ψ_1 from Equation (IV1). This parameter captures the differential impact of retirement on female hiring between eligible and ineligible universities during the funding period. The vertical dashed line marks the actual first-stage estimate, which lies at the 99.3rd percentile of the placebo distribution. Standard errors are clustered by department and funding period.

[25] The reduction in magnitude is partly mechanical. Specifically, in a model with a constant, an indicator variable, and a dependent variable bounded between zero and one, the OLS estimator is given by $\hat{\beta}_1(x \in \{0,1\}, y \in [0,1]) = \bar{y}_1 - \bar{y}_0$, which is constrained within the interval $[-1,1]$. Conversely, when the indicator variable is replaced with a continuous measure bounded between zero and one, it follows that $\hat{\beta}_1(x \in [0,1], y \in [0,1]) \in \mathbb{R}$.

Table 1.6: First Stage Estimates – Validity

	Dependent Variable: Any Woman Getting Tenure					
	Retirement Probabilities			Retirement Age Threshold		
	Fixed	Varying	Varying	≥ 65	≥ 66	≥ 67
	(1)	(2)	(3)	(4)	(5)	(6)
Retire · Post	0.039	0.052	0.055	0.038	0.032	0.031
	(0.079)	(0.085)	(0.082)	(0.052)	(0.053)	(0.052)
Retire · Eligible · Post^{1st}	0.471***	0.429***	0.428***	0.225***	0.207***	0.195***
	(0.096)	(0.098)	(0.093)	(0.057)	(0.057)	(0.058)
Retire · Eligible · Post^{2nd}	0.037	0.055	0.058	0.040	0.034	0.033
	(0.097)	(0.100)	(0.097)	(0.055)	(0.056)	(0.055)
Observations	147,591	147,591	147,591	147,591	147,591	147,591
F-statistic	24.07	19.16	21.18	15.58	13.19	11.77
Fixed Effects						
Department	✓	✓	✓	✓	✓	✓
Field × Year	✓	✓	✓	✓	✓	✓
University × Year	✓	✓	✓	✓	✓	✓
Controls						
Retirement	-	-	✓	-	-	-

Note: The table presents regression results from estimating Equation (IV1) using different approaches to measure departmental retirement. Columns (1)–(3) employ a continuous retirement measure based on the logistic LASSO estimator described in Section 1.3.3. In Column (1), retirement probabilities are computed at the start of the funding period and remain fixed throughout the corresponding panel. Column (2) allows for time-varying retirement probabilities, recalculating them annually for the subsequent five years. Column (3) extends this specification by additionally controlling for time-varying retirement probabilities. Columns (4)–(6) replace retirement probabilities with a binary indicator based on the retirement age thresholds specified in the column headers. All regressions are estimated using a combined dataset constructed by stacking funding period-specific panels (as detailed in Section 1.4) and interacting all fixed effects with funding period indicators. The F-statistic tests the null hypothesis that the coefficient on the triple interaction term is equal to zero. Robust standard errors, clustered by department and funding period, are reported in parentheses. Significance levels are indicated as follows: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Junior Faculty and Ph.D. Hiring Next, I evaluate potential trickle-down effects by analyzing changes in the hiring of women among assistant professors, post-docs, and Ph.D. students. Figure 1.5b shows that the number of female assistant professors remains unchanged following the appointment of a female full professor. Similarly, Figure 1.5c reveals a slight increase in the proportion of female post-docs three years post-funding, though this effect is statistically insignificant.

In contrast, Figure 1.5a highlights a significant increase in the hiring of female Ph.D. students. Again, the figure displays stable pre-trends. Three years after the funding period ends, there is a significant uptake in the hiring of female Ph.D. students. Relative to the pre-funding period and departments in ineligible universities, experiencing a certain retirement within the next five years is associated with a 4.6 percentage-point rise in female Ph.D. recruitment. This effect is almost entirely driven by women pursuing their Ph.D. at their home institution – defined as those who completed their undergraduate studies in the same department – with a rise of 3.9 percentage points.[26]

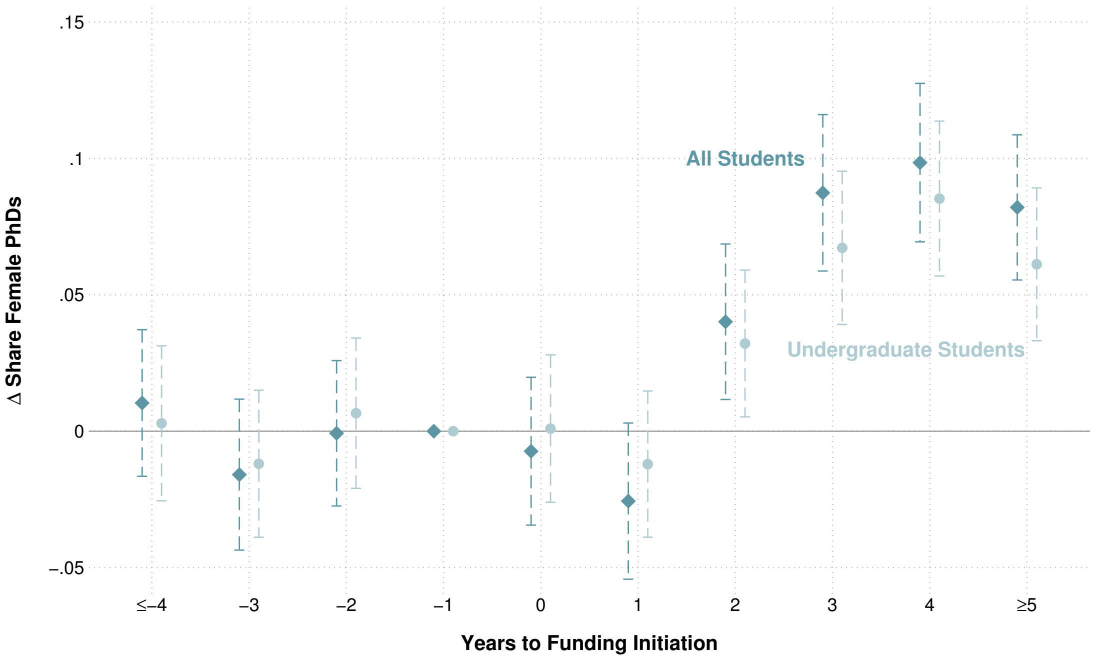
Corresponding 2SLS estimates, combining the first-stage and reduced-form results, are displayed in Panel B of Table 1.7. Each subsidized female professor appointment leads to a 9.8 percentage-point increase (+19%) in the overall share of female Ph.D. students and a 8.3 percentage-point increase (+29%) among those studying at their home institution. I discuss robustness checks addressing weak-instrument concerns in Appendix Section 1.C.2.

Mechanism The observed hiring shift could stem from shifts in preferences among students, professors, or both. For instance, female students may be more likely to pursue PhDs at their home institution after exposure to newly appointed female professors – a pattern consistent with existing evidence on how female role models influence career trajectories (Porter and Serra, 2020; Blau et al., 2010; Ginther et al., 2020). Alternatively, current professors might prioritize recruiting female PhD students in response to the appointment of a female full professor. However, this explanation is less compelling given the absence of significant effects on the hiring of female assistant professors and post-docs. The lack of impact on these groups – who are typically affiliated with other institutions and thus unable to interact with the subsidized professor – further supports the role model hypothesis. Overall, the evidence strongly suggests that role model effects drive these changes.

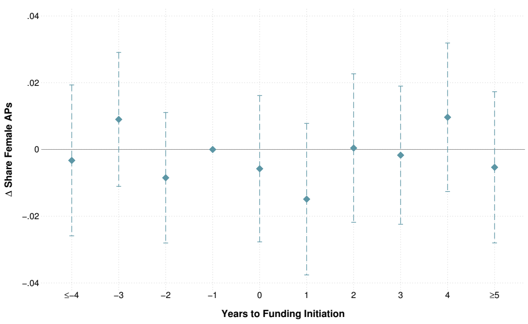
[26] The *Hochschulpersonalstatistik* data allows one to infer the 'home institution' of Ph.D. students through a variable indicating the highest degree awarding institution. From this, I construct the share of female Ph.D. students pursuing their doctoral studies at their 'home institution'.

Figure 1.5: Reduced Form – Effects on Junior Female Hiring

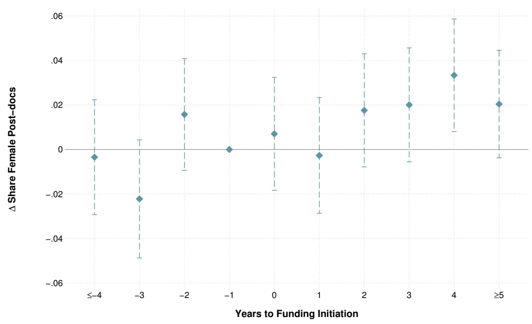
(a) Ph.D. Students



(b) Assistant Professors



(c) Post-docs



Note: Each figure presents reduced-form event study estimates based on the regression framework outlined in Section 1.4. The outcome variable in each figure is the share of appointed women within the group specified in the caption. The post-indicator is replaced by a set of indicators that represent the relative time in years from the start of the funding period, as shown on the horizontal axis. Standard errors are clustered by department and funding period. Bars represent 95% confidence intervals.

Table 1.7: Change in Hiring Patterns

	Junior Faculty		Ph.D. Students	
	Ass. Professor	Post-Doc	Overall	Home
	(1)	(2)	(3)	(4)
Panel A: Reduced Form				
Retire · Post	0.017	0.009	0.004	-0.013
	(0.014)	(0.012)	(0.012)	(0.016)
Retire · Eligible · Post	0.016	-0.003	0.046***	0.039***
	(0.013)	(0.011)	(0.014)	(0.014)
Observations	147,591	147,591	147,591	147,591
Panel B: 2SLS				
Female Hiring	0.034	-0.006	0.098***	0.083***
	(0.121)	(0.109)	(0.028)	(0.029)
Observations	147,591	147,591	147,591	147,591
F-statistic	24.07	24.07	24.07	24.07
Panel C: OLS				
Female Hiring	0.124**	0.190**	0.401***	0.359***
	(0.049)	(0.082)	(0.108)	(0.083)
Observations	147,591	147,591	147,591	147,591
Fixed Effects				
Department	✓	✓	✓	✓
Field × Year	✓	✓	✓	✓
University × Year	✓	✓	✓	✓

Note: The table presents regression results from estimating Equation (IV1) across multiple specifications. All regressions are estimated using a combined dataset constructed by stacking funding period-specific panels (as detailed in Section 1.4) and interacting all fixed effects with funding period indicators. Robust standard errors, clustered by department and funding period, are reported in parentheses. Significance levels are indicated as follows: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

1.5.2 Effects on Collaboration Patterns

Existing research suggests that exposure to underrepresented groups can reduce stereotypes and increase future engagement with those groups (Carrell, Hoekstra and West, 2015). I hypothesize that if negative gender stereotypes exist, changes in gender attitudes may manifest in co-authorship patterns. Specifically, I examine whether exposure to a female professor increases the share of female co-authors. My data allows for dynamic tracking of co-authorship networks over time, enabling a detailed analysis of these patterns.

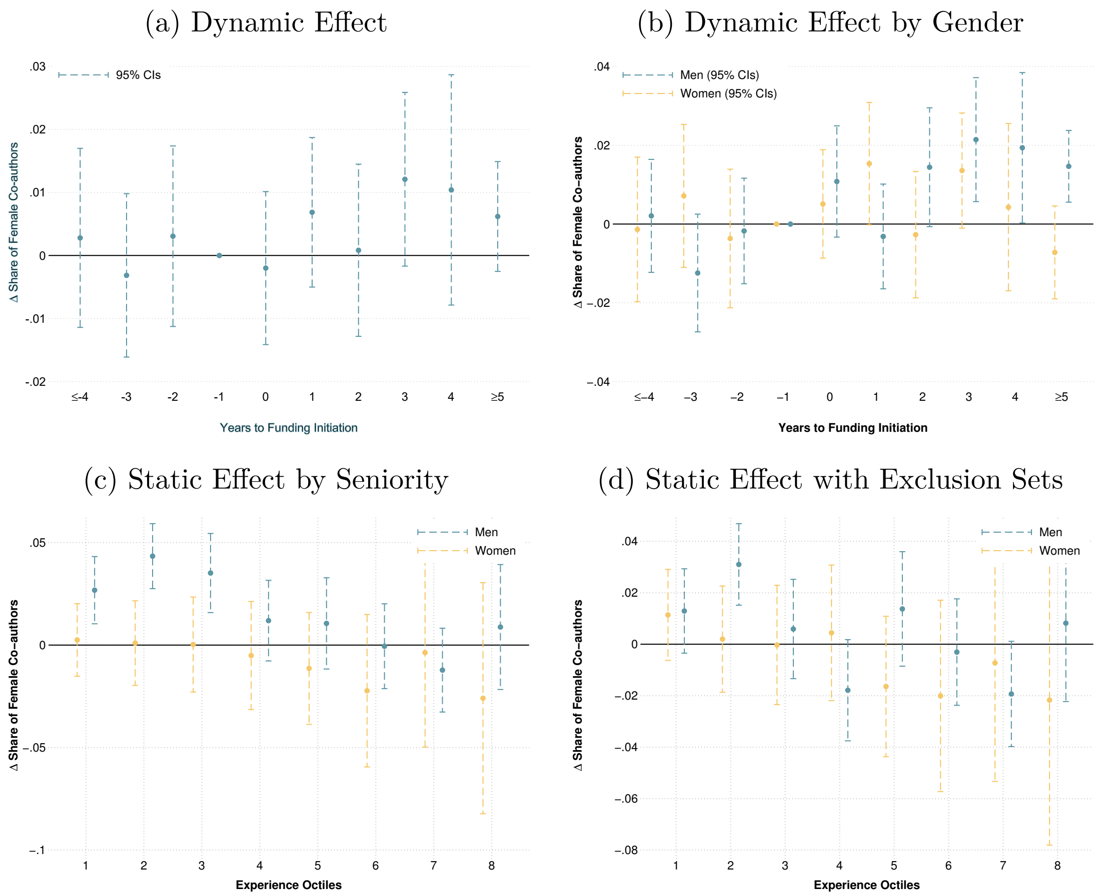
Reduced Form To track changes in collaboration patterns, I use the share of female co-authors as the outcome variable[27] in my empirical design. Figure 1.6a displays how the average share of female co-authors among all department members changes after a woman joins the department. The event study reveals no significant overall increase in female co-authorship.

However, disaggregating these effects by gender in Figure 1.6b shows that while women's co-authorship patterns remain unchanged, men exhibit a slight increase in female co-authorship after being exposed to an additional female professor. This shift emerges two to three years post-exposure, likely reflecting the time required for author matching and publication lags.

Figure 1.6c breaks down these results by seniority, showing that the effect is most pronounced among junior men – defined as those with below-median experience. This aligns with evidence from outside academia suggesting that stereotype malleability declines with age (Gonsalkorale, Sherman and Klauer, 2009; Siyanova-Chanturia et al., 2015). The estimates indicate that a 10 percentage-point increase in retirement probabilities among departments in eligible universities – relative to departments in ineligible universities and the pre-funding period – raises the likelihood of junior men co-authoring with women by 2.9 percentage-points (average of Columns (4) and (5) in Panel A of Table 1.8). The corresponding 2SLS estimates in Panel B of Table 1.8 indicate that hiring a female professor results in a 6.3 percentage-point (average of Columns (4) and (5) in Panel B of Table 1.8) increase in female co-authorships among junior men, representing a 24% rise relative to the pre-funding period average.

[27] To avoid the possibility that results are driven by changes in department composition following a subsidized appointment, I fix the department composition in each funding period-specific dataset in $\tau(g)$.

Figure 1.6: Reduced Form – Effects on Collaboration Patterns



Note: The figure presents reduced-form event study estimates based on the regression framework outlined in Section 1.4, using the share of female co-authors as outcome variable. The post-indicator is replaced by a set of indicators that represent the relative time in years from the start of the funding period, as shown on the horizontal axis. Figures 1.6c and 1.6d restrict the sample by tenure length octiles, as indicated on the horizontal axis. In Figure 1.6d, the share of female co-authors is measured excluding the peer network of the subsidized appointment, as detailed in Section 1.5.2. Standard errors are clustered by department and funding period. Bars represent 95% confidence intervals.

Accounting for Peer Network Instead of reflecting shifts in gender attitudes, the increase in female co-authors could result from faculty members gaining access to the peer networks of subsidized appointments, which are predominantly female (64%). To disentangle these channels, I model the potential peer networks of each affirmative action appointment and calculate co-author shares excluding these networks. Specifically, for all academic work published by professors in some department i, I exclude authors connected to the subsidized appointment joining department i.[28] I narrow the pool of co-authors along four dimensions.

[28] In departments without a subsidized appointment, the share of female co-authors remains unchanged.

Table 1.8: Change in Collaboration Patterns

	All	Women	Men	Men by Seniority (Quartiles)			
				Q1	Q2	Q3	Q4
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A: Reduced Form							
Retire · Post	-0.007	-0.001	-0.009	0.009	0.006	-0.002	-0.008
	(0.007)	(0.008)	(0.008)	(0.008)	(0.009)	(0.010)	(0.06)
Retire · Eligible · Post	0.010	0.005	0.013	0.035***	0.024***	0.005	-0.002
	(0.008)	(0.009)	(0.009)	(0.007)	(0.009)	(0.009)	(0.008)
Observations	147,591	147,591	147,591	147,591	147,591	147,591	147,591
Panel B: 2SLS							
Female Hiring	0.019	0.011	0.028	0.074***	0.051**	0.011	-0.004
	(0.025)	(0.023)	(0.025)	(0.027)	(0.024)	(0.023)	(0.026)
Observations	147,591	147,591	147,591	147,591	147,591	147,591	147,591
F-statistic	24.07	24.07	24.07	24.07	24.07	24.07	24.07
Panel C: OLS							
Female Hiring	0.081**	0.094***	0.075*	0.081*	0.079*	0.077*	0.063
	(0.040)	(0.034)	(0.044)	(0.049)	(0.041)	(0.045)	(0.044)
Observations	147,591	147,591	147,591	147,591	147,591	147,591	147,591
Fixed Effects							
Department	✓	✓	✓	✓	✓	✓	✓
Field × Year	✓	✓	✓	✓	✓	✓	✓
University × Year	✓	✓	✓	✓	✓	✓	✓

Note: The table presents regression results from estimating Equation (IV1) across multiple specifications. All regressions are estimated using a combined dataset constructed by stacking funding period-specific panels (as detailed in Section 1.4) and interacting all fixed effects with funding period indicators. Robust standard errors, clustered by department and funding period, are reported in parentheses. Significance levels are indicated as follows: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

First, I exclude the subsidized appointment itself. Second, I exclude former and future co-authors of subsidized appointments, as well as co-authors of these co-authors. Extending the peer network to include future co-authors might overstate the current network. However, future co-authors could already be part of the women's network today, even if no co-authorship exists yet. Third, I exclude authors who at some point have shared the same affiliation as the subsidized appointment, those employed in the same department during the same year. Fourth, I exclude authors working in the same specialized field as the subsidized appointment. I identify these authors using related works listed in the publication data.[29]

[29] Related works are identified algorithmically by comparing the titles and abstracts of papers. Specifically, the OpenAlex algorithm identifies papers that share common concepts with a

Figure 1.6d shows that applying these exclusion sets substantially attenuates the previous estimates. This suggests that the primary effect is driven by access to the peer networks associated with subsidized appointments. This increase may partly reflect a mechanical effect: when a professor joins a department, it naturally encourages greater collaboration between their existing co-author network and the current members of the department.

However, even after accounting for peer networks, I still observe a statistically significant increase in the share of female collaborators among junior men compared to departments with low retirement probabilities. This residual effect might provide evidence that gender attitudes are malleable through increased exposure to women, particularly among younger male scholars.

1.5.3 Effects on Quality and Direction of Research

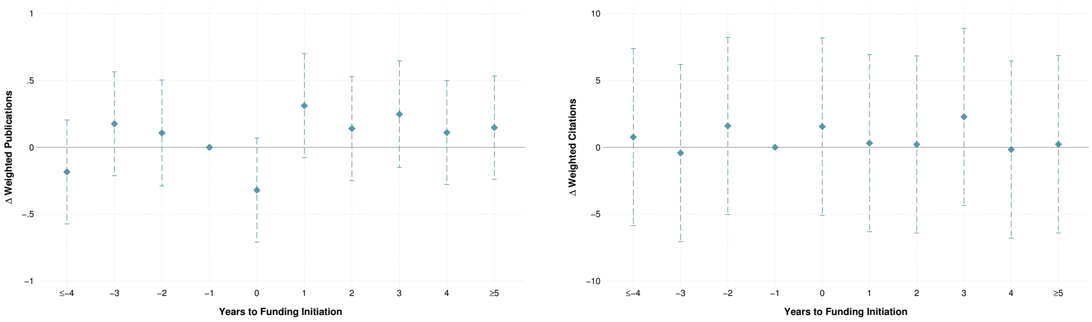
Diversity may affect the performance of existing department members. A range of studies examines how diversity impacts performance outside the academic context. For instance, Ahern and Dittmar (2012), Matsa and Miller (2013), and Nygaard (2011) examine the effects of Norway's board composition quota on firm performance and governance, yielding mixed results. Kim and Starks (2016) find that gender diversity on U.S. corporate boards enhances firm valuation, primarily due to the contributions of female directors. In Italy, Flabbi et al. (2019) show that female corporate leadership improves overall firm performance, positively affecting the upper end of the female wage distribution while negatively impacting the lower end. Hoogendoorn, Oosterbeek and Van Praag (2013) conduct a field experiment on business teams, revealing that mixed-gender teams outperform male-dominated teams in both profit and sales.

Publication Quality To test whether diversity affects scholarly output, I analyze whether a subsidized appointment influences the quality of publications by department members. Figures 1.7a and 1.7b show that the addition of a female professor has no impact on average publication quality of the department, as measured by either journal impact factor or citations.

given paper. Each work in OpenAlex is linked to several concepts sourced from a repository of approximately 65,000 concepts from Wikidata. For technical details on how concepts are assigned to papers, refer to OpenAlex's technical documentation available at OpenAlex's technical documentation.

Figure 1.7: Reduced Form – Effects on Research Output

(a) Journal Impact Factor (b) Citations



Note: Each figure presents reduced-form event study estimates based on the regression framework outlined in Section 1.4, using the research metric specified in the caption as outcome variable. The post-indicator is replaced by a set of indicators that represent the relative time in years from the start of the funding period, as shown on the horizontal axis. Robust standard errors, clustered by department and funding period, are reported in parentheses. Bars represent 95% confidence intervals.

Direction of Research Increasing gender diversity might also broaden the range of research questions pursued, as women often prioritize different policy areas compared to men (Dolado, Felgueroso and Almunia, 2012; Beneito et al., 2021). For example, surveys among economists indicate that women are generally more supportive of government intervention and environmental regulation, whereas men prioritize economic growth and are less concerned about inequality (Chari and Goldsmith-Pinkham, 2017; May, McGarvey and Kucera, 2018).

To test this hypothesis, I evaluate whether the research direction of departments shifts following the appointment of a female professor. I measure research direction using topic distributions constructed through a two-step procedure. First, I compute year-specific topic distributions for each department, capturing the extent to which researchers work on specific topics. In particular, for each department i in year t, I compute year-specific topic distributions by applying a topic model to all abstracts of papers published by researchers in the department.[30] A detailed description of the topic distribution construction is provided in Appendix Section 1.D.3. Next, I track how these distributions evolve over time by calculating the Mahalanobis distance relative to the pre-funding period. This scalar measure serves as the outcome variable in the empirical framework outlined in Section 1.4.

Figure 1.8a shows that topic distributions remain stable for up to six years after the appointment of a female professor, suggesting no significant shifts in

[30] Again, to avoid that the results are driven by changes in the department composition following a subsidized appointment, I fix the department composition in each funding period-specific panel in $\tau(g)$.

Table 1.9: Change in Publications

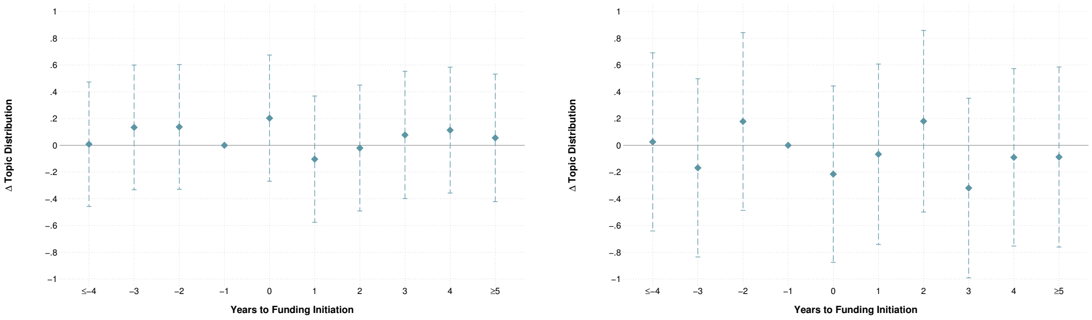
	Publication Outcomes			Research Direction	
	Total	Impact Factor	Citations	All Topics	Female Topics
	(1)	(2)	(3)	(4)	(6)
Panel A: Reduced Form					
Retire · Post	0.105	0.098	0.443	-0.032	0.054
	(0.215)	(0.173)	(2.062)	(0.212)	(0.246)
Retire · Eligible · Post	-0.313	0.106	0.732	0.054	-0.099
	(0.266)	(3.382)	(0.199)	(0.241)	(0.342)
Observations	147,591	147,591	147,591	147,591	147,591
Panel B: 2SLS					
Female Hiring	-0.665	0.225	1.554	0.115	0.210
	(0.521)	(0.194)	(2.930)	(0.412)	(0.617)
Observations	147,591	147,591	147,591	147,591	147,591
F-statistic	24.07	24.07	24.07	24.07	24.07
Panel C: OLS					
Female Hiring	-0.014	-0.012	-0.010	0.184***	0.281**
	(0.009)	(0.008)	(0.008)	(0.064)	(0.091)
Observations	147,591	147,591	147,591	147,591	147,591
Fixed Effects					
Department	✓	✓	✓	✓	✓
Field × Year	✓	✓	✓	✓	✓
University × Year	✓	✓	✓	✓	✓

Note: The table presents regression results from estimating Equation (IV1) across multiple specifications. All regressions are estimated using a combined dataset constructed by stacking funding period-specific panels (as detailed in Section 1.4) and interacting all fixed effects with funding period indicators. Robust standard errors, clustered by department and funding period, are reported in parentheses. Significance levels are indicated as follows: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

research direction. To further investigate whether researchers are more likely to work on topics traditionally associated with female scholars, I analogously train topic models exclusively on papers authored by women. As evident from Figure 1.8b, the topic distributions again remain stable across the entire event window, indicating no detectable shift toward 'female themes'.

Figure 1.8: Reduced Form – Effects on Direction of Research

(a) Overall Topic Distribution (b) Female Topic Distribution



Note: The figure presents reduced-form event study estimates based on the regression framework outlined in Section 1.4, using the change in topic distributions within departments relative to the pre-funding period as outcome variable. The post-indicator is replaced by a set of indicators that represent the relative time in years from the start of the funding period, as shown on the horizontal axis. Robust standard errors, clustered by department and funding period, are reported in parentheses. Bars represent 95% confidence intervals.

1.5.4 Policy Impact

For policymakers, a key question is the extent to which subsidized appointments contribute to the hiring of female professors who otherwise would not have been appointed. To address this question, it is essential to model the number of female full professors that would have been hired in the absence of the program. To construct this counterfactual, I outline a theoretical framework in Appendix Section 1.C.3. Based on this theoretical framework, I implement a regression framework where I evaluate whether fields with a high share of subsidized hires experience greater increases in female hiring compared to fields with a low share of subsidized hires, relative to the pre-funding period. Drawing on these theoretical derivations and considering all funding periods g, I implement the following regression specification:

$$\Delta f_{jg} = \alpha_j + \alpha_g + \alpha_j \cdot g + \pi f_{jg}^{AA} + \varepsilon_{jg}, \tag{POLICY}$$

where Δf_{jg} represents the change in the share of female hires in field j between the pre- and post-period years corresponding to g and f_{jg}^{AA} denotes the share of affirmative action hires – both directly observable in the data. The specification controls for time-invariant, field-specific factors, α_j, as well as for time-varying effects common to all fields, α_g. Additionally, field-specific time trends, $\alpha_j \cdot g$, allow for temporal variation in α_j over time.

Identification in this framework may fail for several reasons. First, it depends on the theoretical assumptions outlined in Appendix Section 1.C.3, which may not hold in practice – for example, the assumption that candidate pools are always sufficient to fill available positions. Second, the fixed effects in Equation (POLICY) must accurately capture the field-specific baseline change in female hiring, i.e., α_j in Appendix Section 1.C.3. Identification may break down if these baseline trends evolve non-linearly, which is not an unreasonable possibility. A further limitation is that the regression does not account for cross-hiring, where fields recruit candidates from neighboring disciplines. Consequently, estimates from Equation (POLICY) – particularly their magnitudes – should be interpreted with caution.

The parameter of interest, π, captures the average effective conversion rate of affirmative action-funded hires into additional female hires.[31] Analogous to the discussion in the theoretical framework, if affirmative action-funded hires merely replace female hires that would have occurred regardless of the policy, the share of female hires should remain unaffected by affirmative action hires, implying $\pi = 0$, conditional on fixed effects. Analogously, if every affirmative action hire displaces an otherwise male hire, the female hiring share should increase proportionately with the share of affirmative action hires, leading to $\pi = 1$, conditional on fixed effects.

The results in Panel A of Table 1.10 indicate that each affirmative action appointment leads to approximately $\hat{\pi} \approx 0.34$ additional female hires. This suggests that in about two-thirds of cases, departments use the subsidies to appoint women they would have hired anyway. At this rate, roughly 2.9 affirmative action appointments – costing approximately 2.2 million Euros – are needed to generate one additional female hire that would not have been recruited without the program.

The binscatter in Figure 1.9 visualizes this relationship by plotting the residualized shares against each other. The mean across the dots corresponds to the estimate shown in Column (4) of Panel A in Table 1.10. Moreover, the figure suggests that the marginal effectiveness of affirmative action-funded hires decreases as the share of affirmative action hires grows, $\frac{\partial \pi}{\partial f_{j1}^{AA}} < 0$. This reflects a scenario where initial affirmative action hires successfully add new female hires, but beyond a certain point, additional affirmative action hires mostly replace women who would have been hired anyway.

[31] Relating to the theoretical framework π represents the effective conversion rate in field j during period g, that is, $\pi = \frac{1}{N_G} \frac{1}{N_J} \sum_{g \in G} \sum_{j \in J} \bar{\pi}_{jg}$.

Table 1.10: Policy Effect Estimates

	Dependent Variable: Δ Share Female Hiring			
	(1)	(2)	(3)	(4)
Panel A: Sample of Fields				
Share AA Hiring	0.476***	0.374**	0.346*	0.343*
	(0.121)	(0.165)	(0.185)	(0.187)
Observations	198	198	198	198
Number of Clusters	33	33	33	33
Panel B: Sample of Faculties				
Share AA Hiring	0.366*	0.329*	0.295	0.314
	(0.189)	(0.193)	(0.212)	(0.201)
Observations	48	48	48	48
Number of Clusters	8	8	8	8
Fixed Effects				
Funding Period	-	✓	✓	✓
Unit	-	-	✓	✓
Unit-specific Trends	-	-	-	✓

Note: This table presents regression estimates of the impact of affirmative action-funded hires on the change in the share of female hires (Δf_{jg}) using the specification in Equation (POLICY). The regressions in Panel A are estimated on panel data across academic fields and funding periods. The regressions in Panel B are estimated on panel data across academic faculties and funding periods. All regressions include uni-specific fixed effects (α_j), common time effects (α_g), and unit-specific linear trends ($\alpha_j \cdot g$). Robust standard errors, clustered by field or faculty, are reported in parentheses. Significance levels are indicated as follows: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Addressing Potential Cross-Hiring Effects One limitation of the design is that it does not capture the possibility of cross-hiring – that is, fields may recruit candidates from neighboring disciplines. Although the direction of any bias introduced by cross-hiring is unclear a priori, I address this issue by re-estimating Equation (POLICY) at the faculty level. Under the assumption that cross-hirings occur only among closely related fields (e.g., between economics and business, but not between economics and engineering), this approach effectively rules out cross-hiring effects across observations.

Estimates from this exercise are presented in Panel B of Table 1.10. Notably, the sign and magnitude of the estimated effect remain similar to those obtained in the field-level regression, suggesting that each AA appointment translates to approximately $\hat{\pi} \approx 0.31$ additional female professors. However, the results are no longer statistically significant, likely due to the reduced variation and sample size resulting from aggregating fields into faculties.

Figure 1.9: Policy Effect Estimates – Visual

Note: This figure displays a binscatter plot constructed from the residualized values of the dependent variable (Δf_{jg}) and the share of AA-funded hires, f_{jg}^{AA}. To generate the plot, both variables are first residualized with respect to the covariates in Equation (POLICY); the residuals are then binned into equally spaced intervals, with the mean value of the dependent variable computed for each bin. The red cross marks the overall mean share of AA hires across all funding periods, while the two grey arrows indicate the two extreme cases for π.

1.6 Conclusion

This paper studies the impact of hiring a female professor. I address the endogeneity in hiring decisions by leveraging the introduction of the *Professorinnenprogramm*, an affirmative action policy by the German Ministry of Education. The program provides a five-year subsidy of up to 825,000 Euros (165,000 Euros annually) to cover the costs associated with the initial appointment of women to permanent full professorships. Since its inception in 2008, the program has facilitated the appointment of 845 women as full professors, accounting for 12% of all female appointments to full professorships, with a total expenditure of 820 million Euros. For identification, I employ an instrumental variable design using administrative data on all academic personnel employed at public German universities from 2002 to 2023. I utilize the program's requirement that subsidized appointments must eventually be converted into permanent positions, which makes departments with high retirement probabilities during the subsidy period marginally more likely to appoint a woman as a full professor.

My analysis suggests three lessons about the impact of appointing a female professor. First, exposure to a female professor increases the share of female PhD students by 18%. This effect is primarily driven by students who completed their undergraduate studies in the same department, suggesting that female professors act as role models and mentors for aspiring female academics. Notably, I do not observe changes in the hiring patterns for senior academic positions. Second, exposure to a female professor increases the number of female co-authors among junior men, mainly through collaboration with the newly hired woman's peer network. Third, I document that research output and research direction remain unaffected by the presence of an additional female professor.

When considering affirmative action policies, policymakers must weigh these benefits against the estimated costs of 2.2 million Euros per additional female professor.

Bibliography

Abadie, Alberto, Susan Athey, Guido W Imbens, and Jeffrey M Wooldridge. 2023. "When Should You Adjust Standard Errors for Clustering?" *Quarterly Journal of Economics*, 138(1): 1–35.

Ahern, Kenneth R, and Amy K Dittmar. 2012. "The Changing of the Boards: The Impact on Firm Valuation of Mandated Female Board Representation." *Quarterly Journal of Economics*, 127(1): 137–197.

Antecol, Heather, Kelly Bedard, and Jenna Stearns. 2018. "Equal but Inequitable: Who Benefits from Gender-neutral Tenure Clock Stopping Policies?" *American Economic Review*, 108(9): 2420–41.

Bagues, Manuel, Mauro Sylos-Labini, and Natalia Zinovyeva. 2017. "Does the Gender Composition of Scientific Committees Matter?" *American Economic Review*, 107(4): 1207–38.

Bagues, Manuel, Milan Makany, Giulia Vattuone, and Natalia Zinovyeva. 2023. "Women in Top Academic Positions: Is there a Trickle-down Effect?" Unpublished.

Baker, Andrew, Brantly Callaway, Scott Cunningham, Andrew Goodman-Bacon, and Pedro H. C. Sant'Anna. 2025. "Difference-in-Differences Designs: A Practitioner's Guide."

Beaman, Lori, Raghabendra Chattopadhyay, Esther Duflo, Rohini Pande, and Petia Topalova. 2009. "Powerful Women: Does Exposure Reduce Bias?" *Quarterly Journal of Economics*, 124(4): 1497–1540.

Belloni, Alexandre, Victor Chernozhukov, and Christian Hansen. 2014. "High-dimensional Methods and Inference on Structural and Treatment Effects." *Journal of Economic Perspectives*, 28(2): 29–50.

Beneito, Pilar, José E Boscá, Javier Ferri, and Manu García. 2021. "Gender Imbalance across Subfields in Economics: When Does it Start?" *Journal of Human Capital*, 15(3): 469–511.

Bertrand, Marianne, Sandra E Black, Sissel Jensen, and Adriana Lleras-Muney. 2019. "Breaking the Glass Ceiling? The Effect of Board Quotas on Female Labour Market Outcomes in Norway." *Review of Economic Studies*, 86(1): 191–239.

Besley, Timothy, Olle Folke, Torsten Persson, and Johanna Rickne. 2017. "Gender Quotas and the Crisis of the Mediocre Man: Theory and Evidence from Sweden." *American Economic Review*, 107(8): 2204–2242.

Bhalotra, Sonia, Irma Clots-Figueras, and Lakshmi Iyer. 2018. "Pathbreakers? Women's Electoral Success and Future Political Participation." *The Economic Journal*, 128(613): 1844–1878.

Bhavnani, Rikhil R. 2017. "Do the Effects of Temporary Ethnic Group Quotas Persist? Evidence from India." *American Economic Journal: Applied Economics*, 9(3): 105–123.

Blau, Francine D, Janet M Currie, Rachel TA Croson, and Donna K Ginther. 2010. "Can Mentoring Help Female Assistant Professors? Interim Results from a Randomized Trial." *American Economic Review*, 100(2): 348–52.

BMJ, Bundesministerium der Justiz. 2024. "Gesetz über die Versorgung der Beamten und Richter des Bundes (Beamtenversorgungsgesetz – BeamtVG)." *URL: https://www.gesetze-im-internet.de/beamtvg/BJNR024850976.html*, Accessed: 2024-06-03.

Bundesanzeiger, Bundesministerium für Bildung und Forschung. 2018. "Richtliniezur Umsetzung des Professorinnenprogramms des Bundes und der Länder." *URL: https://www.bundesanzeiger.de/pub/de/amtliche-veroeffentlichung?2*, Accessed: 2022-06-24.

Bundesministerium für Bildung und Forschung. 2008. "Bekanntmachung des Bundesministeriums für Bildung und Forschung von Richtlinien zur Umsetzung des Professorinnenprogramms des Bundes und der Länder zur Förderung der Gleichstellung von Frauen und Männern in Wissenschaft und Forschung an deutschen Hochschulen." *URL: https://www.gwk-bonn.de/presseaktuelles/presse-archiv*, Accessed: 2024-09-19.

Bundesministerium für Bildung und Forschung. 2012. "Bekanntmachung des Bundesministeriums für Bildung und Forschung von Richtlinien zur Umsetzung des Professorinnenprogramms des Bundes und der Länder zur Förderung der Gleichstellung von Frauen und Männern in Wissenschaft und Forschung an deutschen Hochschulen – Professorinnenprogramm II." *URL: https://www.bmbf.de/bmbf/shareddocs/bekanntmachungen/de/2012/12/797_bekanntmachung.html#searchFacets*, Accessed: 2024-09-19.

Bundesministerium für Bildung und Forschung. 2018. “Bekanntmachung – Richtlinie zur Umsetzung des Professorinnenprogramms des Bundes und der Länder zur Förderung der Gleichstellung von Frauen und Männern in Wissenschaft und Forschung an deutschen Hochschulen – Professorinnenprogramm III. Bundesanzeiger vom 21.02.2018.” *URL:* *https://www.bmbf.de/bmbf/shareddocs/bekanntmachungen/de/2018/02/1600_bekanntmachung.html#searchFacets*, Accessed: 2024-09-19.

Bundesministerium für Bildung und Forschung. 2022. “Bekanntmachung – Richtlinie zur Umsetzung des Professorinnenprogramms des Bundes und der Länder zur Förderung der Gleichstellung von Frauen und Männern in Wissenschaft und Forschung an deutschen Hochschulen – Professorinnenprogramm 2030.” *URL: https://www.bmbf.de/bmbf/shareddocs/bekanntmachungen/de/2023/02/2023-02-02-Bekanntmachung-Professorinnenprogramm.html#searchFacets*, Accessed: 2024-09-19.

Bundesregierung, Die. 2023. “Förderkatalog.” *URL: https://foerderportal.bund.de/foekat/jsp/SucheAction.do?actionMode=searchmask*, Accessed: 2023-10-24.

BW-MF, Baden-Württemberg Ministerium der Finanzen. 2022. “Staatshaushaltsplan für 2022 – Einzelplan 14 Ministerium für Wissenschaft, Forschung und Kunst.” Accessed: 2024-06-03.

Card, David, Stefano DellaVigna, Patricia Funk, and Nagore Iriberri. 2020. “Are Referees and Editors in Economics Gender Neutral?” *Quarterly Journal of Economics*, 135(1): 269–327.

Carnes, Molly, Patricia G Devine, Linda Baier Manwell, Angela Byars-Winston, Eve Fine, Cecilia E Ford, Patrick Forscher, Carol Isaac, Anna Kaatz, Wairimu Magua, et al. 2015. “Effect of an Intervention to Break the Gender Bias Habit for Faculty at One Institution: A Cluster Randomized, Controlled Trial.” *Academic Medicine: Journal of the Association of American Medical Colleges*, 90(2): 221.

Carrell, Scott E, Marianne E Page, and James E West. 2010. “Sex and Science: How Professor Gender Perpetuates the Gender Gap.” *Quarterly Journal of Economics*, 125(3): 1101–1144.

Carrell, Scott E, Mark Hoekstra, and James E West. 2015. "The Impact of Intergroup Contact on Racial Attitudes and Revealed Preferences." *NBER Working Paper*, 20940.

Cengiz, Doruk, Arindrajit Dube, Attila Lindner, and Ben Zipperer. 2019. "The Effect of Minimum Wages on Low-Wage Jobs." *Quarterly Journal of Economics*, 134(3): 1405–1454.

Chari, Anusha, and Paul Goldsmith-Pinkham. 2017. "Gender Representation in Economics Across Topics and Time: Evidence from the NBER Summer Institute." *NBER Working Paper*, 23953.

Chattopadhyay, Raghabendra, and Esther Duflo. 2004. "Women as Policy Makers: Evidence from a Randomized Policy Experiment in India." *Econometrica*, 72(5): 1409–1443.

CHE Hochschuldaten. 2024. "Studierende in Deutschland." *URL:* `https://hochschuldaten.che.de/deutschland/studierende-in-deutschland/#:~:text=Die%20meisten%20Studierenden%20(58%2C2,FH)%20eingeschrieben.` , Accessed: 2024-06-03.

Chen, Le-Yu, and Sokbae Lee. 2021. "Binary Classification with Covariate Selection Through ℓ_0-penalised Empirical Risk Minimisation." *Econometrics Journal*, 24(1): 103–120.

Dahl, Gordon B, Andreas Kotsadam, and Dan-Olof Rooth. 2021. "Does Integration Change Gender Attitudes? The Effect of Randomly Assigning Women to Traditionally Male Teams." *Quarterly Journal of Economics*, 136(2): 987–1030.

de Chaisemartin, Clément, and Xavier D'Haultfœuille. 2023. "Two-Way Fixed Effects and Differences-in-Differences Estimators with Several Treatments." *Journal of Econometrics*, 236(2): 105480.

De Paola, Maria, Vincenzo Scoppa, and Rosetta Lombardo. 2010. "Can Gender Quotas Break Down Negative Stereotypes? Evidence from Changes in Electoral Rules." *Journal of Public Economics*, 94(5-6): 344–353.

Deschamps, Pierre. 2018. "Gender Quotas in Hiring Committees: A Boon or a Bane for Women?" *Sciences Po LIEPP Working Paper*, 82.

Destatis, Federal Statistical Office of Germany. 2018. "Hochschulpersonalstatistik." Administrative data. Information on how to request the data can be found at `https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Bildung-Forschung-Kultur/Hochschulen/Methoden/Erlaeuterungen/hochschulen.html`.

Devine, Patricia G, Patrick S Forscher, William TL Cox, Anna Kaatz, Jennifer Sheridan, and Molly Carnes. 2017. "A Gender Bias Habit-Breaking Intervention Led to Increased Hiring of Female Faculty in STEMM Departments." *Journal of Experimental Social Psychology*, 73: 211–215.

Dolado, Juan J, Florentino Felgueroso, and Miguel Almunia. 2012. "Are Men and Women-economists Evenly Distributed across Research Fields? Some New Empirical Evidence." *SERIEs – Journal of the Spanish Economic Association*, 3: 367–393.

Dube, Arindrajit, Daniele Girardi, Oscar Jorda, and Alan M Taylor. 2023. "A Local Projections Approach to Difference-in-Differences Event Studies." *American Economic Review.*

Dupas, Pascaline, Alicia Modestino, Muriel Niederle, Justin Wolfers, et al. 2021. "Gender and the Dynamics of Economics Seminars." *NBER Working Paper*, 28494.

European Commission, Directorate General for Research. 2021. "Gender in Research and Innovation: Statistics and Indicators." *URL: https://op.europa.eu/en/web/eu-law-and-publications/publication-detail/-/publication/67d5a207-4da1-11ec-91ac-01aa75ed71a1*, Accessed: 2022-05-26.

Flabbi, Luca, Mario Macis, Andrea Moro, and Fabiano Schivardi. 2019. "Do Female Executives Make a Difference? The Impact of Female Leadership on Gender Gaps and Firm Performance." *Economic Journal*, 129(622): 2390–2423.

Friedman, Jerome, Trevor Hastie, and Rob Tibshirani. 2010. "Regularization Paths for Generalized Linear Models via Coordinate Descent." *Journal of Statistical Software*, 33(1): 1.

Ginther, Donna K, Janet M Currie, Francine D Blau, and Rachel TA Croson. 2020. "Can Mentoring Help Female Assistant Professors in Economics? An Evaluation by Randomized Trial." Vol. 110, 205–09.

Gonsalkorale, Karen, Jeffrey W Sherman, and Karl Christoph Klauer. 2009. "Aging and Prejudice: Diminished Regulation of Automatic Race Bias Among Older Adults." *Journal of Experimental Social Psychology*, 45(2): 410–414.

Guarino, Cassandra M, and Victor MH Borden. 2017. "Faculty Service Loads and Gender: Are Women Taking Care of the Academic Family?" *Research in Higher Education*, 58(6): 672–694.

Hansen, Bruce. 2022. *Econometrics.* Princeton University Press.

Hochschulverband, Deutscher, ed. 2002–2022. *Hochschullehrer Verzeichnis – Universitäten Deutschland.* De Gruyter Saur.

Hoogendoorn, Sander, Hessel Oosterbeek, and Mirjam Van Praag. 2013. "The Impact of Gender Diversity on the Performance of Business Teams: Evidence from a Field Experiment." *Management Science*, 59(7): 1514–1528.

HRK, Hochschulrektorenkonferenz. 2024. "Hochschulfinanzierung." *URL:* `https://www.hrk.de/themen/hochschulsystem/hochschulfinanzierung/`, Accessed: 2024-06-03.

Janys, Lena. 2024. "Testing the Presence of Implicit Hiring Quotas with Application to German Universities." *Review of Economics and Statistics*, 106(3): 627–637.

Jensen, Robert, and Emily Oster. 2009. "The Power of TV: Cable Television and Women's Status in India." *Quarterly Journal of Economics*, 124(3): 1057–1094.

Kim, Daehyun, and Laura T Starks. 2016. "Gender Diversity on Corporate Boards: Do Women Contribute Unique Skills?" *American Economic Review*, 106(5): 267–71.

Kleemans, Marieke, and Rebecca L Thornton. 2021. "Who Belongs? The Determinants of Selective Membership into the National Bureau of Economic Research." Vol. 111, 117–22.

Lewicki, Pawel. 1983. "Self-image Bias in Person Perception." *Journal of Personality and Social Psychology*, 45(2): 384.

Matsa, David A, and Amalia R Miller. 2013. "A Female Style in Corporate Leadership? Evidence from Quotas." *American Economic Journal: Applied Economics*, 5(3): 136–169.

May, Ann Mari, Mary G McGarvey, and David Kucera. 2018. "Gender and European Economic Policy: A Survey of the Views of European Economists on Contemporary Economic Policy." *Kyklos*, 71(1): 162–183.

Mengel, Friederike, Jan Sauermann, and Ulf Zölitz. 2019. "Gender Bias in Teaching Evaluations." *Journal of the European Economic Association*, 17(2): 535–566.

National Health & Medical Research Council. 2022. "Working towards gender equity in Investigator Grants." *URL: https://www.nhmrc.gov.au/about-us/news-centre/working-towards-gender-equity-investigator-grants*, Accessed: 2024-06-19.

NRW, Gesetz über die Hochschulen des Landes Nordrhein-Westfalen. 2014. "§37a – Gewährleistung der Chancengerechtigkeit von Frauen und Männern bei der Berufung von Professorinnen und Professoren." *URL: https://recht.nrw.de/lmi/owa/br_bes_detail?sg=0&menu=0&bes_id=28364&anw_nr=2&aufgehoben=N&det_id=643733*, Accessed: 2024-06-24.

NRW, Ministerium des Innern des Landes Nordrhein-Westfalen. 2009. "Verordnung über die Lehrverpflichtung an Universitäten und Hochschulen für angewandte Wissenschaften." *URL: https://recht.nrw.de/lmi/owa/br_text_anzeigen?v_id=10000000000000000609*, Accessed: 2024-06-03.

Nygaard, Knut. 2011. "Forced Board Changes: Evidence from Norway." *NHH Department of Economics Discussion Paper*, 5.

Porter, Catherine, and Danila Serra. 2020. "Gender Differences in the Choice of Major: The Importance of Female Role Models." *American Economic Journal: Applied Economics*, 12(3): 226–54.

Priem, Jason, Heather Piwowar, and Richard Orr. 2022. "OpenAlex: A Fully-open Index of Scholarly Works, Authors, Venues, Institutions, and Concepts." *arXiv Pre-print*, 2205.01833.

Roth, Jonathan, Pedro HC Sant'Anna, Alyssa Bilinski, and John Poe. 2023. "What's Trending in Difference-in-Differences? A Synthesis of the Recent Econometrics Literature." *Journal of Econometrics*, 235(2): 2218–2244.

Sarsons, Heather, Klarita Gërxhani, Ernesto Reuben, and Arthur Schram. 2021. "Gender Differences in Recognition for Group Work." *Journal of Political Economy*, 129(1): 101–147.

Scopus, Elsevier. 2023. "CiteScore 2023." *URL:* `https://www.scopus.com/sources`, Accessed: 2024-06-24.

Siniscalchi, Marciano, and Pietro Veronesi. 2020. "Self-image Bias and Lost Talent." *NBER Working Paper*, 28308.

Siyanova-Chanturia, Anna, Paul Warren, Francesca Pesciarelli, and Cristina Cacciari. 2015. "Gender Stereotypes Across the Ages: On-line Processing in School-age Children, Young and Older Adults." *Frontiers in Psychology*, 6: 1388.

Swiss National Science Foundation. 2021. "Quoten für mehr Gleichstellung in der Forschung." *URL:* `https://www.snf.ch/de/kjCKYzJgvuNWbsf2/news/news-210205-quoten-fuer-mehr-gleichstellung-in-der-forschung`, Accessed: 2024-06-19.

Vernos, Isabelle. 2013. "Quotas Are Questionable." *Nature*, 495(7439): 39–39.

Wallon, Gerlind, Sandra Bendiscioli, and Michele S Garfinkel. 2015. "Exploring Quotas in Academia." *URL:* `www.embo.org/documents/science_policy/exploring_quotas.pdf`, Accessed: 2024-06-24.

Whelan, Jennifer, and Robert Wood. 2012. "Targets and Quotas for Women in Leadership: A Global Review of Policy, Practice and Psychological Research." *Gender Equality Project, Centre for Ethical Leadership, Melbourne Business School.*

Wing, Coady, Seth M Freedman, and Alex Hollingsworth. 2024. "Stacked Difference-in-Differences." *NBER Working Paper*, 32054.

1.A Additional Tables

Table 1.A.1: Public Universities by State

No.	State	University	No.	State	University
1	BB	Brandenburg University of Technology	43	NI	Osnabrück University
2	BB	European University Viadrina Frankfurt	44	NI	Technical University of Braunschweig
3	BB	Film University Babelsberg	45	NI	University of Göttingen
4	BB	University of Potsdam	46	NI	University of Hildesheim
5	BE	Free University of Berlin	47	NI	University of Lüneburg
6	BE	Humboldt University of Berlin	48	NI	University of Oldenburgurg
7	BE	Technical University of Berlin	49	NI	University of Vechta
8	BW	Heidelberg University	50	NI	University of Veterinary Medicine
9	BW	Karlsruhe Institute of Technology	51	NW	Bielefeld University
10	BW	University of Freiburg	52	NW	German Sport University Cologne
11	BW	University of Hohenheim	53	NW	Ruhr University Bochum
12	BW	University of Konstanz	54	NW	RWTH Aachen University
13	BW	University of Mannheim	55	NW	Technical University of Dortmund
14	BW	University of Stuttgart	56	NW	University of Bonn
15	BW	University of Tübingen	57	NW	University of Cologne
16	BY	Bundeswehr University Munich	58	NW	University of Duisburg-Essen
17	BY	Catholic University of Eichstätt-Ingolstadt	59	NW	University of Dusseldorf
18	BY	Technical University of Munich	60	NW	University of Hagen
19	BY	University of Augsburg	61	NW	University of Münster
20	BY	University of Bamberg	62	NW	University of Paderborn
21	BY	University of Bayreuth	63	NW	University of Siegen
22	BY	University of Erlangen-Nuremberg	64	NW	University of Wuppertal
23	BY	University of Munich	65	RP	University of Administrative Sciences
24	BY	University of Passau	66	RP	University of Kaiserslautern
25	BY	University of Regensburg	67	RP	University of Koblenz and Landau
26	BY	University of Ulm	68	RP	University of Mainz
27	BY	University of Würzburg	69	RP	University of Trier
28	HB	University of Bremen	70	SH	Kiel University
29	HE	Goethe University Frankfurt	71	SH	University of Flensburg
30	HE	Technical University of Darmstadt	72	SH	University of Lübeck
31	HE	University of Giessen	73	SL	Saarland University
32	HE	University of Kassel	74	SN	Chemnitz University of Technology
33	HE	University of Marburg	75	SN	Dresden University of Technology
34	HH	HafenCity University Hamburg	76	SN	Freiberg University of Mining
35	HH	Hamburg University of Technology	77	SN	Leipzig University
36	HH	Helmut Schmidt University	78	ST	University Halle-Wittenberg
37	HH	University of Hamburg	79	ST	University Magdeburg
38	MV	University of Greifswaldd	80	TH	Bauhaus University Weimar
39	MV	University of Rostock	81	TH	Ilmenau University of Technology
40	NI	Clausthal University of Technology	82	TH	University of Erfurt
41	NI	Hannover Medical School	83	TH	University of Jena
42	NI	Leibniz University Hannover			

Note: The table presents a list of all public universities along with their corresponding states included in the analysis, as listed in the 'Hochschulpersonalstatistik' (Destatis, 2018).

Table 1.A.2: Subsidy Characteristics by Faculty and Field

			Subsidized Hirings		Subsidy Characteristics		
No.	Faculty	Field	Share	Total	Duration	Amount	Regular
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
1	Humanities	Media Studies	5.57	24	4.68	351	0.71
2	Humanities	Language Studies	5.10	22	4.84	336	0.91
3	Humanities	German Studies	5.10	22	4.63	299	0.73
4	Humanities	History	3.71	16	4.52	300	0.81
5	Humanities	Philosophy	2.32	10	4.26	298	0.60
6	Humanities	Theology	1.39	6	4.74	324	0.83
7	Sports Sciences	Sport	1.62	7	4.63	353	0.57
8	Social Sciences	Educational Sciences	8.58	37	4.81	344	0.89
9	Social Sciences	Business	6.73	29	4.65	346	0.90
10	Social Sciences	Sociology	5.80	25	4.29	301	0.72
11	Social Sciences	Legal Sciences	4.18	18	4.80	367	0.83
12	Social Sciences	Psychology	3.94	17	4.49	335	0.76
13	Social Sciences	Political Sciences	3.25	14	4.51	310	0.71
14	Social Sciences	Economics	2.09	9	4.92	394	0.78
15	Natural Sciences	Biology	5.80	25	4.72	356	0.76
16	Natural Sciences	Chemistry	5.34	23	4.76	347	0.78
17	Natural Sciences	Mathematics	3.71	16	4.90	367	0.69
18	Natural Sciences	Physics	2.32	10	4.92	381	0.60
19	Natural Sciences	Geography	2.09	9	4.31	338	0.56
20	Life Sciences	Medicine	3.94	17	4.81	333	0.41
21	Life Sciences	Dentistry	0.46	2	4.92	375	0.50
22	Agricultural Sciences	Biotechnology	1.62	7	4.92	389	1.00
23	Agricultural Sciences	Forestry	0.93	4	4.44	313	0.75
24	Agricultural Sciences	Veterinary Medicine	0.23	1	2.00	171	0.00
25	Agricultural Sciences	Nutrital Sciences	0.00	0	-	-	-
26	Engineering	Architecture	3.25	14	4.88	362	0.79
27	Engineering	Computer Science	3.02	13	4.35	342	0.23
28	Engineering	Mechanical Engineering	1.39	6	4.92	379	1.00
29	Engineering	Electrical Engineering	1.39	6	4.92	369	0.33
30	Engineering	Civil Engineering	1.16	5	4.82	345	0.40
31	Engineering	Mining	0.23	1	4.92	375	1.00
32	Art Sciences	Visual Arts	2.09	9	4.68	311	0.89
33	Art Sciences	Musicology	1.62	7	4.35	307	0.86

Note: The table presents a list of all fields and their corresponding faculties included in the analysis, as listed in the 'Hochschulpersonalstatistik' (Destatis, 2018). Columns (4) and (5) provide the share and number of subsidized appointments by field, while columns (6) to (8) detail the subsidy duration, amount, and type by field.

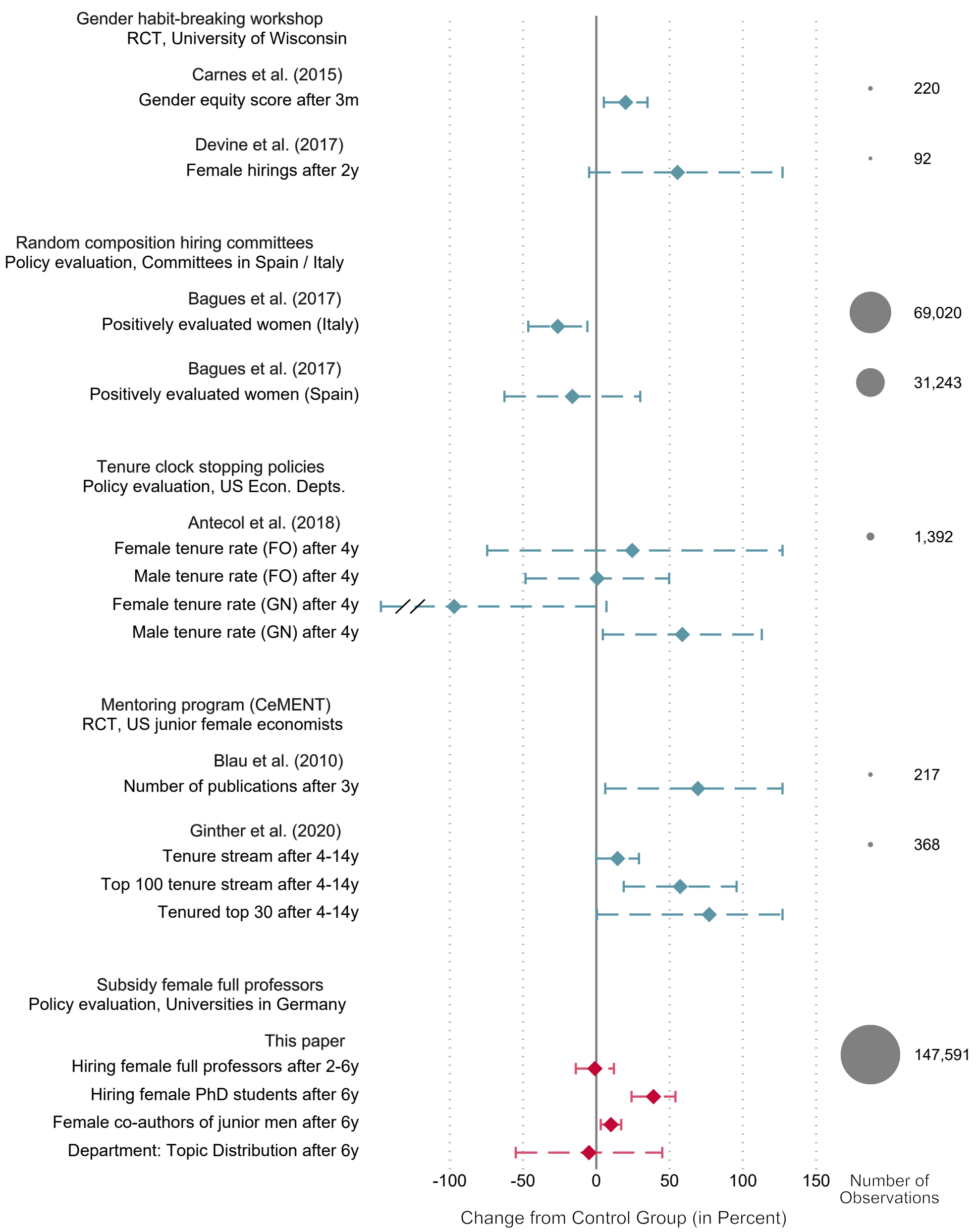
Table 1.A.3: Full Professor Characteristics by Gender

	Men			Women			Difference
	Mean	SD	N	Mean	SD	N	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A: Field							
Humanities	0.170	0.376	183,244	0.314	0.464	45,314	0.145***
Sports	0.012	0.107	183,244	0.001	0.098	45,314	-0.002***
Social Sciences	0.238	0.426	183,244	0.302	0.459	45,314	0.063***
Natural Sciences	0.297	0.457	183,244	0.185	0.388	45,314	-0.112***
Health Sciences	0.024	0.154	183,244	0.049	0.216	45,314	0.025***
Agricultural Sciences	0.028	0.164	183,244	0.028	0.164	45,314	0.000
Engineering	0.200	0.400	183,244	0.083	0.276	45,314	-0.118***
Arts	0.024	0.154	183,244	0.049	0.216	45,314	0.025***
Panel B: Compensation							
C4	0.315	0.465	183,244	0.152	0.359	45,314	-0.143***
C3	0.169	0.374	183,244	0.164	0.370	45,314	0.008***
W3	0.326	0.469	183,244	0.349	0.477	45,314	-0.001
W2	0.152	0.359	183,244	0.273	0.445	45,314	0.111***
Full-time	0.981	0.137	183,244	0.967	0.179	45,314	-0.014***
Panel C: Financing Source							
Regular Budget	0.876	0.330	183,244	0.818	0.385	45,314	-0.053***
DFG Funds	0.006	0.075	183,244	0.009	0.095	45,314	0.003***
EU Funds	0.002	0.046	183,244	0.005	0.069	45,314	0.002***
Excellence Initiative	0.006	0.078	183,244	0.010	0.099	45,314	0.003***
Panel D: Leadership Positions							
Rector	0.000	0.015	42,289	0.000	0.121	13,066	0.000
Prorector	0.003	0.058	42,289	0.006	0.121	13,066	0.003***
President	0.000	0.005	42,289	0.000	0.121	13,066	0.000
Vice-President	0.004	0.067	42,289	0.008	0.121	13,066	0.004***
Chancellor	0.000	0.000	42,289	0.000	0.000	13,066	0.000
Panel E: Pre-Tenure Position							
Ass. Prof. w/o TT	0.035	0.184	36,229	0.062	0.241	10,376	0.027***
Ass. Prof. with TT	0.012	0.108	36,229	0.024	0.154	10,376	0.012***
W2 w/o TT	0.034	0.180	36,229	0.046	0.209	10,376	0.012***
W2 with TT	0.008	0.087	36,229	0.011	0.103	10,376	0.003***
Habilitation	0.602	0.490	36,229	0.541	0.498	10,376	-00.060***
Habilitation (equivalent)	0.207	0.405	36,229	0.202	0.402	10,376	0.003***
Panel F: Individual Characteristics							
Age	51.825	8.190	183,244	49.163	7.806	45,314	-2.678***
Age Tenure	40.775	5.010	114,156	41.231	5.089	25,074	0.595***
Age Highest Degree	37.564	3.969	26,451	38.902	4.471	7,201	1.339***
German	0.925	0.264	183,244	0.911	0.285	45,314	-0.013***
PhD Highest Degree	0.338	0.473	41,391	0.400	0.490	12,690	0.061***
Habilitation Highest Degree	0.639	0.480	41,391	0.567	0.495	12,690	-0.071***

Note: The table shows descriptive statistics for the sample of full professors from 2008–2022. The difference reported in column (7) is the coefficient obtained by regressing an indicator for women on the respective variable controlling for year fixed effects. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

1.B Additional Figures

Figure 1.B.1: Literature Overview



Note: Effect and sample sizes for the evaluation of the gender habit-breaking workshop correspond to Table 3 in Carnes et al. (2015) and Table 1 in Devine et al. (2017). Effect and sample sizes for the random allocation of hiring committees in Italy and Spain are taken from Table 1 in Bagues, Sylos-Labini and Zinovyeva (2017). Effect and sample sizes for the evaluation of tenure clock stopping policies are taken from Table 2 in Antecol, Bedard and Stearns (2018). Effect and sample sizes for the evaluation of the CeMENT mentoring program are retrieved from Table 2 in Blau et al. (2010) and Table 3 in Ginther et al. (2020). Treatment effects across studies are made comparable by considering the main specification of each paper and computing the percent increase from the pre-treatment control-group mean.

Figure 1.B.2: Employment Plan University of Mannheim

Ministerium für Wissenschaft, Forschung und Kunst

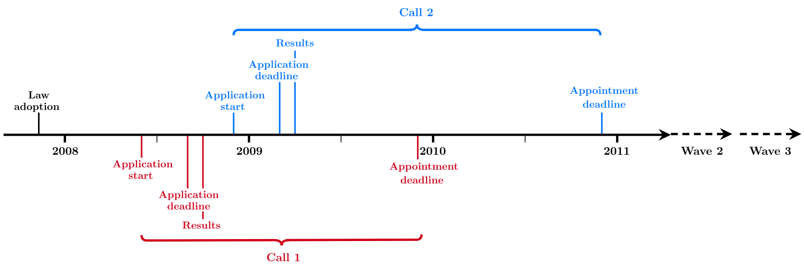
1420 Universität Mannheim

Tit. Bes.Gr. Entg.Gr.	FKZ	Bezeichnung	Stellenzahl 2019	2020	2021
682 01	133	**Stellenplan für Beamtinnen und Beamte im Landesbetrieb Universität Mannheim**			
		1. Vgl. Vermerke bei Kap. 1402 Tit. 422 01 und 428 01. 2. Die in der Stellenübersicht im Wirtschaftsplan für Arbeitnehmerinnen und Arbeitnehmer aufgeführten Stellen dürfen, soweit es dienstlich notwendig ist, bzgl. Dienstarten und Wertigkeit anderweitig bis Entg.Gr. 14 TV-L besetzt werden. Voraussetzung ist Kostenneutralität und Einhaltung des Stellensolls.			
		a) Planstellen für Beamtinnen und Beamte im Landesbetrieb			
W 3		Rektor/Präsident	1,0	1,0	1,0
W 3		Kanzler	1,0	1,0	1,0
W 3		Universitätsprofessor	140,0	152,0	151,0
		kw nach Ablauf der Förderung, spätestens ab 01.01.2021 4)	* 0,0	* 1,0	* 0,0
		kw nach Ablauf der Förderung, spätestens ab 01.08.2043 6)	* 0,0	* 1,0	* 1,0
		kw nach Ablauf der Förderung, spätestens ab 01.02.2037 5)	* 0,0	* 1,0	* 1,0
		kw nach Ablauf der Förderung, spätestens ab 01.08.2032 1)	* 0,0	* 1,0	* 1,0
		kw nach Ablauf der Förderung, spätestens ab 01.01.2020 5)	* 1,0	* 0,0	* 0,0
		kw nach Ablauf der Förderung, spätestens ab 01.01.2022 1)	* 1,0	* 0,0	* 0,0
		kw nach Ablauf der Förderung, spätestens ab 01.01.2022 2)	* 3,0	* 3,0	* 3,0
W 2		Universitätsprofessor	3,0	3,0	3,0
W 1		Professor als Juniorprofessor	57,5	61,5	56,5
		kw nach Ablauf der Förderung, spätestens ab 01.01.2021 9)	* 2,0	* 2,0	* 0,0
		kw nach Ablauf der Förderung, spätestens ab 01.01.2021 7)	* 0,0	* 1,0	* 0,0
		kw nach Ablauf der Förderung, spätestens ab 01.09.2020 8)	* 0,0	* 1,0	* 0,0
		kw nach Ablauf der Förderung, spätestens ab 01.01.2025 11)	* 0,0	* 2,0	* 2,0
		kw nach Ablauf der Förderung, spätestens ab 01.03.2020 10)	* 1,0	* 1,0	* 0,0
A 16		Leitender Regierungsdirektor	0,0	1,0	1,0
A 16		Leitender Akademischer Direktor	1,0	1,0	1,0
A 16		Leitender Bibliotheksdirektor	1,0	1,0	1,0
A 15		Regierungsdirektor	5,0	5,0	5,0
A 15		Akademischer Direktor	7,0	7,0	7,0
A 15		Bibliotheksdirektor	3,0	3,0	3,0
A 14		Oberregierungsrat	2,0	1,0	1,0
A 14		Akademischer Oberrat	41,5	41,5	41,5
A 14		Oberbibliotheksrat	5,0	5,0	5,0
A 13		Regierungsrat	7,0	7,0	7,0
A 13		Akademischer Rat 3)	60,5	60,5	60,5
A 13		Bibliotheksrat	3,0	3,0	3,0
A 13		Oberamtsrat (Bi)	3,0	3,0	3,0

-970-

Note: The figure displays an excerpt from the 2019 budget of the state of Baden-Wuerttemberg, listing the Employment Plan for the University of Mannheim (Ministerium für Wissenschaft, Forschung und Kunst Baden-Württemberg, 2020).

Figure 1.B.3: Application Timeline



Note: The figure exemplarily presents the timeline of application steps for the first two funding periods, as outlined in Gemeinsame Wissenschaftskonferenz (2013–2022). Each funding period includes an application phase, followed by a funding phase during which successfully evaluated universities can use subsidies to appoint up to three women to full professor positions.

Figure 1.B.4: Job Advertisement Example – University of Tübingen

EBERHARD KARLS UNIVERSITÄT TÜBINGEN

Philosophische Fakultät

Professur (W3) für Kunstgeschichte

Am Kunsthistorischen Institut der Universität Tübingen ist zum 01.10.2019 eine

Professur (W3) für Kunstgeschichte

zu besetzen.

Die Stelleninhaberin/Der Stelleninhaber soll das Fach in Forschung und Lehre in großer Breite vertreten. Erwartet wird ein ausgewiesener Forschungsschwerpunkt im Bereich der Kunst des Mittelalters; einschlägige Kompetenzen in der Architekturgeschichte sind erwünscht, aber nicht Voraussetzung. Neben der Beteiligung an allen kunsthistorischen Studiengängen werden die Fähigkeit und Bereitschaft zur interdisziplinären Zusammenarbeit und insbesondere zur Mitwirkung in interdisziplinären Forschungsverbünden der Fakultät erwartet.

Einstellungsvoraussetzungen sind die Habilitation oder gleichwertige wissenschaftliche Leistungen, international beachtete Publikationen sowie nachgewiesene didaktische Eignung.

Diese Professur wird im Rahmen des Professorinnenprogramms III des Bundes und der Länder ausgeschrieben. Eine Besetzung der Stelle erfolgt vorbehaltlich der Zuweisung der im Professorinnenprogramm III beantragten Mittel.

The professorship is being advertised as part of the Professorinnenprogramm III of the German federal and state governments. The position will be filled subject to the allocation of the funds requested through the Professorinnenprogramm III.

Qualifizierte internationale Wissenschaftlerinnen und Wissenschaftler sind ausdrücklich aufgefordert, sich zu bewerben.

Schwerbehinderte werden bei gleicher Eignung bevorzugt berücksichtigt.

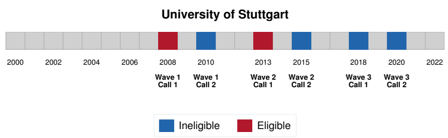
Bewerbungen mit den üblichen Unterlagen (Lebenslauf, Zeugnisse, Schriftenverzeichnis, Verzeichnis der abgehaltenen Lehrveranstaltungen) sowie den selbst verfassten Monographien und bis zu 5 Aufsätzen sind möglichst in elektronischer Form **bis zum 15.03.2019** zu richten an bewerbung@philosophie.uni-tuebingen.de (Postanschrift: Dekan der Philosophischen Fakultät, Keplerstr. 2, 72074 Tübingen). Rückfragen können direkt an den Dekan gerichtet werden (Prof. Dr. Jürgen Leonhardt, juergen.leonhardt@uni-tuebingen.de).

Note: The figure displays a job advertisement from the Art History department at the University of Tübingen (Eberhard Karls Universität Tübingen, 2019), intended to be funded through the *Professorinnenprogramm*.

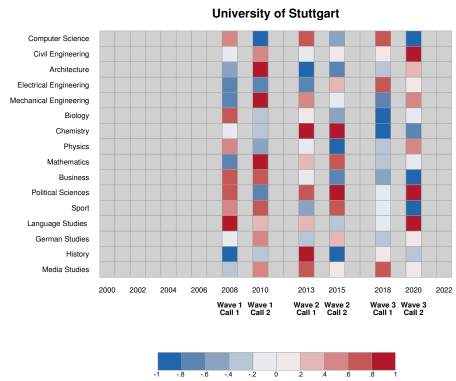
Figure 1.B.5: Identifying Variation

Panel A: University of Stuttgart

(a) Eligibility Status

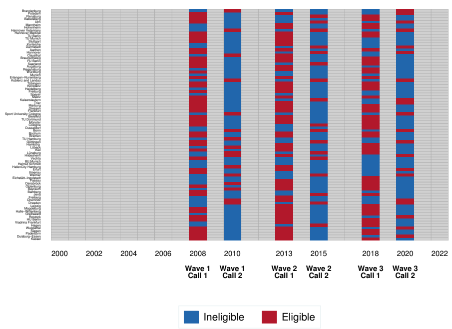


(b) Residual Retirement Probabilities

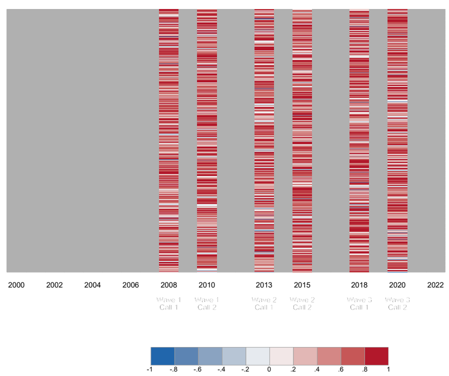


Panel B: All Universities

(c) Eligibility Status



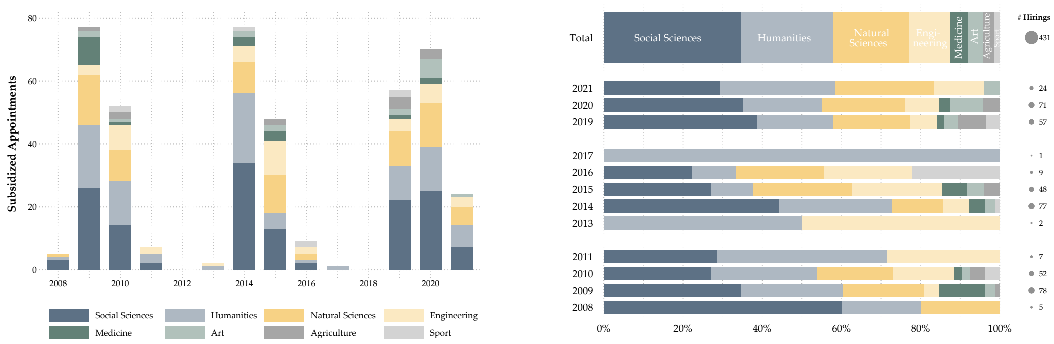
(d) Residual Retirement Probabilities



Note: Panel A of the figure illustrates the identifying variation for the University of Stuttgart. Figure 1.B.5a displays the eligibility status of the University of Stuttgart across funding periods. Figure 1.B.5b depicts residualized variation in departmental retirement probabilities for the University of Stuttgart across funding periods. Panel B of the figure illustrates the identifying variation across all universities. Figure 1.B.5c displays the eligibility status of universities across funding periods. Figure 1.B.5d depicts residualized variation in departmental retirement probabilities across funding periods, with each row representing a department within a university.

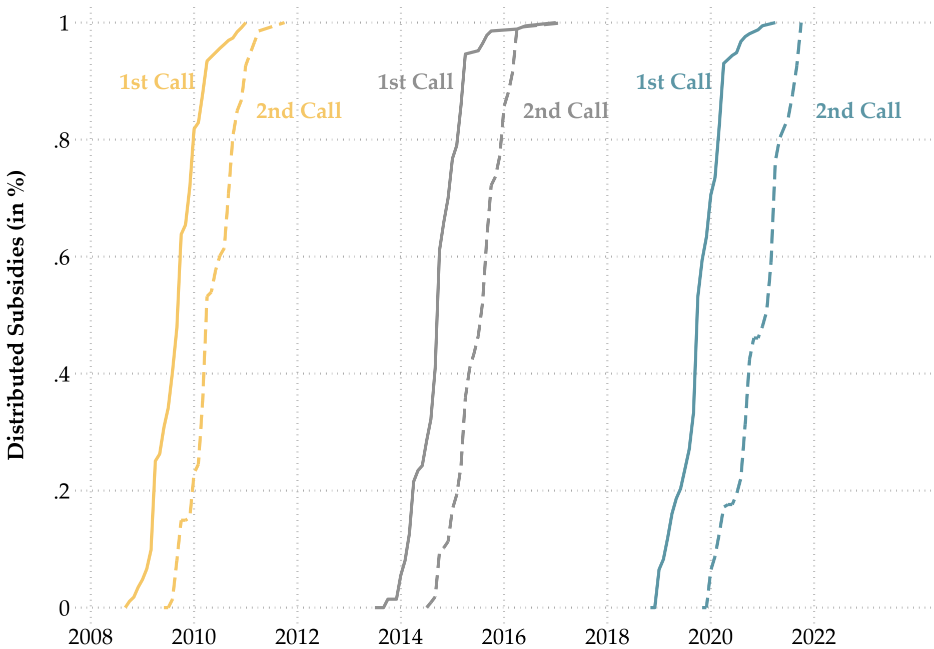
Figure 1.B.6: Affirmative Action Appointments by Faculty and Year

(a) Absolute (b) Relative



Note: The figure shows the number of affirmative action appointments by year and faculty across public universities, as detailed in Appendix Table 1.A.1. In total, the sample includes 431 subsidized appointments of women to full professor positions. Figure 1.B.6a presents the absolute number of subsidized appointments, while Figure 1.B.6b displays these numbers as a proportion of the total number of subsidized appointments per year.

Figure 1.B.7: Distributed Funds by Funding Period and Year



Note: The figure illustrates the share of funds utilized over time for each funding period. The share is calculated by summing the subsidies granted to all types of universities over time, as recorded in the Federal Government's funding portal (Bundesregierung, 2023), and dividing this total by the budgetary resources allocated for each funding period, as detailed in Table 1.3.

1.C Additional Analyses

1.C.1 Text Analysis

To explore this possibility, I conduct a text analysis on all available application documents, which I gather by systematically searching all university web pages for *Professorinnenprogramm* application documents. In total, I collect 247 documents: 143 covering eligible universities, 103 covering ineligible ones.

The analysis proceeds in two steps. First, I evaluate the semantic similarity of application documents from positively and negatively evaluated universities. In particular, I evaluate whether the tone of application documents differs between the two cases or whether they use different language to support their application. I measure semantic similarity using three measures. Subjectivity measures the degree to which a piece of text expresses personal opinions, feelings, or judgments, rather than factual information. It ranges from 0 to 1, where 0 indicates an objective, factual statement and 1 indicates a highly subjective, opinionated statement. Polarity is a measure of the sentiment expressed in a piece of text. It ranges from -1 to 1, where negative values indicate negative sentiment and positive values indicate positive sentiment. Lastly, I provide a measure of language similarity. To this end, I represent each application document as a word embedding. An embedding is a vector representation of a text body in continuous space. Application documents with similar embeddings are also likely to use similar language. To test for differences in embeddings between application documents of eligible and ineligible universities, I first retrieve the word embedding of each article using a pre-trained language model[32]. Next, I extract the first principal component across all application document embeddings. I standardize all three measures to mean zero and standard deviation one, such that a one unit increase corresponds to a one standard deviation increase of the respective measure.

To test for statistical differences along these measures, I estimate the following regression equation

$$Y_{ug} = \alpha_u + \alpha_g + \beta \text{Eligible}_{ug} + \varepsilon_{ug} \tag{1.1}$$

where Y_{ug} indicates some text metric of application document submitted by university u in funding period g. Eligible_{ug} is an indicator equaling one if

[32] In particular, I use the 'paraphrase-multilingual-MiniLM-L12-v2' language model, which paraphrases multilingual sentences and paragraphs as a 384 dimensional dense vector space.

university u is positively evaluated in funding period g. Through α_u and α_g I account for unobserved university-specific and time-specific effects.

The estimates shown in Columns (1)–(3) of Table 1.C.1 indicate that the application documents do not differ along either dimension. Across all three measures, I document a small and statistically insignificant effect, indicating that application documents from eligible and ineligible universities use similar semantics and language.

Table 1.C.1: Text Analysis of Application Documents

	Semantics			Topic Distribution	
	1st Principal Component	Subjectivity	Polarity	1st Principal Component	2nd Principal Component
	(1)	(2)	(3)	(4)	(5)
Eligible University	-0.027	0.067	0.024	0.031	-0.252
	(0.056)	(0.084)	(0.278)	(0.637)	(0.206)
Observations	134	134	134	134	134
Fixed Effects					
University	✓	✓	✓	✓	✓
Funding Period	✓	✓	✓	✓	✓

Note: This table shows estimates from regressing various text-based metrics on an indicator of university eligibility. The sample includes all publicly available application documents of the *Professorinnenprogramm*. Columns (1)–(3) consider semantic metrics as described in Section 1.C.1. Columns (4)–(5) consider the first two principal components of the topic model trained on the application documents. All specifications include university and funding period fixed effects. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Next, moving beyond semantics, I aim to analyze whether the themes of application documents differ by eligibility status. In a first step, I display the most frequently used words in the application documents in Appendix Figure 1.C.1. The size of each word is proportional to its relative frequency within the application documents. Unsurprisingly, the application documents most frequently mention 'women'. To analyze the content structure in more detail, I proceed by training a topic model on the application documents. A topic model is a statistical model designed to discover abstract topics within a collection of documents or texts. It is employed in natural language processing and machine learning to identify the underlying themes or topics prevalent in a set of documents. The goal is to automatically extract meaningful patterns and associations among words for categorizing and understanding the content of text documents. Intuitively, a topic model algorithm computes a word embedding for each document and then clusters documents close in vector space.

Figure 1.C.1: Application Documents – Wordcloud

Note: The figure displays wordclouds depicting the most frequently used words in the application documents discussed in Appendix Section 1.C.1. All application documents have been translated into English, and common stopwords have been excluded. The size of each word is proportional to its relative frequency within the documents.

I train a topic model[33] on all available application documents. After training, I extract a topic distribution for each application document along the identified topics. To compare the topic distribution of application documents by eligibility status, I extract the first two principal components of the topic distribution and use them as outcome variables in Equation (1.1).

1.C.2 Weak Instrument Considerations

It is well known that t-ratio tests over-reject when instruments are weak (Bound, Jaeger and Baker, 1995; Staiger and Stock, 1997). The discussion on dealing with potentially weak instruments revolves around two parameters: the first-stage F-statistic and the endogeneity coefficient ρ, measuring the correlation between structural and first-stage residuals. Within this framework, a high degree of endogeneity calls for a strong instrument, i.e., a high first-stage F-statistic. In contrast, ‘low’ endogeneity is reconcilable with a low first-stage F-statistic. In particular, conventional (unadjusted) IV standard errors sufficiently account for weak instruments unless endogeneity is ‘extraordinarily high’, defined as $|\rho| > .565$ (Angrist and Kolesár, 2021). However, because it might be challenging to bound ρ a priori, numerous frequentist methods exist to adjust standard errors and confidence intervals for potential inference distortions (Anderson and Rubin, 1949; Lee et al., 2022).

[33] I use the ’BERTopic’ Python module with default settings.

I address potential weak instrument concerns in two ways. First, I report 95-percent confidence intervals $[\hat{\rho}_L, \hat{\rho}_U]$ of the endogeneity parameter ρ. Appendix Tables 1.C.2 and 1.C.3 show that my specification exhibits moderate to high levels of endogeneity, exceeding the threshold of $|\rho| > .565$ when considering my main specification. The high degree of endogeneity might not be surprising given that the hiring of professors is a highly endogenous process. At the same time, the high degree of endogeneity justifies my instrumental variable approach and offers an explanation for the stark difference between OLS and 2SLS estimates observed in Tables 1.7 and 1.8.

Complementing the bounding exercise on ρ, Appendix Tables 1.C.2 and 1.C.3 reports p-values of the Anderson-Rubin F-test (Anderson and Rubin, 1949) as well as tF-adjusted standard errors (Lee et al., 2022). The procedure by Anderson and Rubin yields confidence intervals with undistorted coverage for any pair of values ρ and F. On the other hand, tF-adjusted standard errors assume a worst-case endogeneity scenario, i.e., $|\rho| = 1$, and accordingly adjust the conventional 2SLS standard errors by an adjustment factor based on the first-stage F-statistic and the considered significance level.[34] Under both procedures, my results remain significant at the 1-percent level even when considering a worst-case endogeneity scenario of $|\rho| = 1$ as assumed when computing tF-adjusted standard errors.

[34] Both procedures yield correct coverage under arbitrarily weak instruments; however, the expected length of the Anderson-Rubin confidence interval is infinite, while the corresponding tF interval is finite (Lee et al., 2022).

Table 1.C.2: Change in Hiring Patterns – Weak IV

	Junior Faculty		Ph.D. Students	
	Ass. Professor	Post-Doc	Overall	Home
	(1)	(2)	(3)	(4)
Panel A: 2SLS Estimate				
Female Hiring	0.034	-0.006	0.098***	0.083***
	(0.121)	(0.109)	(0.028)	(0.029)
Observations	147,591	147,591	147,591	147,591
Panel B: Weak IV Considerations				
Endogeneity Parameter ρ				
$\max\{\lvert\hat{\rho}_L\rvert, \lvert\hat{\rho}_U\rvert\}$	0.589	0.612	0.472	0.491
Anderson-Rubin Inference				
p-value	0.831	0.764	0.032	0.021
tF-adjusted Standard Errors				
5-percent Significance	(0.146)	(0.142)	(0.034)	(0.033)
1-percent Significance	(0.161)	(0.178)	(0.041)	(0.041)
Fixed Effects				
Department	✓	✓	✓	✓
Field × Year	✓	✓	✓	✓
University × Year	✓	✓	✓	✓

Note: Panel A displays 2SLS estimates based on Equation (IV2). Panel B reports three measures to discover and account for the presence of weak instruments. First, I report a bound on the endogeneity parameter ρ by following Online Appendix Section A.8.3 of Lee et al. (2022). In particular, I use 95-percent tF confidence interval endpoints $[\hat{\beta}_L, \hat{\beta}_U]$ to compute the endpoints $\rho(\hat{\beta}_L)$ and $\rho(\hat{\beta}_U)$. Second, I report p-values of the Anderson-Rubin F-test of endogenous regressors (Anderson and Rubin, 1949). Third, I construct tF-adjusted standard errors for 5-percent and 1-percent significance levels using first-stage F-statistics and critical values provided in Lee et al. (2022). Robust standard errors, clustered by department and event, are reported in parentheses. Significance levels are indicated as follows: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 1.C.3: Change in Collaboration Patterns – Weak IV

	All	Women	Men	Men by Seniority (Quartiles)			
				Q1	Q2	Q3	Q4
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A: 2SLS Estimate							
Female Hiring	0.019	0.011	0.028	0.074***	0.051**	0.011	-0.004
	(0.025)	(0.023)	(0.025)	(0.027)	(0.024)	(0.023)	(0.026)
Observations	147,591	147,591	147,591	147,591	147,591	147,591	147,591
Panel B: Weak IV Considerations							
Endogeneity Parameter ρ							
$\max\{\lvert\hat{\rho}_L\rvert, \lvert\hat{\rho}_U\rvert\}$	0.464	0.764	0.452	0.552	0.489	0.452	0.689
Anderson-Rubin Inference							
p-value	0.214	0.343	0.151	0.051	0.073	0.907	0.858
tF-adjusted Standard Errors							
5-percent Significance	(0.034)	(0.032)	(0.035)	(0.038)	(0.033)	(0.033)	(0.036)
1-percent Significance	(0.045)	(0.041)	(0.044)	(0.048)	(0.045)	(0.045)	(0.047)
Fixed Effects							
Department	✓	✓	✓	✓	✓	✓	✓
Field × Year	✓	✓	✓	✓	✓	✓	✓
University × Year	✓	✓	✓	✓	✓	✓	✓

Note: Panel A displays 2SLS estimates based on Equation (IV2). Panel B reports three measures to discover and account for the presence of weak instruments. First, I report a bound on the endogeneity parameter ρ by following Online Appendix Section A.8.3 of Lee et al. (2022). In particular, I use 95-percent tF confidence interval endpoints $[\hat{\beta}_L, \hat{\beta}_U]$ to compute the endpoints $\rho(\hat{\beta}_L)$ and $\rho(\hat{\beta}_U)$. Second, I report p-values of the Anderson-Rubin F-test of endogenous regressors (Anderson and Rubin, 1949). Third, I construct tF-adjusted standard errors for 5-percent and 1-percent significance levels using first-stage F-statistics and critical values provided in Lee et al. (2022). Robust standard errors, clustered by department and event, are reported in parentheses. Significance levels are indicated as follows: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

1.C.3 Quantifying the Impact of Affirmative Action

Model Framework Consider a model of academic fields $j \in J$ observed over two time periods, $t \in \{0, 1\}$. In the initial period ($t = 0$), each field has T_{j0} available positions and hires F_{j0} women. Over time, the total number of positions evolves according to a field-specific factor γ_j, so that in period $t = 1$ the total number of positions in field j is

$$T_{j1} = (1 + \gamma_j)T_{j0}.$$

Similarly, the model accounts for field-specific female hiring trends, captured by γ_j^F, such that

$$F_{j1} = (1 + \gamma_j^F)F_{j0}.$$

Modelling Affirmative Action In period $t = 1$, an affirmative action (AA) initiative is introduced. This policy provides subsidies for female hires but does not increase the total number of available positions, so that T_{j1} remains unchanged. Let F_{j1}^{AA} denote the number of AA-funded female hires in field j. AA hires constitute only a fraction of the female hires that would have occurred in the absence of the policy, $F_{j1}^{AA} < (1 + \gamma_j^F)F_{j0}$. The total number and share of female hires in period $t = 1$ can thus be written as

$$F_{j1} = (1 + \gamma_j^F)F_{j0} + \pi\, F_{j1}^{AA},$$

and

$$f_{j1} = \frac{1 + \gamma_j^F}{1 + \gamma_j} f_{j0} + \pi\, f_{j1}^{AA}.$$

Here $\pi \in [0, 1]$ captures the degree to which AA-funded hires substitute for other hires. Two extreme cases illustrate this interpretation. If $\pi = 0$, each AA hire fully replaces a woman who would have been hired anyway, so the total number of female hires remains unchanged. If $\pi = 1$, each AA hire displaces an otherwise male hire.

Candidate Pool Constraint Thus far, the models assume that there is an unlimited candidate pool in each field. However, in practice, the number of suitable candidates is likely to be constrained – for instance, due to quality thresholds. To account for this, each field is assumed to have a time-variant pool of suitable candidates, with C_{jt}^F and C_{jt}^M denoting the numbers of available and suitable female and male candidates at time t, respectively. These pools are always sufficient to fill the available positions, i.e.,

$$C_{jt}^{F} \geq F_{jt} \quad \text{and} \quad C_{jt}^{M} \geq M_{jt} \quad \forall j, t.$$

In period $t = 1$, the maximum possible additional female hires attributable to AA are constrained by the number of available female candidates. Specifically, AA hires cannot exceed

$$\bar{F}_{j1}^{AA} = C_{j1}^{F} - (1 + \gamma_j^F) F_{j0}.$$

If $\pi F_{j1}^{AA} > \bar{F}_{j1}^{AA}$, all AA hires exceeding $\bar{F}_{j1}^{AA}$ must substitute for women who would have been hired anyway. To reflect this constraint, I define the effective conversion rate $\bar{\pi}_j$ as

$$\bar{\pi}_j \equiv \pi \frac{\bar{F}_{j1}^{AA}}{F_{j1}^{AA}}.$$

which scales π by the proportion of AA hires that do not exceed the constraint.

Correspondingly, the change in the share of female hires from period $t = 0$ to $t = 1$ is given by

$$\Delta f_j = f_{j1} - f_{j0} = \underbrace{\left[\frac{1 + \gamma_j^F}{1 + \gamma_j} - 1 \right] f_{j0}}_{\equiv \alpha_j} + \bar{\pi}_j f_{j1}^{AA}.$$

The first term, α_j, represents the baseline change in the female hiring share driven by the differential growth rates of female versus overall hires. The second term, $\bar{\pi}_j f_{j1}^{AA}$, captures the additional increase in female hires attributable to AA-funded positions.

1.D Additional Data

1.D.1 Alternative Retirement Measures

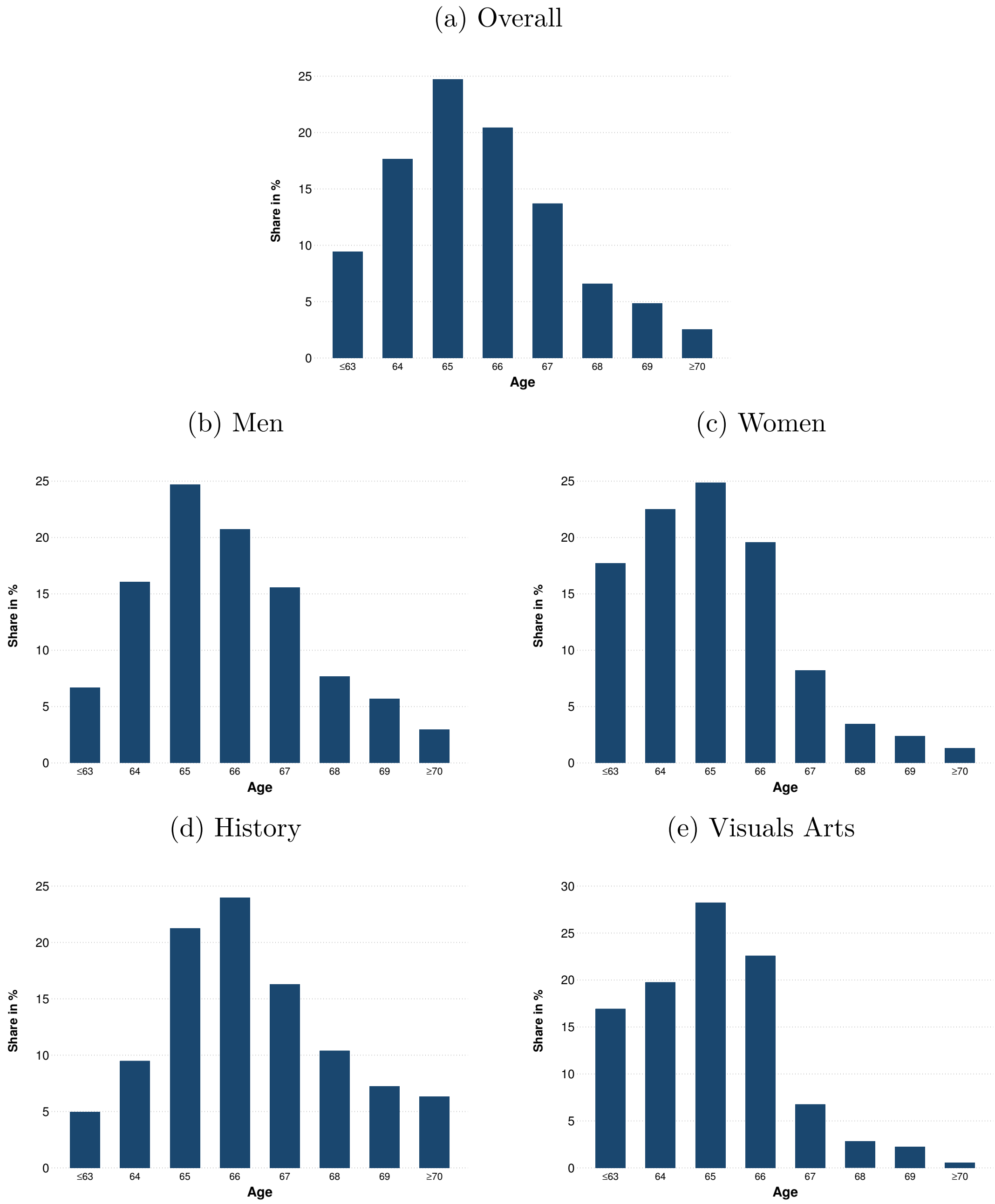
In Section 1.5.1, I show that my findings remain robust when using alternative retirement measures. Specifically, I construct binary indicators based on whether any department member reaches a certain age threshold, such as the statutory retirement age. While these binary measures yield estimates of similar size to the continuous measure, they are less statistically significant. I attribute this to two factors. First, my continuous approach accounts for cases where multiple professors in a department are nearing retirement by aggregating individual retirement probabilities. A binary model, on the other hand, cannot differentiate between departments with several impending retirements and those with only one. Second, a binary measure discards variation in retirement timing that the continuous measure is able to capture.

In particular, a binary approach would be reasonable if all professors retired precisely at the statutory retirement age. However, in Germany, professors have considerable flexibility in deciding when to retire. As shown in Appendix Figure 1.D.1a, most professors retire at the age of 65, around 25%.[35] Besides this, retirement ages vary widely, with the distribution being notably right-skewed: only around 25% of professors retire before 65, while around 50% retire after. Therefore, for example, a binary indicator with a cutoff at 65 would misclassify about 25% of retirement decisions as false negatives.

Retirement timing also varies across other dimensions. Appendix Figures 1.D.1b and 1.D.1c indicate that women retire substantially earlier than men. Similarly, Appendix Figures 1.D.1d and 1.D.1e reveal substantial differences across academic disciplines: historians tend to postpone retirement as long as possible, while art professors often retire early. A binary retirement indicator fails to account for these nuances, imposing an overly simplistic model on the data-generating process. In contrast, the continuous measure captures this variation, leading to higher statistical power in subsequent analyses. Therefore, the main analysis is based on this measure.

[35] The statutory retirement age has gradually increased, starting at 65 for individuals born before 1946 and reaching 67 for those born after 1964. Most professors retiring between 2000 and 2010 were still subject to the 65 or 66 statutory retirement age.

Figure 1.D.1: Retirement Probability Distributions



Note: The figure illustrates the retirement probability distributions of full professors across various subgroups. Figure 1.D.1a displays the overall share. Figures 1.D.1b and 1.D.1c break down this data by gender, while Figures 1.D.1d and 1.D.1e distinguishes between history and arts departments. All figures are based on the population of professors who retired between 2000 and 2010 and were employed at public German universities. The data are sourced from the Hochschulpersonalstatistik, as detailed in Section 1.3.1.

1.D.2 Matching Research Output

The ‘Hochschullehrerverzeichnis’ is an annual directory that lists all German university professors along with their affiliations and descriptions of their disciplines (Hochschulverband, 2002–2022). Appendix Figure 1.D.2 shows an example excerpt of the first entry of the 2008 HLV, which comprises 750 pages of similar layout. I first digitized all the directories covering the years 2002–2022 using optical character recognition. The first entry shown in Appendix Figure 1.D.2 provides an overview of the typical structure of each entry:

> **Aach**, Til; Dr.-Ing., Prof. RWTH Aachen; *Signalverarbeitung u. Prozeßrechentechnik, Bildverarbeitung, med. Bildverarbeitung, Mustererkennung*; di: RWTH, Fak. f. Elektrotechnik u. Informationstechnik, Lst. für Bildverarbeitung, Sommerfeldstraße, 52056 Aachen, T: (0241) 8027860, F: 8022200, til.aach@lfb.rwth-aachen.de; www.isip.uni-
> luebeck.de

This entry lists the name (Aach, Til), title (Dr.-Ing.), position (Prof.), institution (RWTH Aachen), academic discipline (Signalverarbeitung u. Prozeßrechentechnik, Bildverarbeitung, Mustererkennung), and contact information (di: RWTH, Fak. f. Elektrotechnik u. Informationstechnik, Lst. für Bildverarbeitung, Sommerfeldstraße, [...]).

The objective is to extract the position, department, and institution from each entry. To achieve this, I utilize classification algorithms trained using 1,000 randomly selected and manually classified entries. For each algorithm, I manually define a set of categories to choose from. For institutions, the potential targets include all public universities in Germany as listed in Appendix Tables 1.A.1, while the set of potential departments corresponds to those listed in the HPS as listed in Appendix Table 1.A.2. If a university does not have a specific department – for example, if a university lacks an art history department – the set of potential departments is limited to those that are actually present at that university (as identified through the HPS). The position categories include full professor, assistant professor, and other professor. In the latter category, I classify emeritus and honorary professors. If the algorithm assigns multiple positions, the highest one is assigned. For instance, in the example provided, the algorithms correctly infer the position (full professor), department (computer science), and institution (RWTH Aachen).

In total, I classify entries for 1.2 million individuals, averaging approximately 60,000 entries per year. Next, I match these entries over time by identifying individuals with the same name, department, and university. If a direct match is not found, I narrow the criteria to just name and department to account for potential changes in affiliation. Throughout, I retain only individuals with unique matches. In the next step, I combine this panel with research output data obtained from OpenAlex. For each professor identified in the 'Hochschullehrerverzeichnis', I search for researchers with the same name and affiliation in the OpenAlex data. In cases of multiple matches, I manually verify and assign matches by comparing publication records. By aggregating the resulting panel by department and year, I can track departmental research output and collaboration patterns over time.

Figure 1.D.2: Exemplary Excerpt from the 'Hochschullehrerverzeichnis'

Note: The figure provides an example excerpt from the *Hochschullehrerverzeichnis* in 2008 (Hochschulverband, 2002–2022). Each entry lists the professor's name, university, and field of specialization. The pages from each volume are digitized using optical character recognition. The entries are matched over time by identifying individuals with the same name, department, and university. If a direct match cannot be established, the criteria are narrowed to name and department to account for potential changes in affiliation. For further details, refer to Appendix Section 1.D.2.

1.D.3 Measuring Changes in Research Direction

To measure changes in research direction, I first construct department-specific topic distributions using a topic model – an unsupervised machine learning technique designed to uncover latent themes in textual data. This approach allows each text to be represented by a distribution of topics, providing a more nuanced and detailed representation compared to binary classifications. I utilize the BERTopic algorithm developed by Grootendorst (2022). Appendix Figure 1.D.3 provides an overview of the steps involved in this analysis.

I begin by collecting all abstracts of papers published by professors working at Germany's public universities during the sample period. For each academic discipline, I train a separate topic model to identify and describe the topics within this field. For each field f, I randomly select 10,000 papers authored by researchers in field f and published between 2000 and 2022. To ensure equal representation of abstracts across years, I stratify the randomization process by year. The topic model is then trained on the entire set of abstracts from all years to ensure a consistent and time-invariant set of topics for each field. Notably, results from models trained on different datasets, such as annual subsets, are inherently not comparable, as explained below.

The topic model algorithm involves two key steps. First, it represents each abstract as a dense vector in continuous space, known as an embedding. For this purpose, I utilize the pre-trained multilingual language model 'paraphrase-multilingual-MiniLM-L12-v2' (Reimers and Gurevych, 2019). This sentence-transformer maps text to a 384-dimensional dense vector space, supports over 50 languages, and considers the context in which words appear within sentences. Second, the topic model algorithm clusters embeddings that are sufficiently close to each other while being distinctly separated from other groups of embeddings. Each cluster represents a distinct topic. The number of topics, or clusters, is determined by setting hyperparameters that define what constitutes sufficiently close and sufficiently far distances between embeddings. For technical details, interested readers are referred to Grootendorst (2022). To objectively select these hyper-parameters, I perform cross-validation to find the set that maximizes the topic model's coherence score, a metric used to assess the quality and interpretability of the topics generated by the model (Mimno et al., 2011). This process results in a representative topic distribution for each field, denoted as $\mathbf{X}(f)$, and a covariance matrix $\mathbb{V}_{\mathbf{X}(f)}$ describing the correlation of topics within a field. For instance, papers covering labor economics are more likely to touch on thematic areas from public economics rather than from monetary economics. I utilize this substitutability information when calculating how topic distributions change over time. Lastly, I assign labels to each topic. This step is purely for human understanding and does not influence how the model assigns topics. Initially, I identify a collection of keywords and documents that most accurately depict each topic using a term frequency–inverse document frequency (TF-IDF) approach, which highlights their significance. These selected keywords and documents are then fed into OpenAI's ChatGPT-4 (OpenAI, 2024), which I prompted to generate a concise description of the topic in three words.

Once the topic model is trained, it can be used to predict the topic distribution of previously unseen abstracts. The model converts each provided abstract into an embedding and assesses its alignment within the clusters of identified topics from the training phase. This process enables us to predict the topic distribution across all academic work published during the sample period. To ensure each professor's equal weighting in computing departmental topic distributions, I first average the paper-specific topic distributions by author and year, resulting in annual topic profiles for each professor. Subsequently, these individual profiles are averaged by department and year to produce departmental topic distributions.

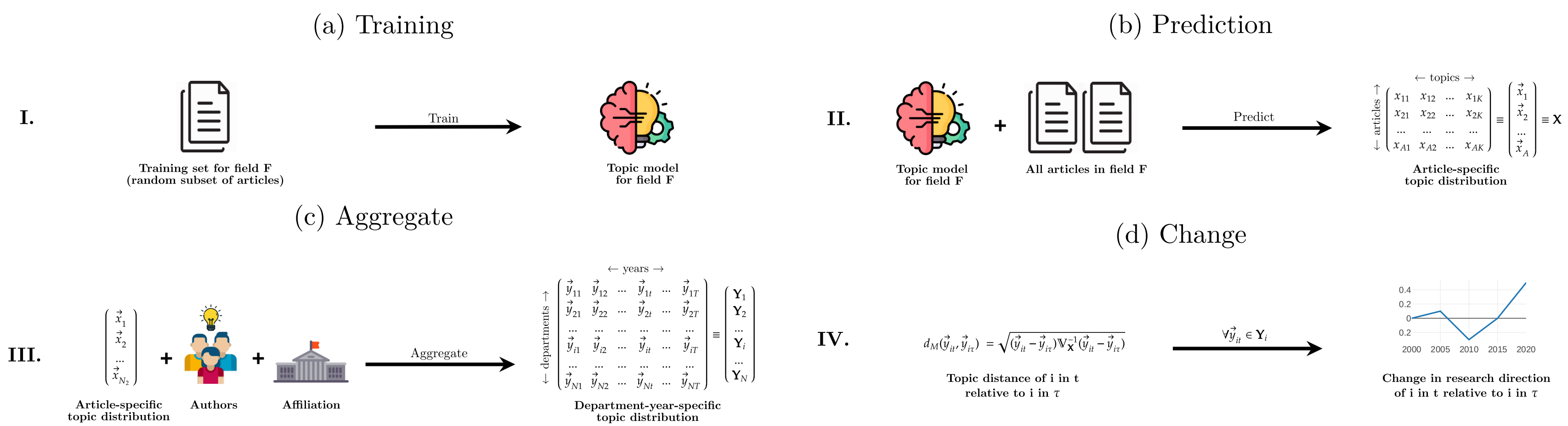
Next, I analyze whether these average topic distributions differ between departments that appoint a female professor and those that do not. I measure changes in topic distributions within departments across years via the Mahalanobis distance:

$$d_M \equiv d_M(\vec{y}_{it}, \vec{y}_{i\tau}) = \sqrt{(\vec{y}_{it} - \vec{y}_{i\tau})\mathbb{V}_{\mathbf{X}}^{-1}(\vec{y}_{it} - \vec{y}_{i\tau})}$$

Here, the vectors $\vec{y}_{it}$ and $\vec{y}_{i\tau}$, represent the topic distribution of department i in year t and the pre-funding period $\tau \equiv \tau(g)$, respectively. The Mahalanobis distance assumes that these vectors are drawn from some distribution $\mathbf{X}$ on $\mathbb{R}^K$ with covariance matrix $\mathbb{V}_{\mathbf{X}}$, which I replace by the sample analogs obtained from the field-specific topic models.

A unit increase in d_M indicates that department i's topic distribution in year t deviates by one standard deviation from its distribution in τ. The measure is zero if the topic distribution remains constant over time and diverges quadratically to infinity as the distance between topic distributions increases. Unlike other measures, the Mahalanobis distance allows accounting for different degrees of substitutability between topics by weighting the distance using the inverse of the covariance matrix, $\mathbb{V}_{\mathbf{X}}^{-1}$. For instance, shifts from labor economics to public economics are weighted less compared to shifts from labor economics to monetary economics in the distance calculation. I use d_M as the outcome measure in the regression framework described in Section 1.4.

Figure 1.D.3: Constructing Field-Specific Topic Distribution



Note: The figure schematically describes the construction of field-specific topic distributions as described in Appendix Section 1.D.3. First, abstracts from a stratified sample of 10,000 papers per field are used to train separate topic models for each academic discipline, ensuring consistent topic identification across years (Appendix Figure 1.D.3a). The trained models are then applied to predict the topic distribution of unseen abstracts (Appendix Figure 1.D.3b). To calculate departmental topic distributions, the topic profiles for each professor are averaged by year and then further aggregated at the department level (Appendix Figure 1.D.3c). Finally, changes in departmental topic distributions relative to the pre-funding period are measured using the Mahalanobis distance (Appendix Figure 1.D.3d).

Bibliography

Anderson, Theodore W, and Herman Rubin. 1949. "Estimation of the Parameters of a single Equation in a Complete System of Stochastic Equations." *Annals of Mathematical Statistics*, 20(1): 46–63.

Angrist, Joshua, and Michal Kolesár. 2021. "One Instrument to Rule Them All: The Bias and Coverage of Just-ID IV." *NBER Working Paper*, 29417.

Antecol, Heather, Kelly Bedard, and Jenna Stearns. 2018. "Equal but Inequitable: Who Benefits from Gender-neutral Tenure Clock Stopping Policies?" *American Economic Review*, 108(9): 2420–41.

Bagues, Manuel, Mauro Sylos-Labini, and Natalia Zinovyeva. 2017. "Does the Gender Composition of Scientific Committees Matter?" *American Economic Review*, 107(4): 1207–38.

Blau, Francine D, Janet M Currie, Rachel TA Croson, and Donna K Ginther. 2010. "Can Mentoring Help Female Assistant Professors? Interim Results from a Randomized Trial." *American Economic Review*, 100(2): 348–52.

Bound, John, David A Jaeger, and Regina M Baker. 1995. "Problems with Instrumental Variables Estimation when the Correlation between the Instruments and the Endogenous Explanatory Variable is Weak." *Journal of the American Statistical Association*, 90(430): 443–450.

Bundesregierung, Die. 2023. "Förderkatalog." *URL: https://foerderportal.bund.de/foekat/jsp/SucheAction.do?actionMode=searchmask*, Accessed: 2023-10-24.

Carnes, Molly, Patricia G Devine, Linda Baier Manwell, Angela Byars-Winston, Eve Fine, Cecilia E Ford, Patrick Forscher, Carol Isaac, Anna Kaatz, Wairimu Magua, et al. 2015. "Effect of an Intervention to Break the Gender Bias Habit for Faculty at One Institution: A Cluster Randomized, Controlled Trial." *Academic Medicine: Journal of the Association of American Medical Colleges*, 90(2): 221.

Destatis, Federal Statistical Office of Germany. 2018. "Hochschulpersonalstatistik." Administrative data. Information on how to request the data can be found at `https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Bildung-Forschung-Kultur/Hochschulen/Methoden/Erlaeuterungen/hochschulen.html`.

Devine, Patricia G, Patrick S Forscher, William TL Cox, Anna Kaatz, Jennifer Sheridan, and Molly Carnes. 2017. "A Gender Bias Habit-Breaking Intervention Led to Increased Hiring of Female Faculty in STEMM Departments." *Journal of Experimental Social Psychology*, 73: 211–215.

Eberhard Karls Universität Tübingen. 2019. "Professur (W3) für Kunstgeschichte." *URL: https://uni-tuebingen.de/fakultaeten/philosophische-fakultaet/fakultaet/ausschreibungenstellenangebote/professuren/w3/*, Accessed: 2024-09-19.

Gemeinsame Wissenschaftskonferenz. 2013–2022. "Presse-Archiv." *URL: https://www.bmbf.de/bmbf/shareddocs/bekanntmachungen/de/2008/03/327_bekanntmachung.html#searchFacets*, Accessed: 2024-09-19.

Ginther, Donna K, Janet M Currie, Francine D Blau, and Rachel TA Croson. 2020. "Can Mentoring Help Female Assistant Professors in Economics? An Evaluation by Randomized Trial." Vol. 110, 205–09.

Grootendorst, Maarten. 2022. "BERTopic: Neural Topic Modeling With a Class-based TF-IDF Procedure." *arXiv Pre-print*, 2203.05794.

Hochschulverband, Deutscher, ed. 2002–2022. *Hochschullehrer Verzeichnis – Universitäten Deutschland.* De Gruyter Saur.

Lee, David S, Justin McCrary, Marcelo J Moreira, and Jack Porter. 2022. "Valid t-ratio Inference for IV." *American Economic Review*, 112(10): 3260–90.

Mimno, David, Hanna Wallach, Edmund Talley, Miriam Leenders, and Andrew McCallum. 2011. "Optimizing Semantic Coherence in Topic Models." 262–272.

Ministerium für Wissenschaft, Forschung und Kunst Baden-Württemberg. 2020. "Stellenplan für Beamtinnen und Beamte im Landesbetrieb Universität Mannheim." *URL: https://www.statistik-bw.de/shp/2020-21/pages/Epl14/ST/epl14_1420_st.pdf*, Accessed: 2024-09-19.

OpenAI. 2024. "ChatGPT-4: Conversational AI powered by OpenAI." Accessed: 2024-06-27.

Reimers, Nils, and Iryna Gurevych. 2019. “Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks.” Association for Computational Linguistics.

Staiger, Douglas, and James H Stock. 1997. “Instrumental Variables Regression with Weak Instruments.” *Econometrica*, 557–586.

Chapter 2

Leveling the Playing Field: Knowledge Production in the Digital Age

(joint with Jens Oehlen)

2.1 Introduction

The creation of new ideas is the central pillar of modern economic growth (Romer, 1990; Jones, 1995). New insights are generated using existing knowledge (Mokyr, 2011) and, in particular, knowledge created by scientific 'giants' (Azoulay, Graff Zivin and Wang, 2010; Iaria, Schwarz and Waldinger, 2018) which ultimately fuels industrial innovation (Ahmadpoor and Jones, 2017; Bryan and Ozcan, 2021). With the rise of the internet, the marginal cost of distributing scientific articles has dramatically declined. However, access to the latest research is still severely restricted. Only about 20 percent of peer-reviewed academic journals are published under open access – the practice of providing online access to scientific information

* We are grateful to Antonio Ciccone, Matthew Gentzkow, Abhishek Nagaraj, Torsten Persson, Giuseppe Sorrenti, Carolyn Stein, David Strömberg and Ulf Zölitz, for helpful comments and encouragement as well as César Hidalgo and Megan MacGarvie for insightful discussions. We thank audiences at the NBER SI SSF 2023, Munich SI 2023, 7th IZA Workshop on the Economics of Education, WICK #10 at Collegio Carlo Alberto, University of Mannheim, Stanford University, and Stockholm University for comments. Jens Oehlen gratefully acknowledges funding from the Tom Hedelius foundation. The authors declare that we have no relevant or material financial interests that relate to the research described in this paper. All errors are our own.

free of charge.[1] The remaining 80% of journals are only available behind – often very expensive – paywalls.

To what extent do access restrictions inhibit further knowledge production? Despite potentially grave impacts, rigorous evidence on this question is surprisingly scant. The key reason is that researchers' journal-access is typically tied to the academic institutions they are associated with. Hence, any comparison across researchers with different journal-access conditions would be subject to endogeneity.

In this paper, we overcome this challenge by focusing on a natural experiment. We study how the consumption and production of new scientific insights are affected when vast amounts of existing knowledge become freely available through *Sci-Hub*. Sci-Hub is an online media tool developed in Almaty, Kazakhstan, that offers free access to most scientific articles worldwide. Launched in 2011, the website has garnered a global audience with roughly 3 million paper downloads per day.[2] However, Sci-Hub traffic across the world is not randomly distributed. We, therefore, isolate quasi-exogenous variation through social networks using an instrumented difference-in-differences framework. Akin to papers in the existing media literature (Enikolopov, Makarin and Petrova, 2020; Müller and Schwarz, 2023), we argue that social connections often drive technology adoption. Sci-Hub was created in Almaty, Kazakhstan, without large marketing budgets. Hence, knowledge of its existence spread mainly via social networks, leading to increased website traffic particularly in regions with stronger social ties to Almaty. For one such network, we have high-quality data: anonymized friendship links measured by Facebook. This allows us to examine the effect of Sci-Hub on knowledge creation under the identifying assumption that scientific outcomes in regions with different degrees of social connectedness to Almaty would have followed parallel trends without the rise of Sci-Hub.

Our empirical analysis relies primarily on two key datasets. The first consists of server log files from Sci-Hub, covering the period from its launch in 2011 through 2017. These logs capture approximately 300 million access requests worldwide, with each entry recording the timestamp, article accessed, and—crucially—the geolocation of the user's IP address. Using this information, we construct a

[1] Own calculations based on Scopus data from 2020.
[2] Source: sci-hub.se/stats, late 2022. For comparison, JSTOR counted approximately 600,000 daily downloads in 2019 (source: about.jstor.org/librarians/journals/, accessed on 14th of January 2023). PubMed received approximately 3 million searches and 2.5 million unique visitors per day in 2017 (Fiorini, Lipman and Lu, 2017).

dynamic, global measure of Sci-Hub usage intensity at the sub-national level. The second dataset comes from OpenAlex, the successor to the Microsoft Academic Graph. OpenAlex provides comprehensive global data on scientific publications. We use it to build a panel dataset of subnational regions, capturing both citations to closed-access papers and the geographical distribution of newly authored scientific articles from 2000 to 2022.

Leveraging our large data, we start by documenting four facts. First, we show that monetary restrictions are pervasive, yet particularly binding for top-quality journals. On average, only 20% of journals operate under open access regimes and the figure drops to 9% in the top percentile of all journals, as measured by impact factor. If scientists had bulk access through their libraries and institutions, access restrictions would not significantly hinder the spread and production of scientific knowledge. However, our second fact speaks against an equal distribution of access. We find that institutions in less developed regions are much less likely to have JSTOR subscriptions, a proxy for institutional bulk access.[3] Third, the unequal distribution of access does not simply mimic an unequal distribution of demand for high-quality knowledge. Our analysis of the freely available articles on Sci-Hub yields that most downloads per researcher stem predominantly from developing and emerging countries. Differences in demand are particularly large for high-quality journals: researchers from low-income countries are four times more likely to download papers from the top 1% of journals than researchers in high-income countries. Finally, we document significant differences in the production of high-quality research between less and highly-developed regions. Among top journals, close to 90% of papers are written by authors based in developed countries, while the share is reduced to approximately 50% at below-median-quality journals. Taken together, these empirical patterns motivate the question of whether and to what extent access restrictions *cause* the unequal distribution of high-quality knowledge production.

Next, to initiate our causal analysis, we demonstrate that social connectedness to the Almaty region is a strong predictor for Sci-Hub usage. An increase in connectedness to Almaty by 1% is associated with a 0.34% higher Sci-Hub traffic with an F-statistic of approximately 40. We conduct several tests on the validity of the identification strategy. First, we show that social .connectedness is not associated with differential trends in scientific outcomes in the ten years prior to the launch of Sci-Hub. Second, we run horse races with connectedness to

[3] JSTOR includes access to over 2800 academic journals (https://about.jstor.org/librarians/journals/, 2024)

major cities in countries neighboring Kazakhstan. We consistently find that Almaty is a strong predictor of Sci-Hub traffic, whereas other regions show no or slightly negative correlation. Third, we run placebo regressions using all other subnational regions for which Facebook provides data. In this exercise, again, Almaty emerges as a robust predictor, alleviating concerns of Facebook connections per se predicting Sci-Hub take-up. Fourth, the same picture emerges when estimating placebo reduced-form equations: connectedness to Almaty predicts changes in scientific outcomes after 2011, whereas connectedness to other regions does not. Throughout our analyses, we control for a host of covariates, including year-by-country fixed effects and subnational fixed effects. Hence, all identifying variation is the differential impact of connectedness on subnational regions within a country over time.

Following the platform's launch, regions with higher Sci-Hub traffic began referencing more paywalled papers. Doubling Sci-Hub traffic leads to a five-percentage-point increase in the share of references to closed-access publications (+7.4%). We show that the largest increase in references accrues to papers published most recently and in higher-ranked journals. Notably, we estimate decreases in references to low-quality journals. This is consistent with the theory that open access enables scientists to screen papers based on their complete merit rather than relying solely on titles and abstracts (McCabe and Snyder, 2021). This suggests two key benefits of the platform. Sci-Hub enables scientists to read and reference significantly more frontier research, which would not have been possible in its absence. At the same time, the informational value of citations has increased since researchers cite fewer papers 'unseen'. This is particularly important in light of recent research documenting how citations as a performance metric aid institutions in hiring and promotion of scientists (Hager, Schwarz and Waldinger, 2023).

Lastly, we investigate the potentially beneficial effects of Sci-Hub on follow-on research. Using the same strategy outlined before, we demonstrate that regions with greater connectedness to Almaty not only reference more high-quality works, but also receive more citations themselves. Comparing publications published in 2010 versus 2015, papers from a region with twice as many friendship links as another see a differential increase in citations of almost 10%. However, we do not find that these papers are published in relatively higher-ranking journals nor that the research topic distribution shifts toward the frontier. These findings suggest that quality improvements likely take more time to manifest in these other dimensions. Alternatively, gatekeeping mechanisms may be at play, where editors have yet to recognize the enhanced quality of work, allowing for more publication success in higher-ranking journals. Finally, we do not observe evidence for greater spillovers of scientific insights to industry use. Yet, we remain cautious in drawing firm conclusions from these results because the underlying patent data may not capture global innovation activities accurately.

We contribute to several strands of literature. First, we add to studies on the economics of science. Much of the earlier work in this field has focused on understanding the academic publishing industry more broadly (McCabe, 2002; Bergstrom and Bergstrom, 2004; Jeon and Menicucci, 2006) including the role of open access journals (McCabe and Snyder, 2005). In recent years, the literature has become more empirical and examined how research quantity and quality are affected by peers (Waldinger, 2012), intellectual property rights (Williams, 2013; Murray et al., 2016; Biasi and Moser, 2021), international cooperation (Iaria, Schwarz and Waldinger, 2018; Yin et al., 2021; Jia et al., 2022), income inequality (Agarwal and Gaule, 2020) and competition (Hill and Stein, 2021). We add to this literature by examining the effect of a key pillar of knowledge creation: access to previous knowledge.

While we are not the first to study the relevance of open access, most prior empirical research on open access has focused on the effects on specific journals or papers rather than on *researchers*.[4] Moreover, the large majority of papers do not rely on (quasi-)experimental variation. Notable exceptions are Davis et al. (2008), and Davis (2011), who vary open access status for specific papers experimentally. They find that open access papers gain more views and downloads, but not citations.[5] McCabe and Snyder (2014) use a difference-in-differences design with journal-level variation and find increases in citations of approximately 8% when journals move from paid to open access.[6] However, their conclusions are drawn from a sample of journals mostly publishing work in ecology, botany, and biology. The paper closest to our work, Bryan and Ozcan (2021), shows that open access mandates imposed by the National Institutes of Health (NIH) significantly

[4] For a systematic review, see Langham-Putrow, Bakker and Riegelman (2021).

[5] The absence of effects on citations is likely a result of the selected study periods. In Davis et al. (2008) citations were measured only one year after publication, leaving only a very limited time period for realization of citation differences. In Davis (2011), on the other hand, the control papers were moved from closed to open access within one year in 19 of the 20 participating journals.

[6] Consistent with our empirical results, open access decreased citations to journals of lower quality. In a follow-up study, McCabe and Snyder (2021) investigate this seemingly surprising result more closely. They argue that some scientists previously cited closed-access publications based on abstract inspection only. Once journals moved to open access, closer inspection of actual content likely prevented such “cites unseen” citations.

increased industry-use of biomedical academic research. However, they do not find an effect on scientific citations. A likely cause for these heterogeneous results is differences across scientific fields. In our analysis across all fields, we find an average impact of open access on follow-on science.[7] Additionally, we advance the scope and quality of existing evidence by focusing on a global natural experiment with long time horizons.

Finally, we add to the literature studying the effects of media. Initially documenting the broader effects of specific technologies such as radio (Strömberg, 2004; Yanagizawa-Drott, 2014; Adena et al., 2015), TV (Gentzkow, 2006; DellaVigna and Kaplan, 2007; Enikolopov, Petrova and Zhuravskaya, 2011; Durante, Pinotti and Tesei, 2019) and the spread of the internet (Falck, Gold and Heblich, 2014; Guriev, Melnikov and Zhuravskaya, 2021), more recent work has focused on specific digital tools such as Twitter (Müller and Schwarz, 2023; Cagé et al., 2022), Facebook (Müller and Schwarz, 2021), VKontakte (Enikolopov, Makarin and Petrova, 2020; Bursztyn et al., 2019) or Craigslist (Seamans and Zhu, 2014; Djourelova, Durante and Martin, 2021) with a tremendous variety of different outcomes. Here, we focus on a novel digital platform, an academic file-sharing website, that is widely used across the world. We are unaware of other studies documenting the causal effects of digital media on scientific outcomes.

The paper is structured as follows. First, we give a brief account of the background. Then, we outline the data construction in Section 2.3. Section 2.4 discusses the empirical strategy, and the results are shown in Section 2.5. Section 2.6 concludes.

2.2 Background

Reading research published in non-open-access journals requires previous payment for specific articles or a journal subscription. Subscriptions can be costly because five publishers control 56 percent of the market (Sample, 2012; Stoy, Morais and Borrell-Damián, 2019). Hence, there is substantial variation in access to research across universities and countries. While publishers partly serve an economically meaningful purpose – ensuring quality scientific standards, curating and disseminating academic work – they cannot internalize the benefits of offering free access. As a result, knowledge through openly accessible publications is likely an under-provided public good.

[7] Consistent with the absence of effects on biomedical research in Bryan and Ozcan (2021), we also attain the smallest effect sizes for medical and biochemical research.

Inhibited by access restrictions, in 2011, a former student from Almaty, Kazakhstan, founded Sci-Hub. Sci-Hub is a so-called shadow library, an online platform that contains illicit collections of scientific papers downloadable for free by anyone with an internet connection. Sci-Hub is by far the world's largest and most prominent shadow library. In 2016, it hosted more than 50 million academic papers covering roughly 85% of all closed-access papers, and in 2017 the platform had roughly 500,000 daily visitors (Bohannon, 2016; Himmelstein et al., 2018). By late 2022, the website counted approximately 3 million daily downloads worldwide.[8] To put these numbers into perspective, the traffic is comparable in magnitude to websites such as JSTOR or PubMed. JSTOR counted approximately 600,000 daily downloads in 2019[9] whereas PubMed received approximately 3 million searches and 2.5 million unique visitors per day in 2017 (Fiorini, Lipman and Lu, 2017).

Despite the large traffic, academic file-sharing platforms are still not known by a large number of researchers. A survey by Segado-Boj, Martín-Quevedo and Prieto-Gutiérrez (2022) reached out to roughly ninety thousand scientists around the world to document the use of pirated document repositories. Even in the arguably positively selected sample of 3,300 respondents, only a little over half indicated ever having used such a platform. The remainder did not partly because of ethical concerns (46%), but also simply because they didn't know such platforms existed (36%).

Sci-Hub was neither the first nor is currently the only shadow library. While other shadow libraries existed beforehand, they either focused on hosting illicit copies of academic books, like Library Genesis, or were only available to tech-savvy users. Sci-Hub obtains scholarly work through leaked authentication credentials for educational institutions (Elbakyan, 2017). These credentials enable Sci-Hub to use institutional networks and gain access to the content of restricted-access journals. Academic work through this channel is subsequently incorporated into the Sci-Hub database and made available through the website. The ease of use was likely a key factor for Sci-Hub becoming the most prominent shadow library for journal publications. In Appendix Figure 2.B.1 we illustrate Sci-Hub's front page.

Despite its rapid spread, Sci-Hub was not met with unequivocal appreciation. Large publishers pushed back against the platform in courts around the world.

[8] Source: sci-hub.se/stats, accessed on 26th of November, 2022.
[9] Source: about.jstor.org/librarians/journals/, accessed on 14th of January, 2023

As a result, Sci-Hub lost numerous legal disputes, and the platform had to cycle through at least 54 different domain names. In particular, the Eastern District Court of Virginia (2017) *"[...] ordered that any person or entity in privity with Sci-Hub [...], including any Internet search engines, web hosting, and Internet service providers, [...], and domain name registries, cease facilitating any or all domain names and websites through which Defendant Sci-Hub engages in unlawful practices."* Yet, to this date, the platform has remained online.

2.3 Data

Our main analysis relies on an annual global panel of subnational units from 2000 to 2022. The panel results from three primary data sources. First, we use publicly available log files from Sci-Hub that record micro-level download activity from 2011 to 2013 and 2015 to 2017. For each download, we know the date and geographic location of the download and the work retrieved. We observe more than 300 million download requests across 100,000 unique geographic locations within our observation period. Second, we collect data on global scholarly output. Drawing on data from OpenAlex, the successor to Microsoft Academic Graph, we construct for each sub-national unit measures on publications, citations, and references. For all measures, we distinguish between open- and restricted-access status as well as quality and field of research. Third, to implement our identification strategy, we add information on social network linkages between sub-national regions and Almaty, where Sci-Hub was originally founded. These data are drawn from an anonymized snapshot of all active Facebook users and their friendship networks.

2.3.1 Measuring Sci-Hub Activity

Sci-Hub log files were made available in three batches. First, logs of Sci-Hub usage from September 1, 2015, through February 29, 2016, were released as part of a descriptive study in Science (Bohannon, 2016). Log files for 2017 were released on January 18 and updated on May 15, 2018. Finally, log files from 2011 to 2013 were released on January 27, 2020. Overall, the log files cover 1,394 days of Sci-Hub usage, and 300 million recorded resolved requests.

The log files contain three unprocessed pieces of information for all resolved requests.[10] First, they record the exact download date of each request from which we identify the corresponding download year. Second, data entries include the geographical location from which the download was made based on the IP address of the download device. Unfortunately, it is impossible to determine whether the location determined from the IP address matches the actual location of the Sci-Hub user. For example, the two locations diverge if a virtual private network (VPN) is used. While VPN usage likely introduces noise, it is unlikely to invalidate our identification strategy and bias our results. First, VPNs were not as ubiquitous and easy to use as they are today. Second, and more importantly, for our results to be affected, VPN usage would need to (1) differently change in high versus low connected sub-national units to Almaty after the introduction of Sci-Hub (conditional on all covariates) while also (2) being correlated with our outcomes of interest. So far, we do not have any evidence of this backdoor mechanism. Moreover, Elbakyan herself has stated that less than 3% of Sci-Hub users relied on VPNs (Bohannon, 2016). After pre-processing the log files, we observe downloads across more than 100,000 unique geographic locations, which we spatially aggregate into subnational units in a final step. The reason we aggregate data by regions, as opposed to institutions, is that we cannot directly link downloads to individual institutions. The third entry in the log files is the DOI of the downloaded paper that allows attaching paper- and journal-specific characteristics to each download.

Figure 2.1 shows the daily number of resolved requests across the time span for which log files are available. Comparing the horizontal axis labeling between Panels (a) and (b) shows the rapid increase in Sci-Hub usage from its onset in late 2011 to our last observations in late 2017. The oscillating pattern reflects usage peaks during the week and a leveling off of research activity on weekends. Days with zero requests represent server outages. On average, each researcher performs 4.1 downloads, a total of 217 downloads per research institution (see Panel A of Appendix Table 2.A.1).

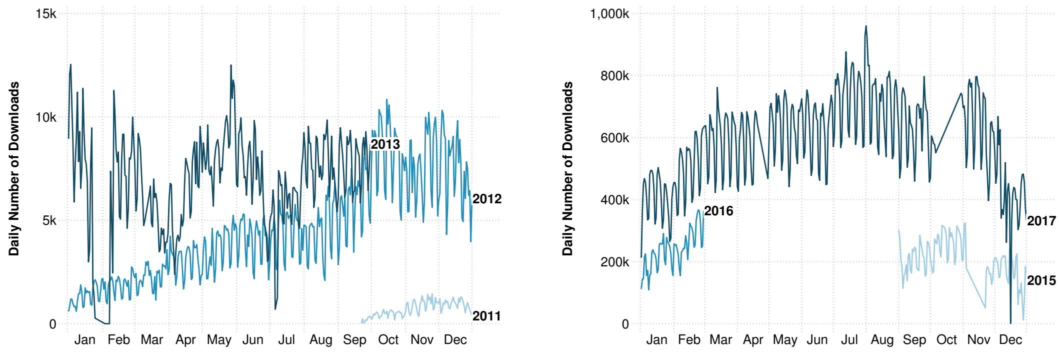
2.3.2 Measuring Global Research Output

To construct outcome measures, we draw on OpenAlex. OpenAlex is a fully open catalog of global research output. The platform replaced Microsoft Academic

[10] Appendix Figure 2.B.2 shows the structure of an entry in the Sci-Hub log-files and describes how it is subsequently processed.

Figure 2.1: Sci-Hub Downloads over Time

(a) 2011-2013 (b) 2015-2017



Note: The figure shows the average daily Sci-Hub downloads by year. The figure includes all downloads recorded in Sci-Hub log files from 2011 to 2013 and 2015 to 2017.

Graph (MAG), which was discontinued at the end of 2021. Its database was initially based on MAG's existing records, but subsequently, coverage was improved by incorporating data from Crossref, ORCID, Pubmed, arXiv, and DOAJ, among many others. OpenAlex hosts all kinds of scholarly output, including journal articles, books, datasets, and theses. At the end of 2022, OpenAlex indexed close to 300 million works.

Recent bibliometric studies show that OpenAlex significantly increased MAG's coverage (Scheidsteger and Haunschild, 2022), which already, before its discontinuation, outperformed other subscription-based platforms such as Scopus, Web of Science and Dimension in terms of coverage (Martín-Martín et al., 2021). With Google Scholar unavailable for bulk data usage, OpenAlex appears to be the most suitable alternative to studying global research patterns.

To construct measures of global research output, we download a snapshot of the entire OpenAlex database as of August 2022 (roughly 300 gigabytes of compressed data). The unit of observation within OpenAlex's database is a scholarly work, a journal article, a book, a dataset, or a thesis. To each work, multiple pieces of publication-specific information are attached. Importantly, this includes the publication year, the host venue (in most cases, journals), and a list of referenced works.[11] The list of referenced works allows us to back out the number and quality of citations for each work. In our main analyses, we focus specifically on journal publications and exclude non-scholarly works.

[11] OpenAlex provides several other pieces of information. A complete list of available characteristics can be found here.

Each article is connected to a set of authorship objects, representing an author and their affiliated institution at the time of publication. Based on the affiliation of authors and the geolocation of institutions[12], we assign publications to sub-national units. Each work is only counted once per institution for articles with multiple co-authors from the same affiliation. If an author has multiple affiliations across sub-national units, the publication is assigned to each sub-national unit separately. Appendix Figure 2.B.3 gives an overview of the information we extract from each entry in OpenAlex. The key output measures we construct are the number of references and citations. For clarity, we denote references as citations from an author in a given region to *other* papers – we interpret referencing as a measure of knowledge consumption. Citations, on the other hand, are citations *received* by an author in a given region from other researchers. Here, we treat citations as a measure of scientific quality and impact.[13] Summary statistics are presented in Appendix Table 2.A.2.

Finally, to trace out potential impacts on research topics and direction, we construct a text-based measure of similarity to the research frontier. In particular, for each scientific field and year, we train a topic model on all papers in the top percentile of the citation distribution. For each other article, we then compute the Mahalanobis distance to these top publications. A detailed description of this procedure is provided in Appendix Section 2.D.2.

Matching Open-access Status, Quality, and Field We corroborate each work with journal-specific metrics provided by Scopus' yearly ranking of peer-reviewed journals.[14] All journal measures retrieved through Scopus are fixed in 2011[15] to rule out that our results are driven by time trends in any of these metrics. For example, in 2011 the journal ranking list included 19,941 journals, identifiable by the time-invariant 'International Standard Serial Number'.

We extract three key measures. First, Scopus computes a measure of scientific influence for each scholarly journal that accounts for the number of citations received by a journal and the importance or prestige of the journals from which

[12] For each of the 109,000 institutions covered by OpenAlex, a separate database provides a mapping from institution identifiers to geolocations.

[13] We acknowledge that citations are an imperfect measure of quality. Nonetheless, citations are correlated with several meaningful characteristics that imply greater quality. Specifically, they are correlated with positive peer reviews (Card and DellaVigna, 2020), perceived influence (Teplitskiy et al., 2022), and how much a given paper impacts the language of subsequent papers (Gerrish and Blei, 2010).

[14] Scopus is Elsevier's abstract and citation database

[15] 2011 is the earliest year for which Scopus journal metrics are available.

such citations come. Based on this citation score, journals are assigned field-specific quality percentiles. Second, Scopus reports open-access status for covered journals. Open-access status is based on whether the journal is listed in the Directory of Open Access Journals and/or the Directory of Open Access Scholarly Resources.[16] Third, journals are assigned fields based on the 'All Science Journal Classification' (ASJC) system. In total, there are 333 possible minor fields, which can be aggregated into 27 major fields. Finally, all journal metrics are matched to works from OpenAlex based on the ISSN, which is recorded in both data sources.

Additional Measures In addition, we utilize the OpenAlex database to construct educational measures describing the scientific landscape in sub-national units. Precisely, we measure the number of researchers in sub-national units as of 2010 by counting the unique number of authors recorded in OpenAlex between 2008 and 2012. Moreover, we construct measures for the number of research institutions[17] per sub-national unit, the number of research institutes above the 95th percentile per sub-national unit (measured by citations), and whether a sub-national unit has any research institute.

Aggregation The final step aggregates publication, citation, and reference data across years and sub-national units. Panels A and B of Appendix Table 2.A.3 provide summary statistics on the number of research institutes and researchers in sub-national units. Panels C, D, and E of Appendix Table 2.A.3 give an overview of global research activity across sub-national units. A researcher produces, on average, 1.53 publications per year, of which 67% are published in peer-reviewed journals, 56% of which are open-access. Each paper references, on average, 17 publications, of which 32% are open-access publications. The mean number of citations is 14.47, most originating from peer-reviewed publications.

2.3.3 Measuring Connectedness to Almaty

To measure social ties between sub-national units we use the Social Connectedness Index (CON) as introduced by (Bailey et al., 2018). The index builds on aggregated and anonymized information from the universe of Facebook (FB)

[16] We acknowledge that increasingly, journals offer mixed open-access policies where authors can pay a fee to have their publication openly accessible. For example, 'Nature' charges authors up to 9,500 Euros to make research papers free to read.

[17] Research institutions include universities and other organizations, such as non-profits, government institutions, archives, or corporations, with which authors are affiliated.

friendships as of April 2016. Given Facebook's scale, with 2.1 billion active users, the index provides a large-scale representation of global friendship networks measurable at a sub-national level.

In particular, the Social Connectedness Index, constructed as follows,

$$CON_i^j = \frac{\text{Facebook Friends}_{i,j}}{\text{Facebook Users}_i \cdot \text{Facebook Users}_j} \text{ with } \max_{i,j} CON_i^j = 1,000,000$$

measures the relative probability of a FB friendship between sub-national unit i and sub-national unit j.[18] Sub-national units for European countries are based on the European Nomenclature of Territorial Units or Statistics (NUTS2, 2018). Countries outside Europe are divided into sub-national units based on the Database of Global Administrative Areas (GADM1 Version 2.8, 2015). Countries with a population of less than 1 million are not divided. For each pair of sub-national units, we observe CON_i^j. For example, sub-national unit i with twice the social connectedness index of sub-national unit i' would be twice as likely to have a friend in sub-national unit j.

Using the Social Connectedness Index has two caveats. First, the Social Connectedness Index is not available for other periods. In that sense, we are limited to cross-sectional variation.[19] Second, the Social Connectedness Index is unavailable for countries that restrict FB usage. Figure 2.2 Panels (b) and (c) give a spatial overview of raw and residualized connectedness between subnational regions and Almaty. Notably, there is no information on Russia, China, and Iran, among others.

Appendix Table 2.A.1 Panel B provides summary statistics of CON_i^j for Almaty, Nur-Sultan (the Kazakh capital), Kazakhstan[20], and all other capitals in Central Asia.

2.3.4 Additional Data Sources

We extend the panel with many additional variables that primarily function as control variables. First, we collect global nighttime light emission data at

[18] Note that the index contains a small amount of random noise and is rounded to the nearest integer to ensure that no single individual or friendship link can be identified from the data.
[19] We discuss threats to identification in greater detail in Section 2.4.
[20] The Social Contentedness Index for Kazakhstan results from aggregating sub-national connectedness measures of Kazakhstan weighted by their population shares. In particular, the index can be aggregated to larger geographical units using the following formula: $CON_i^j = \sum_{r_i} \sum_{r_j} \text{PopShare}_{r_i} \times \text{PopShare}_{r_j} \times CON_{r_i}^{r_j}$.

a resolution of 30 arc-seconds to create a proxy for differences in economic development (Li et al., 2020). Second, we utilize gridded population data at a resolution of 30 arc seconds (CIESIN, 2020). Both measures are projected on sub-national units. Third, we gather geographic details for each sub-national unit. Specifically, we compute the latitude and longitude of each sub-national unit's geographic centroid and the distance of each centroid to Almaty. We also compute measures for the area of a sub-national unit and whether a sub-national unit contains a country's capital. Finally, we classify countries into developed, emerging, and developing regions to gauge heterogeneous effects. To tie our hands, the classification is based on data by the International Monetary Fund (2011), and the United Nations (2011). The geographic distribution is shown in Appendix Figure 2.B.4.

2.3.5 Dealing with Zero Observations

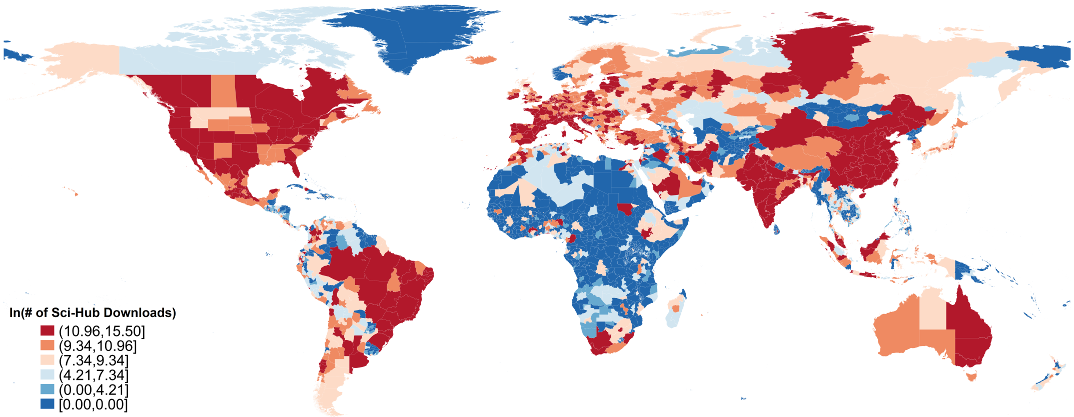
All count variables with a skewed distribution are transformed using the natural logarithm, adding one in case of zero observations. As a robustness test, we additionally apply the inverse hyperbolic sine transformation with $\text{arcsinh}(Y_{it}) = \ln\left(Y_{it} + (Y_{it}^2 + 1)^{1/2}\right)$. We are aware that marginal effects from linear regressions using $\log(1 + Y)$ or $\text{arcsinh}(Y)$ transformations with zero observations can be sensitive to the scaling of the outcome if treatment affects the extensive margin (Chen and Roth, 2022; Mullahy and Norton, 2022).[21] However, in our setting, the main effect is likely to operate through the intensive margin, attenuating concerns that the estimates are distorted due to scale dependence. In particular, Sci-Hub affects existing research dynamics but is unlikely to impact research dynamics in regions with no prior research output.[22]

[21] In particular, Chen and Roth (2022) show that if the scale of non-zero values is large, a change from a zero to a typical non-zero value of the outcome has a huge impact, with the treatment effect placing substantial weight on the extensive margin.

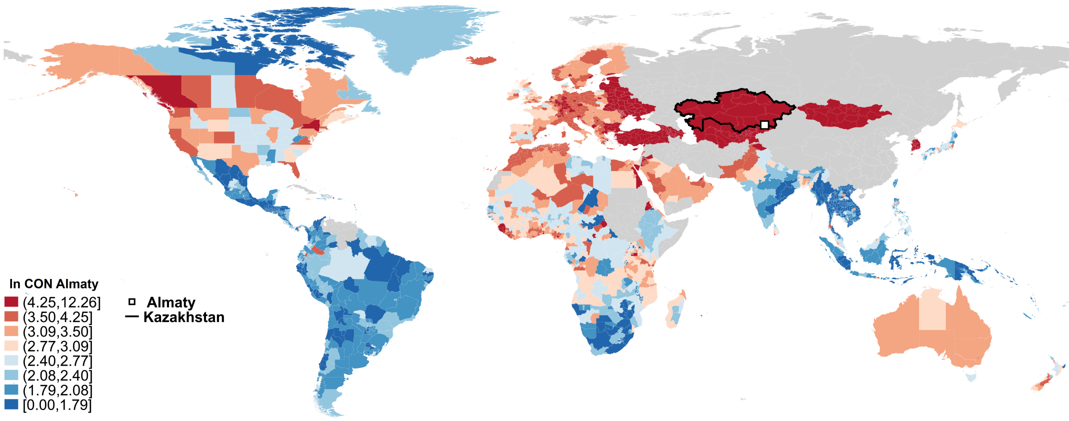
[22] In Appendix Table 2.A.4, we also show that Sci-Hub downloads do not correlate with the probability of (first-time) entering the academic landscape, implying that treatment does not affect the extensive margin.

Figure 2.2: Descriptive by Sub-national Units

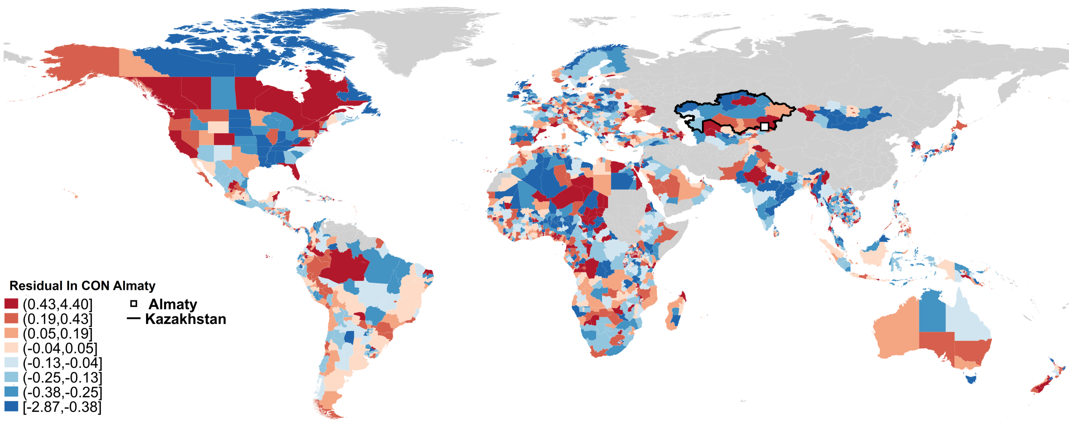
(a) Sci-Hub Downloads



(b) Social Ties to Almaty



(c) Residualized Social Ties to Almaty



Note: Panel (a) shows the spatial distribution of Sci-Hub downloads across sub-national units. Panel (b) depicts how social ties to Almaty vary across sub-national units. Panel (c) depicts the residualized variation (conditional on country fixed effects). The borders of Kazakhstan are marked by a black line. The location of Almaty is marked by the white square outlined in black.

2.4 Empirical Strategy

To identify the causal effect of Sci-Hub on knowledge consumption and creation, we apply an instrumented difference-in-differences framework. The first difference we harness is time. Sci-Hub only gained traction after 2011, so we compare observation units in the years before and after the platform's launch. The second difference is Sci-Hub intensity across sub-national regions. However, the geography of Sci-Hub web traffic is likely endogenous to knowledge creation, our outcome variables of interest. To circumvent endogeneity, we capture exogenous variation in the number of Sci-Hub downloads using social connectedness to Almaty, Kazakhstan. We rely on an anonymized snapshot of all Facebook friendships between subnational regions to construct the instrument.

Former Kazakh student Alexandra Elbakyan founded Sci-Hub in Almaty. We posit that individuals with pre-existing social ties to Almaty were more likely to be early adopters of Sci-Hub, as knowledge of the platform spread mainly by word-of-mouth. Relying on path dependence in technology adoption (Arthur, 1989), we argue that early exposure to Sci-Hub continues to be a strong predictor of sub-national Sci-Hub usage today (akin to Enikolopov, Makarin and Petrova, 2020; Müller and Schwarz, 2023). In the case of Sci-Hub, technological path dependence may have been particularly strong because diffusion outside of social networks was severely hampered by legal actions to stop the site from operating. In practice, we estimate the following first-stage equation:

$$\begin{aligned} \ln \mathrm{Down}_{it} = {} & \alpha_i + \alpha_{c(i)t} + \\ & + \beta_1 \ln \mathrm{CON}_i^{\mathrm{Almaty}} \times \mathbb{1}_{t>2010} + \sum_n \delta_1^{(n)} \ln \mathrm{CON}_i^{\mathrm{n}} \times \mathbb{1}_{t>2010} \\ & + \boldsymbol{X}_{i2010} \boldsymbol{\gamma}_t + \varepsilon_{it} \end{aligned} \tag{IV1}$$

where $\ln \mathrm{Down}_{it}$ is the log number of Sci-Hub downloads in sub-national region i in year t. Our instrument is constructed as the log of social connectedness between region i and Almaty interacted with a post-2010 dummy. Additionally, we control for the social ties of region i with all neighboring country capital regions n of Almaty[23], each interacted with a post-2010 dummy. Therefore, we isolate the idiosyncratic variation of connectedness to Almaty that cannot be attributed to, for example, general friendship linkages to metropolitan areas in Central Asia.

[23] Neighboring country capitals of Almaty are Nur-Sultan, Bishkek, Ashgabat, Tashkent, and Moscow (for which no FB user data exist).

The specification rigorously controls for potential unobserved factors influencing both Sci-Hub downloads and social ties to Almaty. Specifically, it includes subnational region fixed effects, α_i, capturing time-invariant regional characteristics, and country-year fixed effects, $\alpha_{c(i)t}$, accounting for country-specific factors that vary over time (e.g., national higher education reforms). Finally, we control flexibly for several covariates[24] measured in 2010 interacted with year dummies. Unexplained variation is captured by the error term ε_{it}, clustered at the sub-national level.

In the second step, we use predicted Sci-Hub intensity from Equation (IV1) to estimate the following two-stage least squares regression:

$$\begin{aligned}\ln Y_{it} = {} & \alpha_i + \alpha_{c(i)t} + \\ & + \beta_2 \ln \widehat{\mathrm{Down}}_{it} + \sum_n \delta_2^{(n)} \ln \mathrm{CON}_i^{\mathrm{n}} \times \mathbb{1}_{t>2010} \\ & + \boldsymbol{X}_{i2010}\boldsymbol{\phi}_t + \eta_{it}\end{aligned} \tag{IV2}$$

Here, Y_{it} constitutes scientific outcomes, but mainly the share of references to restricted-access journals *from* region i and the log number of citations *to* region i. The coefficient of interest is β_2. The control variables are akin to Equation (IV1).

Identifying Assumption The identifying assumption is that in the absence of Sci-Hub, high versus low connected regions to Almaty would have followed parallel trends in scientific outcomes. This implies that conditional on covariates and fixed effects, social ties to Almaty are orthogonal to η_{it} in Equation (IV2).

Reverse causality A key limitation of the design is that our measure of social connectedness is built on a Facebook snapshot from 2016. We implicitly assume that the network structure has been stable over the years. The existing literature supports this assumption (see, e.g., Kuchler, Russel and Stroebel 2022). It is also doubtful that Sci-Hub shaped the Facebook network structure meaningfully. The overall fraction of scientists in the general population would need to be

[24] The list of control variables includes measures for (1) education (any research institute, number of research institutes, number of research institutes in the 95-100 percentile range, number of researchers in 2010), (2) geography (latitude and longitude of geographical center, distance to Almaty, capital status, area), (3) population (population in 2010), and (4) development (nighttime light emission in 2010).

unreasonably large. Similarly, Bailey et al. (2021) show that even large-scale international trade appears to be no key driver of network formation on Facebook.

2.5 Results

In this section, we present the main results on the relationship between Sci-Hub downloads and subsequent knowledge creation.

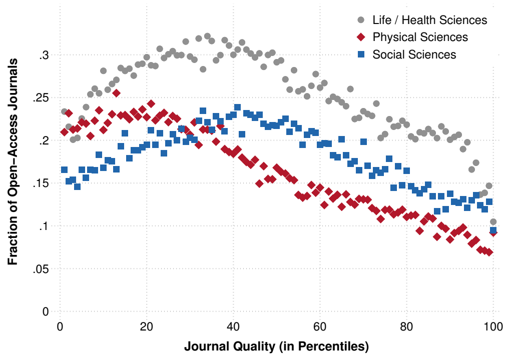
2.5.1 Motivating Facts

Before diving into the causal analysis, we document several empirical facts to motivate our causal analysis. First, we use journal-level data. We ask, how is open access status distributed across journals? We find that on average only 20% of all journals provide free access to published articles. Beyond this first data moment, Figure 2.3 Panel (a) shows large heterogeneity in open access regimes across two dimensions: field and journal quality. We document that open access is most prevalent in the life and health sciences and slightly less so in the physical and social sciences. Consistently across fields, we find that the number of open-access journals dwindles toward the top of the journal quality distribution. In the highest cited percentile of journals only 9% operate under open access. Scientific knowledge is not only highly restricted across fields but these restrictions are particularly severe for knowledge residing in top journals. In Appendix Figure 2.B.5 we further document that open-access journals have become gradually more common over the past decade but remain a small share of all journals.

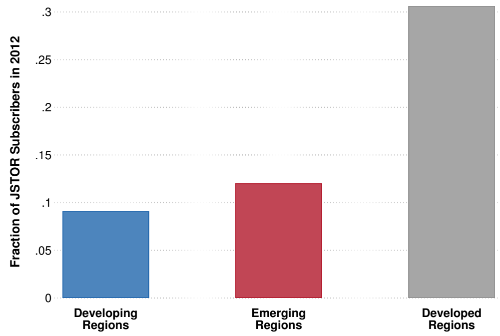
If scientists had universal access through affiliated libraries, these paywalls would not necessarily harm the consumption and production of new scientific insights. However, in Figure 2.3 Panel (b) we show that this appears not to be the case. We proxy for library access using institutional JSTOR subscriptions in 2012. JSTOR is an online library covering roughly 12 million items and access to over 2800 journals. While incomplete, bulk access through JSTOR still allows researchers to read a large number of scholarly works without individual fees. We find that JSTOR subscriptions are largely unequally distributed across universities. While 30% of all institutions in developed regions have subscriptions, the fraction is reduced to roughly 10% in less-developed regions. In Appendix Figure 2.B.6 we show that the unequal distribution of JSTOR access across regions of different

Figure 2.3: Four Facts

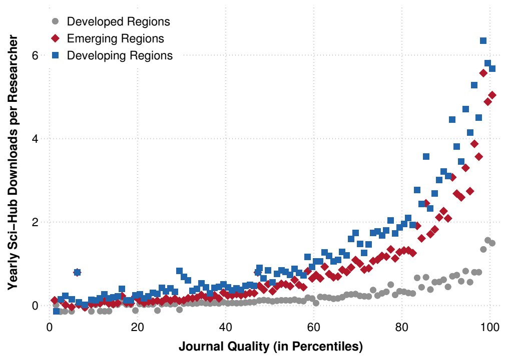
(a) Fraction of Open-Access Journal by Journal Quality across Fields



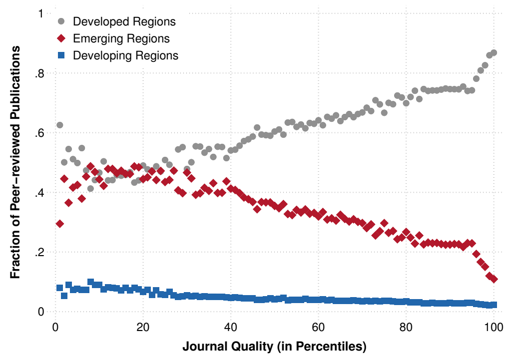
(b) Fraction of JSTOR Subscribers by Region in 2012



(c) Average Yearly Sci-Hub Downloads per Researcher by Journal Quality and Region



(d) Fraction of Peer-reviewed Publications by Journal Quality across Regions



Note: Panel (a) shows the fraction of open-access journals by quality across fields pooled from 2011 to 2022 accounting for year fixed effects. Panel (b) shows the fraction of JSTOR subscribers per research institute across developing, emerging, and developed regions in 2012. Panel (c) shows the average annual Sci-Hub downloads per researcher by journal quality in the different regions. The sample includes all peer-reviewed scientific papers recorded in Sci-Hub log files from 2011 to 2013 and 2015 to 2017. Panel (d) shows the fraction of peer-reviewed publications by journal quality across regions. The figure includes all publications between 2000 and 2022 that are recorded in OpenAlex and are assigned to a journal.
Sources: Journal access and quality data are from Scopus. Journals are declared as open-access status if the journal is listed in the Directory of Open Access Journals and/or the Directory of Open Access Scholarly Resources. Journal quality percentiles are based on the average number of citations from peer-reviewed articles per publication. Country classifications of sub-national units into developed, emerging, and developing regions is based on data by the International Monetary Fund (2011), and the United Nations (2011). JSTOR subscription data come from the JSTOR website as recorded by the Internet Archive in 2012 and the underlying number of institutes from OpenAlex.

economic levels holds even when fixing the quality of institutions. Comparing universities with similar citation levels, the probability of a JSTOR subscription still depends largely on the economic environment.

Does the unequal distribution of bulk access simply mimic heterogeneous demand for scientific articles? To answer this question, we turn to the Sci-Hub data. For each downloaded paper, we add information on the respective journal's quality.

In Figure 2.3 Panel (c), we show the distribution of downloaded papers by varying degrees of journal quality. Unsurprisingly, we find that articles from top journals are downloaded disproportionately often. We further disaggregate downloads by different origins. The data clearly shows that Sci-Hub traffic per researcher is much higher in lesser-developed regions of the world. Individuals in developing regions download four times as many papers (per researcher) than individuals in highly developed regions. This suggests that demand for closed-access papers exists beyond legitimate channels and is large. Moreover, the differential traffic indicates that the constraints are particularly binding for scholars in less developed regions of the world.

Finally, we turn to the production of scientific knowledge. In Figure 2.3 Panel (d), we show fractions of peer-reviewed publications by papers' origins and respective journal quality. We find that most papers written originate from industrialized, developed regions. This is true across different levels of quality, but it is increasing among top journals. While roughly 50% of papers in below-median-level journals stem from developed regions, this fraction increases to close to 90% in the top one percentile of journals. The remainder of papers is predominantly written in middle-income countries. This suggests that the least developed regions lack the means to conduct scientific activities at a larger scale and researchers from middle-income countries face difficulties publishing in the highest echelons of scientific journals. These patterns are shaped by a multitude of different factors. Yet, in the subsequent analyses, we show that access-restrictions play a meaningful role in explaining the geography of scientific knowledge production.

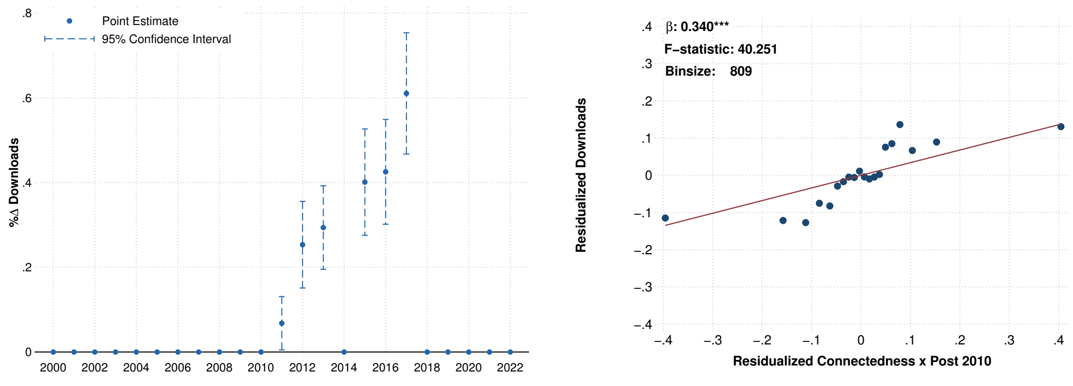
2.5.2 Effects on Knowledge Consumption

To what extent does Sci-Hub affect scientists in their research downstream? In this section, we isolate the effect of the platform on a measurable scientific outcome: references. We argue that once scientists learn of Sci-Hub and use the platform extensively, they start referencing more paywalled papers in their articles – Sci-Hub reshapes global knowledge consumption.

First Stage To make a causal claim, we rely on the identification strategy outlined in Section 2.4. First, we estimate equation (IV1) to show that connectedness to Almaty is a meaningful driver of Sci-Hub traffic. The dynamic event study estimates are shown in Figure 2.4 Panel (a). According to the point

Figure 2.4: First Stage – Visual Evidence

(a) Event Study (b) Binned Scatterplot



Note: Panel (a) shows point estimates and confidence intervals of the dynamic effects corresponding to the specification in Table 2.1 Panel A column (8). Panel (b) plots the residuals and coefficient estimate of the corresponding static difference-in-differences model. Standard errors are clustered by subnational region. Bars represent 95% confidence intervals.

estimates, connectedness is a strong and highly significant predictor that grows in magnitude over time. Note that by construction, we cannot estimate pre-trend coefficients because both the platform and downloads did not yet exist before 2011. Moreover, we, unfortunately, do not observe granular download data in 2014 and after 2017. Particularly in recent years, it is not clear how the correlation would behave if data were available. On the one hand, we would expect social networks' importance to decline in the long run. However, recent survey evidence in an arguably positively selected sample still documents a lack of knowledge about pirating websites as one of the leading factors for not having used such services (Segado-Boj, Martín-Quevedo and Prieto-Gutiérrez, 2022).

Complementary to the event study, Figure 2.4 Panel (b) shows a binned scatterplot of the first-stage correlation, again focusing on our most demanding specification. The figure illustrates the range of variation and provides evidence that the linear model is a good approximation of the data. The corresponding static estimates are presented in Table 2.1. The most demanding specification in Panel A Column (8) suggests that an increase in connectedness by 1% is associated with a 0.34% higher Sci-Hub traffic with an F-statistic of approximately 40. Conditional on connectedness to neighboring country capitals and educational metrics, the coefficient remains consistent when introducing additional control variables. In Appendix Table 2.A.5 we show that the first stage is not sensitive to applying the inverse hyperbolic sine transformation.

Table 2.1: First Stage Estimates

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
ln CON Almaty × Post 2010	0.617***	0.646***	0.755***	0.458***	0.297***	0.304***	0.341***	0.340***
	(0.020)	(0.019)	(0.075)	(0.076)	(0.052)	(0.053)	(0.054)	(0.054)
Observations	41,341	41,341	40,440	40,440	40,440	40,440	40,440	40,440
Number of Clusters	2,437	2,437	2,384	2,384	2,384	2,384	2,384	2,384
F-statistic	912.118	1180.114	100.849	36.685	32.264	32.807	40.154	40.251
Fixed Effects								
Sub-national	-	✓	✓	✓	✓	✓	✓	✓
Year × Country	-	-	✓	✓	✓	✓	✓	✓
CON Neighb. Capitals	-	-	-	✓	✓	✓	✓	✓
Controls in 2010 × Year FE								
Education	-	-	-	-	✓	✓	✓	✓
Geography	-	-	-	-	-	✓	✓	✓
Population	-	-	-	-	-	-	✓	✓
Development	-	-	-	-	-	-	-	✓

Note: The table displays regression results from Equation (IV1) across various specifications. Standard errors are clustered at the sub-national level. Significance levels are indicated as follows: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Design Validity We perform several exercises to support our identification strategy. A key concern is that the observed correlation is not an artifact of connectedness to Almaty, but of being more connected in general. We provide two pieces of evidence against this argument. First, we run a horse race. In particular, we regress the log number of Sci-Hub downloads on connectedness to Almaty, the unofficial capital of Kazakhstan, simultaneously accounting for connectedness to other regions with capital cities in Central Asia. The results of this exercise are shown in Table 2.2. We find that connectedness to Almaty is the only consistent, positive and large predictor of Sci-Hub downloads. All remaining coefficients are small and close to zero or even negative. This is true for direct neighboring capital cities, Column (5), and more distant, non-neighboring capital cities, Column (6). For all remaining analyses, we continue to use the first-stage estimates from Column (5) with direct neighboring capitals as controls to capture the idiosyncratic variation of connectedness to Almaty and not Central Asia.

Second, we re-estimate the first-stage equation by independently considering social ties to all other sub-national units (while controlling for social ties to neighboring capital regions). This exercise allows us to compare the estimate for Almaty with all other regions in our data. In Figure 2.5, it is evident that the Almaty correlation is a highly distinct outlier in the near-normal distribution

Table 2.2: First Stage Estimates – Horse Race

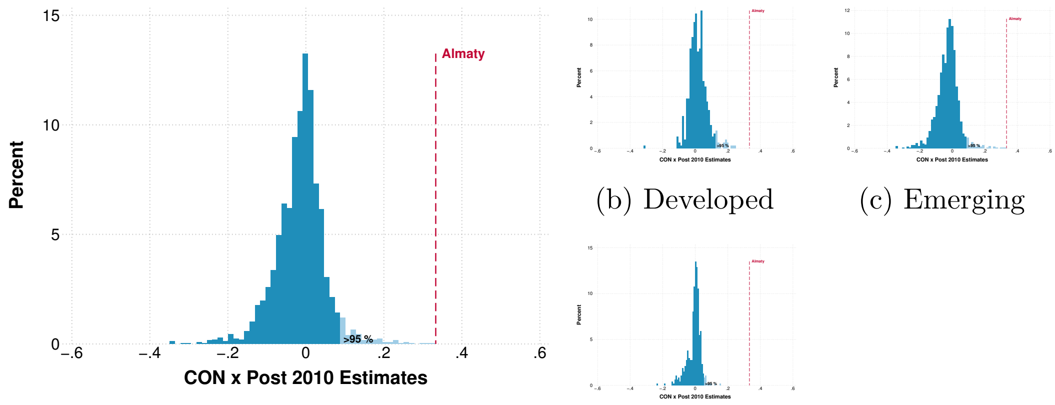
	Dependent Variable: ln Downloads					
	(1)	(2)	(3)	(4)	(5)	(6)
ln CON Almaty × Post 2010	0.274***	0.285***	0.322***	0.307***	0.340***	0.319***
	(0.042)	(0.048)	(0.050)	(0.047)	(0.054)	(0.056)
ln CON KAZ excl. Almaty × Post 2010	–	-0.022	–	–	–	–
	(–)	(0.054)	(–)	(–)	(–)	(–)
ln CON Nur-Sultan × Post 2010	–	–	-0.074*	–	-0.069*	-0.072*
	(–)	(–)	(0.039)	(–)	(0.039)	(0.039)
ln CON Bishkek × Post 2010	–	–	–	-0.067*	-0.059*	-0.066*
	(–)	(–)	(–)	(0.036)	(0.036)	(0.036)
ln CON Ashgabat × Post 2010	–	–	–	-0.020	-0.017	-0.018
	(–)	(–)	(–)	(0.014)	(0.014)	(0.014)
ln CON Tashkent × Post 2010	–	–	–	0.033	0.049	0.033
	(–)	(–)	(–)	(0.038)	(0.040)	(0.042)
ln CON Dushanbe × Post 2010	–	–	–	–	–	0.005
	(–)	(–)	(–)	(–)	(–)	(0.030)
ln CON Ulaanbaatar × Post 2010	–	–	–	–	–	0.041
	(–)	(–)	(–)	(–)	(–)	(0.037)
ln CON Kyiv × Post 2010	–	–	–	–	–	0.027
	(–)	(–)	(–)	(–)	(–)	(0.051)
Observations	40,440	40,440	40,440	40,440	40,440	40,440
F-statistic	42.063	34.805	41.232	41.923	40.251	32.035
Fixed Effects						
Sub-national	✓	✓	✓	✓	✓	✓
Year × Country	✓	✓	✓	✓	✓	✓
Controls in 2010 × Year FE	✓	✓	✓	✓	✓	✓

Note: The table displays regression results from Equation (IV1) across various specifications. Standard errors are clustered at the sub-national level. Significance levels are indicated as follows: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

of placebo estimates. We conclude that diffusion through social networks was driven by social links to Almaty, which cannot be explained by connectedness to similar regions in Central Asia or network connectedness in general.

Reduced Form We depict the dynamic reduced form in Figure 2.6 Panel (a). With the launch of Sci-Hub in 2011, we see a quick and quantitatively large rise in the share of references to restricted-access publications from highly connected regions. Scientists start referring to previously restricted works at much greater rates. Based on the point estimates, doubling a region's connectedness to Almaty is associated with an increase of roughly twelve percentage points in the share of restricted-access references in the later sample periods. The event study also shows that regions with different levels of connectedness are not on

Figure 2.5: First Stage – Placebo Effects of Connectedness on Sci-Hub



(b) Developed (c) Emerging

(a) All Regions (d) Developing

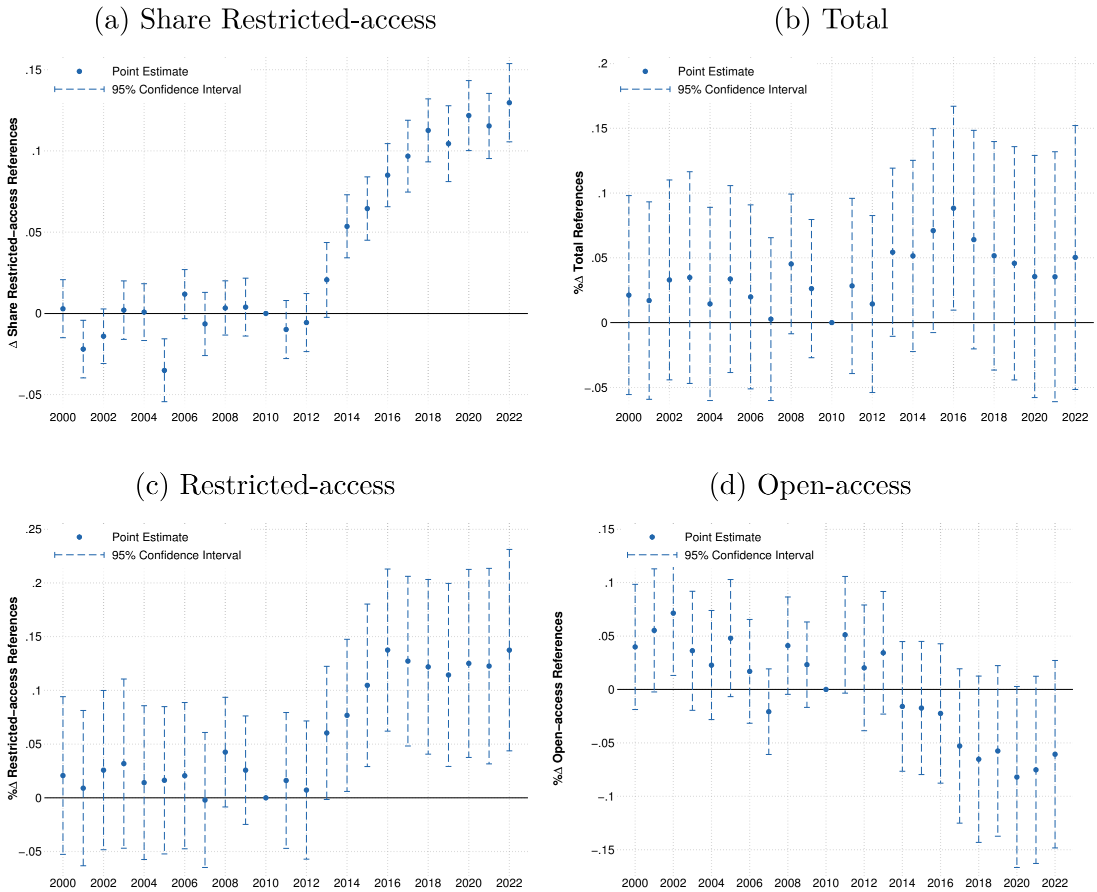
Note: Panel (a) shows the distribution of point-estimates when re-estimating Equation IV1 by iteratively replacing social connectedness to Almaty with social ties to all other sub-national units. We replace social ties to Almaty's neighboring capital regions with respective other neighboring capital regions. For each region, we control for social ties to respective neighboring capital cities. Panels (b) to (d) show the distribution of point-estimates within specific regions. Classification of sub-national units into developed, emerging, and developing regions is based on data by the International Monetary Fund (2011), and the United Nations (2011). In all figures, the dotted red line corresponds to the point estimate in column 8 of Table 2.1.

diverging outcome trajectories before the Sci-Hub launch. Instead, we identify considerably stable pre-trend coefficients before 2011 that are overall close to zero. This reassures that the parallel trends assumption appears to hold, at least in the pre-period. The static equivalents to the dynamic reduced form effects are displayed in Panel A of Table 2.3. In the static reduced form, we find an average increase in restricted-access references of a little less than 5% when a region doubles connectedness to Almaty. Note that the sample here is restricted to years before the launch of Sci-Hub and years in the post-period for which we observe Sci-Hub downloads (2011-2013 and 2015-2017). Within this subsample the static reduced form coefficient equals the average of the event study coefficients for 2011-2013 and 2015-2017 which explains the smaller magnitude.[25]

As before, we also conduct a placebo exercise. In particular, we estimate the static reduced form coefficient for connectedness to all other regions in our data. The result is depicted in Figure 2.7. Akin to the first-stage placebo estimates, we find that the uptake in closed-access references is driven by connectedness to Almaty and appears not to be explained by connectedness to other regions.

[25] Further note sample differences between Columns (1)–(3) and (4). Since the outcome in Column (4) is a share, all region-year observations with zero entries in total references are dropped.

Figure 2.6: Effects of Connectedness on References

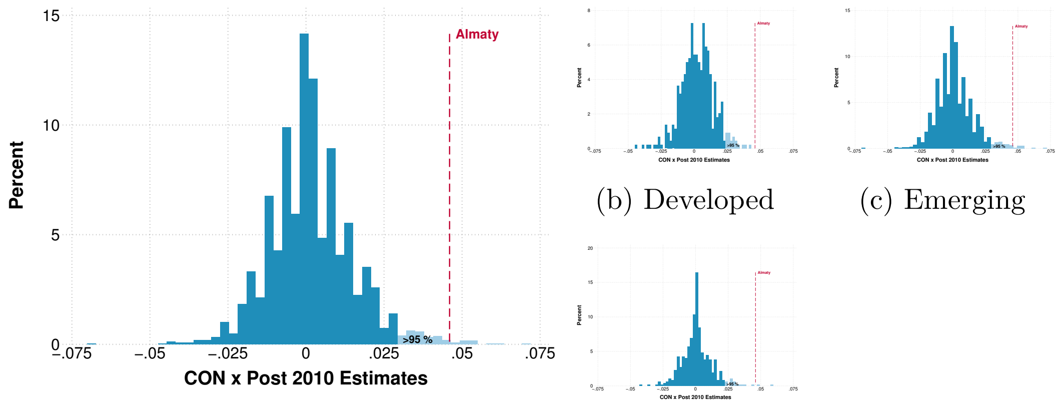


Note: The figure shows reduced form event study estimates for the outcomes and specification displayed in Table 2.3 Panel A. The post-2010 indicator is replaced with a full set of annual indicators, omitting 2010, the year before Sci-Hub was established. Standard errors are clustered by subnational region. Bars represent 95% confidence intervals.

Returning to Figure 2.6, in Panel (b), we further show no effect of connectedness to Almaty on the total number of references – scientists do not appear to consume more papers. Instead, we find a pattern of substitution. Connected researchers read more paywalled work and reference more of these in their research (Panel (c)). This comes at the expense of references to open-access publications. Panel (d) indicates a drop in these references in the post-period. Note that the shift in reference patterns occurs two to three years after the launch of Sci-Hub. This is consistent with lower usage rates in the early years but is also consistent with academic publication lags.

IV Combining our first stage and reduced form results, Panel B of Table 2.3 displays the 2SLS estimates on references for our most demanding specification. We find that doubling Sci-Hub traffic is associated with a 4.6% point increase in

Figure 2.7: Placebo Effects of Connectedness on References



(a) All Regions

(b) Developed

(c) Emerging

(d) Developing

Note: Panel (a) shows the distribution of point-estimates when re-estimating the reduced form effect by iteratively replacing social connectedness to Almaty with social ties to all other sub-national units. We replace social ties to Almaty's neighboring capital regions with respective other neighboring capital regions. The outcome is the share of restricted-access references. Panels (b) to (d) show the distribution of point-estimates within specific regions. Classification of sub-national units into developed, emerging, and developing regions is based on data by the International Monetary Fund (2011), and the United Nations (2011). In all figures, the dotted red line corresponds to the point estimate in Panel A column 4 of Table 2.3.

the share of restricted-access references. Note that this is a pooled estimate for the post-period in which we observe Sci-Hub downloads (2011-2013 and 2015-2017). Since the reduced form effect is particularly strong in later years (post 2017), we would, in all likelihood, obtain even larger estimates if more recent Sci-Hub data were available. This becomes evident when implementing the two-sample 2SLS approach in which we keep the otherwise missing years (Angrist and Krueger, 1992; Inoue and Solon, 2010). In Appendix Table 2.C.2 we find that doubling Sci-Hub is associated with an 8% point increase in the share of restricted-access references using the alternate approach. Finally, in Appendix Section 2.C.1 we discuss and show robustness to weak-IV considerations.

Table 2.3: Change in Reference Patterns

	Number of References			Share Restricted-access References
	Total	Open-access	Restricted-access	
	(1)	(2)	(3)	(4)
Panel A: Reduced Form				
ln CON Almaty × Post 2010	0.038	-0.026	0.066	0.047***
	(0.041)	(0.032)	(0.040)	(0.010)
Observations	40,440	40,440	40,440	19,420
Panel B: 2SLS				
ln Downloads	0.111	-0.077	0.193	0.046***
	(0.121)	(0.094)	(0.121)	(0.012)
Observations	40,440	40,440	40,440	19,420
F-statistic	40.251	40.251	40.251	30.898
Panel C: OLS				
ln Downloads	-0.014*	-0.012*	-0.010	0.002**
	(0.009)	(0.008)	(0.008)	(0.001)
Observations	40,440	40,440	40,440	19,420
Fixed Effects				
Sub-national	✓	✓	✓	✓
Country × Year	✓	✓	✓	✓
CON Neighb. Capitals	✓	✓	✓	✓
Controls in 2010 × Year FE	✓	✓	✓	✓

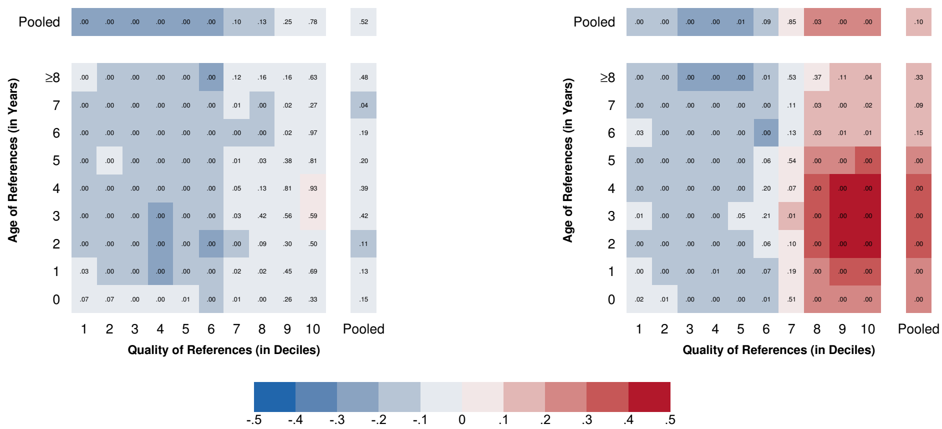
Note: The table displays regression results from Equation (IV2) for various reference measures. Across all panels, the sample is limited to years for which download data are available. Standard errors are clustered at the sub-national level. Significance levels are indicated as follows: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Heterogeneity by Quality and Age We have previously shown that the overall number of references is not affected by Sci-Hub. Instead, scientists switch from open- to more closed-access publications in their reference lists. Next, we ask, which exact types of works are being substituted. To answer this question, we break down all references along two dimensions, the quality deciles of their respective journal and the relative age of the publication (the difference in the publication year between a referenced article and the referencing article). We then run the baseline 2SLS regression on all these subgroups of different references. The estimates are plotted in Figure 2.8. We observe an intuitive but remarkable pattern: the positive effect on restricted-access references is highly concentrated in high-quality journals (top two deciles) and articles published most recently (two to four years ago). Once researchers learn of Sci-Hub, they start reading and referencing frontier research at much greater rates. On the other hand, Sci-Hub is associated with significant reductions in low-quality references – this can be reconciled by incomplete information about an article's relevance and quality before purchasing it. Prior to Sci-Hub, many scientists likely cited papers solely based on abstracts and titles.[26] Importantly, references to high-quality open-access publications remain unaffected. Hence, scientists appear not to unconsciously select restricted-access publications as references but start citing more high-quality work. Since most high-quality work is paywalled, we then, in turn, document substitution from open- to closed-access papers.

Figure 2.8: Change in Reference Dynamics by Age and Quality

(a) Open-access

(b) Restricted-access

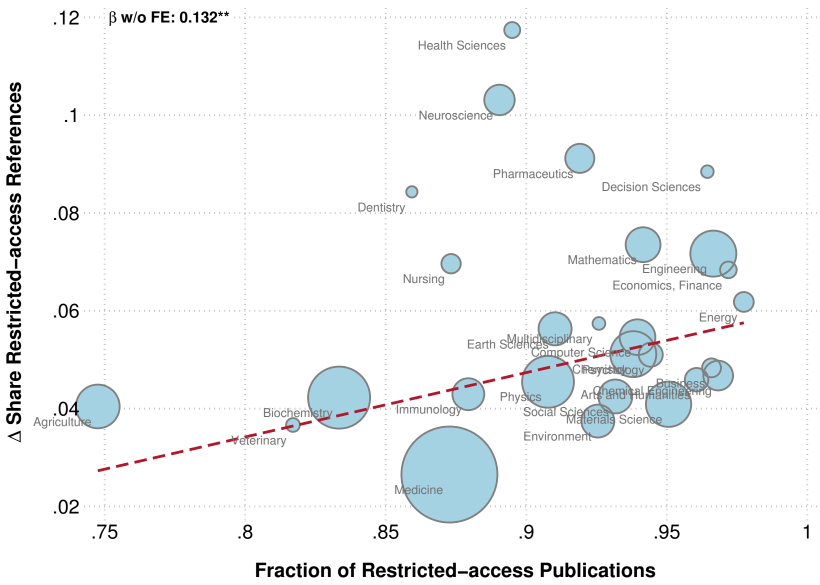


Note: The figure shows disaggregated 2SLS estimates for the number of open-access and restricted-access references according to the specification in Panel B of Table 2.3. Specifically, the number of references is disaggregated by age and quality of the referenced papers. The age corresponds to the year difference between the publication of the referencing paper and the referenced paper. Reference quality deciles are based on journal quality percentiles provided by Scopus, which are based on the average number of times a journal is cited per publication. Each tile represents a separate regression in which the dependent variable is the number of open access or restricted-access references of age a (indicated on the y-axis) and quality q (indicated on the x-axis). Effect sizes are indicated by color codes, with blue indicating a negative effect and red a positive effect. The p-value for each estimate is stated on top of each tile.

Heterogeneity by Field Fields differ in their prevalence of open- versus closed-access journals. We argue that these differences should moderate the impact of Sci-Hub. The intuition is that in fields where there are relatively more restricted-access publications, the pool of suitable closed-access references is also relatively larger. Hence, we should observe quantitatively larger effects in fields with ex-ante greater rates of restriction. We test this in our data. In particular, we estimate separate 2SLS regressions for different fields. In Figure 2.9, we show

[26] This is consistent with theory and evidence by McCabe and Snyder (2021).

Figure 2.9: Change in Reference Dynamics by Field



Note: The figure shows disaggregated 2SLS estimates for the share of restricted-access references as in Panel B of Table 2.3. Each scatter represents a separate regression in which the dependent variable is the share of restricted-access references in a field. Effect sizes are indicated on the vertical axis. The baseline share of open-access journals is displayed on the horizontal axis. The size of each scatter indicates the size of a field, measured by the total number of publications in 2010. A grey outline indicates that the estimate is significant at 5%. The red dotted line is the size-weighted correlation between the baseline share of restricted-access publications and the corresponding 2SLS estimate.

that the increase in the share of restricted-access references is particularly large in fields with higher restriction rates.[27] We confirm that this visual relationship is statistically significant, with the coefficient estimate $\hat{\gamma}_2 = 0.132$ differing significantly from zero (p-value$= 0.03$).[28]

[27] In Appendix Figure 2.B.9 we further show disaggregated 2SLS for the raw numbers of open- and restricted-access references.

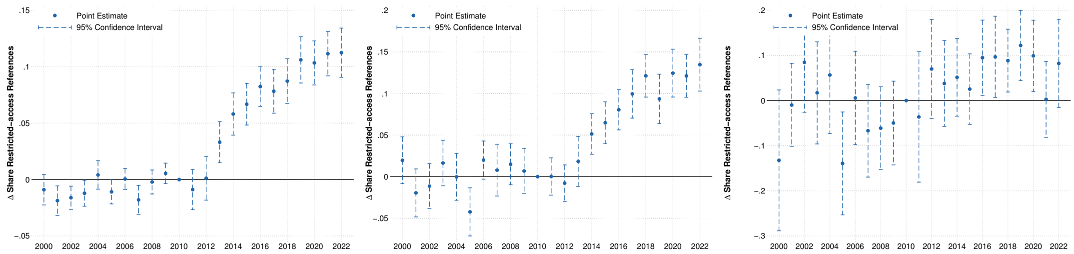
[28] To obtain correct standard errors, we interact Sci-Hub downloads with the field-specific ex-ante rates of access restriction. To implement this design, we construct a combined panel dataset by vertically stacking the time series data for each sub-national unit across all fields. The second-stage regression equation is as follows:

$$\begin{aligned}\text{Ref. Share RA}_{itf} = \alpha_i + \alpha_{c(i)t} + \\ &+ \beta_2 \ln \widehat{\text{Down}}_{it} + \sum_n \delta_2^{(n)} \ln \text{CON}_i^{\text{n}} \times \mathbb{1}_{t>2010} \\ &+ \gamma_2 \ln \widehat{\text{Down}}_{it} \cdot \text{Pub. Share RA}_{f2010} \\ &+ \boldsymbol{X}_{i2010} \boldsymbol{\phi}_t + \eta_{itf}\end{aligned}$$

Here, Ref. Share RA_{itf} represents the share of restricted-access references for field f in sub-national unit i at time t. Similarly, Pub. Share RA_{f2010} denotes the share of publications in field f that were under restricted-access in the pre-Sci-Hub baseline year 2010. All other variables are defined as in Equation (IV2).

Figure 2.10: Reduced Form – Change in References by Region

(a) Developed (b) Emerging (c) Developing



Note: The figure shows reduced form event study estimates for the effect of log connectedness to Almaty on the share of restricted-access references allowing for heterogeneity in developed, emerging, and developing regions. Standard errors are clustered by subnational region. Bars represent 95% confidence intervals.

Heterogeneity by Region Finally, we explore how Sci-Hub affects reference lists in different income regions. In Figure 2.10, we disaggregate the reduced form effect allowing for different responses in developed, emerging, and developing countries (Panels (a), (b) and (c)). We find that increases in the share of references to paywalled papers are driven by developed and emerging regions. Interestingly, the point estimates and dynamics are very similar in both regions, whereas they are absent in developing countries. In the latter, we see no measurable impact on the share of restricted-access references. Note, however, that our instrument lacks relevance for this subgroup (Appendix Table 2.A.6).

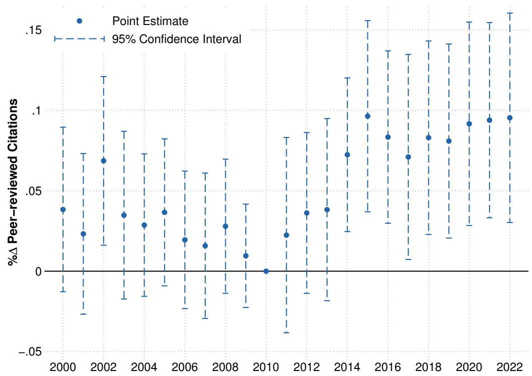
In Appendix Figure 2.B.8 we also show disaggregated results for the number of total references, restricted-access and open-access references. Notably, our analysis reveals that the factors driving the increase in the share of restricted-access references vary between developed and emerging economies. In developed regions, we see a level shift: reference lists get longer and presumably more holistic, due to an increase in the number of restricted-access references. Open-access references remain unaffected. However, in emerging regions, we observe a pattern of pure substitution where open-access references are replaced by restricted-access references without any increase in the total number of both.

2.5.3 Effects on Knowledge Production

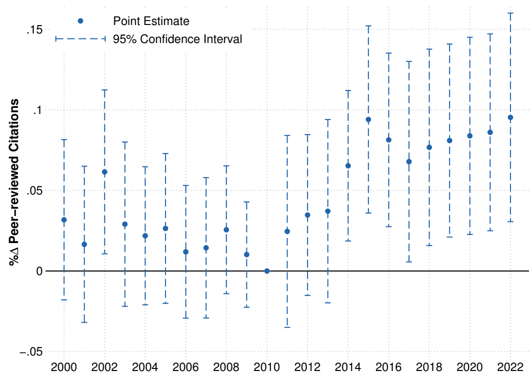
The evidence gathered so far documents that Sci-Hub has profoundly impacted what researchers read and reference. We next examine whether exposure to higher-quality articles, in turn, affects the creation of new scientific insights. To

Figure 2.11: Effects of Connectedness on Citations

(a) Total



(b) Peer-reviewed



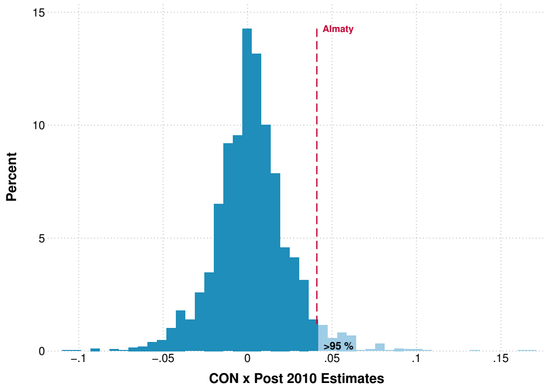
Note: The figure shows reduced form event study estimates for the outcomes and specification displayed in Table 2.4 Panel A. The post-2010 indicator is replaced with a full set of annual indicators, omitting 2010, the year before Sci-Hub was established. Standard errors are clustered by subnational region. Bars represent 95% confidence intervals.

answer that question, we estimate the effect of Sci-Hub on the creation of new scientific works.

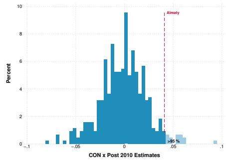
Citations First, we assess the effects of Sci-Hub on citations, a standard quality measure of scientific output. In Figure 2.11 we present reduced form estimates of the effect of connectedness to Almaty on the number of citations accruing to researchers in a given region. If access to frontier research leads to higher-quality works, we would expect increases in citations to regions with higher connectedness. Indeed, this is what we find. To interpret the magnitude, consider two similar regions with the exception that one has twice as many friendship links to Almaty as the other. Comparing publications published in 2010 versus 2015, papers from the higher-connected region see a differential increase in citations of almost 10%. We document similar effects and magnitudes both in the total number of citations, including from non-scholarly sources, and citations from only peer-reviewed journals. In Figure 2.12, we again show that it is specifically connectedness to Almaty that predicts increases in citations whereas other regions generally do not.

In Table 2.4 we show associated 2SLS estimates. Note that, in this exercise, we again lose a substantial fraction of the sample, namely 2014 and all years after 2017. Hence, we do not have sufficient power to reject the null hypothesis of no effect at the standard levels of statistical significance. Nonetheless, the estimate is helpful for interpreting the reduced form effect through the lens of Sci-Hub. On

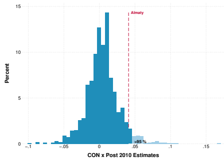
Figure 2.12: Placebo Effects of Connectedness on Citations



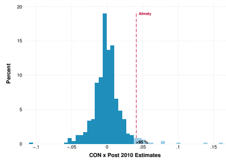
(a) All Regions



(b) Developed



(c) Emerging



(d) Developing

Note: Panel (a) shows the distribution of point-estimates when re-estimating the reduced form effect by iteratively replacing social connectedness to Almaty with social ties to all other sub-national units. We replace social ties to Almaty's neighboring capital regions with respective other neighboring capital regions. The outcome is the log-number of peer-reviewed citations. Panels (b) to (d) show the distribution of point-estimates within specific regions. Classification of sub-national units into developed, emerging, and developing regions is based on data by the International Monetary Fund (2011), and the United Nations (2011). In all figures, the dotted red line corresponds to the point estimate in Panel A column 3 of Table 2.4.

average, doubling Sci-Hub traffic is associated with roughly 12% more citations from peer-reviewed journals. When employing the two-sample 2SLS approach, we estimate an increase of roughly 14% significant at the 90% level (Appendix Table 2.C.3). We also test whether open-access elevates the probability of writing "home-run" papers, articles that reach the 95th or 99th percentile within a field's citation distribution. Yet, we do not find evidence for increases along this margin (Appendix Table 2.A.7).

Finally, we investigate heterogeneity by splitting the sample into regions of different economic development. In particular, we introduce interactions with indicator variables for developed, emerging, and developing countries with connectedness to Almaty. The estimates of the reduced form effect of connectedness on log-transformed citations are presented in Figure 2.13.[29] Allowing for heterogeneous effects, we find positive and significant increases in citations concentrated in high- and middle-income countries following 2011, but not in low-income countries. When we estimate the two-sample IV coefficients (Appendix Table 2.C.4), we find that in emerging regions doubling Sci-Hub downloads is associated with 12.6% more citations from peer-reviewed journals (significant at the 95% level). In high-income regions, we estimate an insignificant increase of 9.1% more citations, whereas

[29] Corresponding static estimates are shown in Appendix Table 2.A.8.

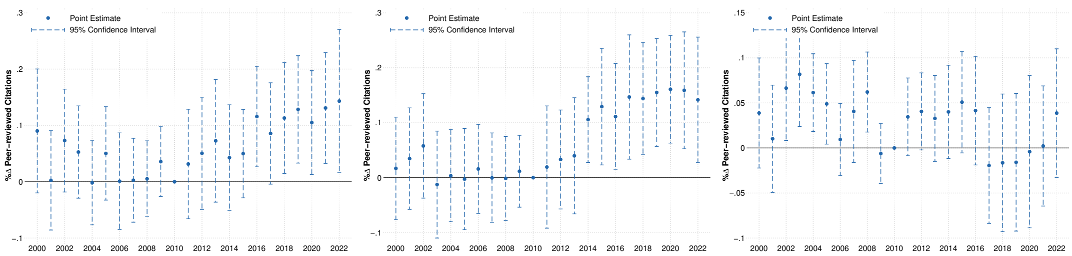
Table 2.4: Change in Citation Patterns

	Number of Citations (log-transformed)			
	Total	Non-peer-reviewed	Peer-reviewed	Cross-field
	(1)	(2)	(3)	(4)
Panel A: Reduced Form				
ln CON Almaty × Post 2010	0.036	0.007	0.041	-0.031
	(0.028)	(0.024)	(0.028)	(0.040)
Observations	40,440	40,440	40,440	40,440
Panel B: 2SLS				
ln Downloads	0.107	0.019	0.121	-0.092
	(0.083)	(0.072)	(0.083)	(0.118)
Observations	40,440	40,440	40,440	40,440
F-statistic	40.251	40.251	40.251	40.251
Panel C: OLS				
ln Downloads	-0.007	0.002	-0.006	-0.011
	(0.006)	(0.006)	(0.006)	(0.009)
Observations	40,440	40,440	40,440	40,440
Fixed Effects				
Sub-national	✓	✓	✓	✓
Country × Year	✓	✓	✓	✓
CON Neighb. Capitals	✓	✓	✓	✓
Controls in 2010 × Year FE	✓	✓	✓	✓

Note: The table displays regression results from Equation (IV2) for various log-transformed citation measures. Across all panels, the sample is limited to years for which download data are available. Standard errors are clustered at the sub-national level. Significance levels are indicated as follows: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Figure 2.13: Reduced Form – Effects on Citations by Region

(a) Developed (b) Emerging (c) Developing



Note: The figure shows reduced form event study estimates for the effect of log connectedness to Almaty on log-transformed citations allowing for heterogeneity in developed, emerging and developing regions. Standard errors are clustered by subnational region. Bars represent 95% confidence intervals.

Taken together, we interpret the results as evidence for quality increases in high Sci-Hub traffic regions. Yet, even if one remains agnostic about whether citations reflect quality: at a minimum, the results imply greater recognition of work from regions previously disadvantaged by access restrictions.

Number of Publications Next, we investigate potential increases in the number of publications. Loosening monetary constraints through free downloads may allow scientists to shift resources. Researchers may be able to hire more research assistants or purchase more scientific equipment. Such investments may then increase research output as measured by the number of publications. To test this idea, we estimate Equation (IV2) using the number of newly written articles in a given region as the main outcome. The corresponding estimates are shown in Table 2.5. Columns (1) and (2) show that we do not find any effects of Sci-Hub on the number of new publications (both peer and non-peer-reviewed articles). The estimates are relatively small and even slightly negative. Doubling Sci-Hub traffic is associated with an insignificant reduction in peer-reviewed publications by roughly 1%.

We also test for distributional shifts. If greater access transitions into better papers, we would not necessarily see more publications but a shift in the publishing outlets. To test this idea we classify publications based on journal quality quintiles measured by impact factors. The disaggregated results are shown in Table 2.5, where Columns (3)–(7) denote ascending quality quintiles. Again, we do not observe significant changes in the distribution of outlets in which researchers publish. If anything, we estimate slight reductions in newly written articles across the full journal-ranking spectrum.

Table 2.5: Change in Publication Patterns

	Total	Peer-reviewed	By Journal Quality (in Quintiles) Q1	Q2	Q3	Q4	Q5
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A: Reduced Form							
ln CON Almaty × Post 2010	-0.007	-0.001	-0.009	-0.009	-0.006	-0.002	-0.008
	(0.026)	(0.025)	(0.014)	(0.015)	(0.019)	(0.021)	(0.021)
Observations	40,440	40,440	40,440	40,440	40,440	40,440	40,440
Panel B: 2SLS							
ln Downloads	-0.020	-0.004	-0.026	-0.027	-0.019	-0.006	-0.023
	(0.077)	(0.074)	(0.041)	(0.045)	(0.056)	(0.061)	(0.061)
Observations	40,440	40,440	40,440	40,440	40,440	40,440	40,440
F-statistic	40.251	40.251	40.251	40.251	40.251	40.251	40.251
Panel C: OLS							
ln Downloads	-0.005	-0.004	0.019***	0.010**	0.004	-0.002	0.003
	(0.006)	(0.006)	(0.004)	(0.005)	(0.005)	(0.005)	(0.005)
Observations	40,440	40,440	40,440	40,440	40,440	40,440	40,440
Fixed Effects							
Sub-national	✓	✓	✓	✓	✓	✓	✓
Country × Year	✓	✓	✓	✓	✓	✓	✓
CON Neighb. Capitals	✓	✓	✓	✓	✓	✓	✓
Controls in 2010 × Year FE	✓	✓	✓	✓	✓	✓	✓

Note: The table displays regression results from Equation (IV2) for various publication measures. Across all panels, the sample is limited to years for which download data are available. Standard errors are clustered at the sub-national level. Significance levels are indicated as follows: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Direction of Research Finally, we attempt to gauge changes in the direction of research. Does scientific activity shift toward topics studied at the research frontier? To test this, we train a topic model using abstracts from papers in the top percentile of the citation distribution for each year and field. The trained model is then applied to predict the topic distributions of all other – previously unseen – papers within the same field and year. We then calculate the similarity between a paper's topic distribution and that of the research frontier using the Mahalanobis distance. These distances are then aggregated across regions and years. Details on the construction of this measure of *frontier distance* are provided in Appendix Section 2.D.2. The estimates, using this measure as the outcome variable, are presented in Table 2.6. Contrary to expectations, we do not observe a convergence of research topics toward the research frontier.

Table 2.6: Similarity to Research Frontier Topic Distribution

	Across Field Similarity	Within Field Similarity (... Sciences)			
		Social	Health	Life	Physical
	(1)	(2)	(3)	(4)	(5)
Panel A: Reduced Form					
ln CON Almaty × Post 2010	0.084*	0.009	-0.015	-0.006	-0.031
	(0.048)	(0.025)	(0.024)	(0.027)	(0.020)
Observations	21,585	21,585	21,585	21,585	21,585
Panel B: 2SLS					
ln Downloads	0.087*	0.009	-0.015	-0.007	-0.032
	(0.051)	(0.027)	(0.024)	(0.028)	(0.023)
Observations	21,585	21,585	21,585	21,585	21,585
F-statistic	27.363	27.363	27.363	27.363	27.363
Panel C: OLS					
ln Downloads	0.003	-0.004	0.001	-0.001	0.002
	(0.004)	(0.003)	(0.002)	(0.003)	(0.001)
Observations	21,585	21,585	21,585	21,585	21,585
Fixed Effects					
Sub-national	✓	✓	✓	✓	✓
Country × Year	✓	✓	✓	✓	✓
CON Neighb. Capitals	✓	✓	✓	✓	✓
Controls in 2010 × Year FE	✓	✓	✓	✓	✓

Note: The table displays regression results from Equation (IV2) for our measure of frontier distance. Across all panels, the sample is limited to years for which download data are available. Standard errors are clustered at the sub-national level. Significance levels are indicated as follows: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

2.5.4 Effects on Migration and Innovation

Thus far, we have documented the impact of Sci-Hub on various scientific outcomes. However, other dimensions, such as scientist migration and industrial innovation, may also be affected. First, we investigate whether improved publication quality expands opportunities or incentives for scientist migration. Second, we examine potential spillover effects of increased access on patent citations to scientific articles.

Migration To investigate migration, we rely on changes in the affiliation of scientists in OpenAlex. We classify moves among research institutions along three dimensions: whether a researcher changes her affiliation within the same subnational region, changes affiliation within the same country, or moves to a developed country. The results are shown in Table 2.7. First, we find no effect of Sci-Hub traffic on the total number of researchers in a region. Second, we investigate outflows based on where the receiving institution is situated. We find no effect on moves within a subnational region, to other institutes and universities in the country, or to universities in developed regions.

Table 2.7: Migration Patterns

	Stock of Researchers	Outflows		
		Subnational	Country	Developed
	(1)	(2)	(3)	(4)
Panel A: Reduced Form				
ln CON Almaty × Post 2010	0.000	0.014	0.003	0.007
	(0.022)	(0.012)	(0.010)	(0.009)
Observations	40,440	40,440	40,440	40,440
Panel B: 2SLS				
ln Downloads	0.000	0.041	0.008	0.020
	(0.065)	(0.035)	(0.029)	(0.028)
Observations	40,440	40,440	40,440	40,440
F-statistic	40.251	40.251	40.251	40.251
Panel C: OLS				
ln Downloads	-0.005	0.004	0.005	0.008**
	(0.004)	(0.003)	(0.003)	(0.003)
Observations	40,440	40,440	40,440	40,440
Fixed Effects				
Sub-national	✓	✓	✓	✓
Country × Year	✓	✓	✓	✓
CON Neighb. Capitals	✓	✓	✓	✓
Controls in 2010 × Year FE	✓	✓	✓	✓

Note: The table displays regression results from Equation (IV2) for various migration measures. Across all panels, the sample is limited to years for which download data are available. Standard errors are clustered at the sub-national level. Significance levels are indicated as follows: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

It is important to note that we did not conduct a large-scale linking exercise for scientists, and therefore our findings may not provide a complete picture of migration opportunities. Specifically, if a scientist changes universities and is assigned a new identifier in OpenAlex, our data analysis may not accurately reflect their migration choices. Hence, we caution readers to keep this limitation in mind when interpreting these results.

Innovation Prior evidence documents that articles published under open-access accrue more patent citations – put differently, spillovers from science to industry are more pronounced (Bryan and Ozcan, 2021). We test whether the spread of Sci-Hub has similarly impacted industry-use of scientific insights along two dimensions. First, we test whether patents from regions with high Sci-Hub usage cite more restricted-access publications. Second, we check whether publications from these regions receive more patent citations.[30]

In Table 2.8 we find no measurable impact of Sci-Hub on the share of restricted-access references in patents. Likewise, we do not detect distributional shifts in references to higher-quality scientific publications. In Table 2.9 we also find no indication that publications from high-intensity Sci-Hub regions attract more patent citations.

There are several plausible explanations that can rationalize these results. First, the analysis is built on relatively 'scarce' data. We only observe patents from OECD countries, as opposed to the near-universe of publications. Secondly, firms' budget-constraints may be less binding than those of individual researchers. In all likelihood, paywall fees take up a far smaller share of research budgets within firms and industrial R&D labs than among scientists. We interpret the fact that references to open-access journals take up only about 5% of all references as evidence in favor of this argument.[31]

[30] We discuss data construction and results in greater detail in Appendix 2.D.1.
[31] See Appendix Table 2.A.10.

Table 2.8: Patents - Share of Restricted-access References

	Share Restricted-access References					
	All	Quality of References				
		Q1	Q2	Q3	Q4	Q5
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Reduced Form						
ln CON Almaty × Post 2010	0.016	0.279**	0.002	-0.005	0.030	-0.011
	(0.018)	(0.139)	(0.077)	(0.065)	(0.036)	(0.021)
Observations	5,280	1,651	2,641	3,388	4,288	4,961
Panel B: 2SLS						
ln Downloads	0.012	0.127*	0.001	-0.003	0.021	-0.008
	(0.014)	(0.073)	(0.043)	(0.042)	(0.028)	(0.015)
Observations	5,280	1,651	2,641	3,388	4,288	4,961
F-statistic	13.839	10.929	9.259	10.834	9.994	14.126
Panel C: OLS						
ln Downloads	0.002	-0.006	-0.002	-0.001	-0.000	-0.003*
	(0.003)	(0.008)	(0.006)	(0.005)	(0.004)	(0.002)
Observations	5,280	1,651	2,641	3,388	4,288	4,961
Fixed Effects						
Sub-national	✓	✓	✓	✓	✓	✓
Country × Year	✓	✓	✓	✓	✓	✓
CON Neighb. Capitals	✓	✓	✓	✓	✓	✓
Controls in 2010 × Year FE	✓	✓	✓	✓	✓	✓

Note: The table displays regression results from Equation (IV2) for the share of restricted-access references in patents among various groups. Across all panels, the sample is limited to years for which download data are available. Standard errors are clustered at the sub-national level. Significance levels are indicated as follows: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 2.9: Patent to Publication Citations

	Total	Peer-reviewed	Quality of Cited Reference				
			Q1	Q2	Q3	Q4	Q5
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A: Reduced Form							
ln CON Almaty × Post 2010	-0.011	-0.010	-0.012**	-0.004	-0.001	0.001	-0.004
	(0.012)	(0.012)	(0.006)	(0.007)	(0.008)	(0.009)	(0.011)
Observations	40,440	40,440	40,440	40,440	40,440	40,440	40,440
Panel B: 2SLS							
ln Downloads	-0.032	-0.028	-0.035**	-0.013	-0.003	0.002	-0.013
	(0.035)	(0.035)	(0.017)	(0.020)	(0.025)	(0.027)	(0.033)
Observations	40,440	40,440	40,440	40,440	40,440	40,440	40,440
F-statistic	40.251	40.251	40.251	40.251	40.251	40.251	40.251
Panel C: OLS							
ln Downloads	-0.004	-0.004	-0.001	0.006**	0.009***	0.002	-0.003
	(0.004)	(0.004)	(0.002)	(0.003)	(0.003)	(0.003)	(0.004)
Observations	40,440	40,440	40,440	40,440	40,440	40,440	40,440
Fixed Effects							
Sub-national	✓	✓	✓	✓	✓	✓	✓
Country × Year	✓	✓	✓	✓	✓	✓	✓
CON Neighb. Capitals	✓	✓	✓	✓	✓	✓	✓
Controls in 2010 × Year FE	✓	✓	✓	✓	✓	✓	✓

Note: The table displays regression results from Equation (IV2) for the log-number of patent to publication citations. Patent citations are restricted to in-text citations referenced by inventors. Across all panels, the sample is limited to years for which download data are available. Standard errors are clustered at the sub-national level. Significance levels are indicated as follows: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

2.6 Conclusion

This paper studies the rise of Sci-Hub, an academic file-sharing website. Using a wealth of data sources, we build a global panel of scientific input and output at the sub-national level that spans two decades. In an instrumented difference-in-differences framework, we show that Sci-Hub has meaningfully shifted global knowledge consumption and production.

Our analysis suggests three tentative lessons about the impact of open access on knowledge creation. First, regions exposed to Sci-Hub see a quantitatively significant rise in the share of references to restricted-access publications. In particular, researchers substitute low-quality references with previously closed-access articles at the research frontier. Second, we document that greater exposure to frontier research has resulted in the production of novel works with higher citation rates – a common measure of scientific quality. Third, we do not find evidence for changing research directions, improved journal outcomes, migration opportunities, or greater industry use of scientific insights being majorly affected. Nonetheless, taken together, our results suggest that open-access research is likely an underprovided public good within academic research. With a slowdown in disruptive science (Park, Leahey and Funk, 2023), the policy takeaway is clear: Governments and funders should continue to actively implement measures reducing closed-access rates.

Bibliography

Adena, Maja, Ruben Enikolopov, Maria Petrova, Veronica Santarosa, and Ekaterina Zhuravskaya. 2015. "Radio and the Rise of the Nazis in Prewar Germany." *Quarterly Journal of Economics*, 130(4): 1885–1939.

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

Ahmadpoor, Mohammad, and Benjamin F Jones. 2017. "The Dual Frontier: Patented Inventions and Prior Scientific Advance." *Science*, 357(6351): 583–587.

Angrist, Joshua D, and Alan B Krueger. 1992. "The Effect of Age at School Entry on Educational Attainment: An Application of Instrumental Variables with Moments from Two Samples." *Journal of the American Statistical Association*, 87(418): 328–336.

Arthur, W Brian. 1989. "Competing Technologies, Increasing Returns, and Lock-in by Historical Events." *Economic Journal*, 99(394): 116–131.

Azoulay, Pierre, Joshua S. Graff Zivin, and Jialan Wang. 2010. "Superstar Extinction." *Quarterly Journal of Economics*, 125(2): 549–589.

Bailey, Michael, Abhinav Gupta, Sebastian Hillenbrand, Theresa Kuchler, Robert Richmond, and Johannes Stroebel. 2021. "International Trade and Social Connectedness." *Journal of International Economics*, 129: 103418.

Bailey, Michael, Rachel Cao, Theresa Kuchler, Johannes Stroebel, and Arlene Wong. 2018. "Social Connectedness: Measurement, Determinants, and Effects." *Journal of Economic Perspectives*, 32(3): 259–80.

Bergstrom, Carl T, and Theodore C Bergstrom. 2004. "The Costs and Benefits of Library Site Licenses to Academic Journals." *Proceedings of the National Academy of Sciences*, 101(3): 897–902.

Biasi, Barbara, and Petra Moser. 2021. "Effects of Copyrights on Science: Evidence from the WWII Book Republication Program." *American Economic Journal: Microeconomics*, 13(4): 218–260.

Bohannon, John. 2016. "Who's Downloading Pirated Papers? Everyone." *Science*, 352(6285): 508–512.

Bryan, Kevin A, and Yasin Ozcan. 2021. "The Impact of Open Access Mandates on Invention." *Review of Economics and Statistics*, 103(5): 954–967.

Bursztyn, Leonardo, Georgy Egorov, Ruben Enikolopov, and Maria Petrova. 2019. "Social Media and Xenophobia: Evidence from Russia." National Bureau of Economic Research.

Cagé, Julia, Nicolas Hervé, Béatrice Mazoyer, et al. 2022. "Social Media Influence Mainstream Media: Evidence from Two Billion Tweets." HAL.

Card, David, and Stefano DellaVigna. 2020. "What Do Editors Maximize? Evidence from Four Economics Journals." *Review of Economics and Statistics*, 102(1): 195–217.

Chen, Jiafeng, and Jonathan Roth. 2022. "Log-like? Identified ATEs Defined with Zero-Valued Outcomes Are (Arbitrarily) Scale-Dependent." *Working Paper*.

CIESIN, Center for International Earth Science Information Network (Columbia University). 2020. "Gridded Population of the World, Version 4 (GPWv4): Population Count, Revision 11t." *URL:* *https://sedac.ciesin.columbia.edu/data/set/gpw-v4-population-count-rev11*, Accessed: 2021-10-10.

Davis, Philip M. 2011. "Open Access, Readership, Citations: A Randomized Controlled Trial of Scientific Journal Publishing." *The FASEB Journal*, 25(7): 2129–2134.

Davis, Philip M, Bruce V Lewenstein, Daniel H Simon, James G Booth, and Mathew JL Connolly. 2008. "Open Access Publishing, Article Downloads, and Citations: Randomised Controlled Trial." *BMJ*, 337.

DellaVigna, Stefano, and Ethan Kaplan. 2007. "The Fox News Effect: Media Bias and Voting." *Quarterly Journal of Economics*, 122(3): 1187–1234.

Djourelova, Milena, Ruben Durante, and Gregory Martin. 2021. "The Impact of Online Competition on Local Newspapers: Evidence from the Introduction of Craigslist." 16130.

Durante, Ruben, Paolo Pinotti, and Andrea Tesei. 2019. "The Political Legacy of Entertainment TV." *American Economic Review*, 109(7): 2497–2530.

Eastern District Court of Virginia, United States District Court. 2017. "Civil Action No. 1:l7cv0726 (LMB/JFA)." *URL:* *https://www.infodocket.*

com/wp-content/uploads/2017/10/18918321195.pdf, Accessed: 2021-10-14.

Elbakyan, Alexandra. 2017. "Some Facts on Sci-Hub that Wikipedia Gets Wrong." *URL: https://engineuring.wordpress.com/2017/07/02/some-facts-on-sci-hub-that-wikipedia-gets-wrong/*, Accessed: 2022-11-30.

Enikolopov, Ruben, Alexey Makarin, and Maria Petrova. 2020. "Social Media and Protest Participation: Evidence from Russia." *Econometrica*, 88(4): 1479–1514.

Enikolopov, Ruben, Maria Petrova, and Ekaterina Zhuravskaya. 2011. "Media and Political Persuasion: Evidence from Russia." *American Economic Review*, 101(7): 3253–85.

Falck, Oliver, Robert Gold, and Stephan Heblich. 2014. "E-lections: Voting Behavior and the Internet." *American Economic Review*, 104(7): 2238–65.

Fiorini, Nicolas, David J Lipman, and Zhiyong Lu. 2017. "Cutting Edge: Towards PubMed 2.0." *eLife*, 6: e28801.

GADM1 Version 2.8, University of Berkeley. 2015. "GADM Database of Global Administrative Areas, Version 2.8." *URL: https://gadm.org/old_versions.html*.

Gentzkow, Matthew. 2006. "Television and Voter Turnout." *Quarterly Journal of Economics*, 121(3): 931–972.

Gerrish, Sean, and David M Blei. 2010. "A Language-Based Approach to Measuring Scholarly Impact." Vol. 10, 375–382.

Guriev, Sergei, Nikita Melnikov, and Ekaterina Zhuravskaya. 2021. "3G Internet and Confidence in Government." *Quarterly Journal of Economics*, 136(4): 2533–2613.

Hager, Sebastian, Carlo Schwarz, and Fabian Waldinger. 2023. "Measuring Science: Performance Metrics and the Allocation of Talent." *Available at SSRN 4472441*.

Hill, Ryan, and Carolyn Stein. 2021. "Race to the Bottom: Competition and Quality in Science." Working Paper.

Himmelstein, Daniel S, Ariel Rodriguez Romero, Jacob G Levernier, Thomas Anthony Munro, Stephen Reid McLaughlin, Bastian Greshake Tzovaras, and Casey S Greene. 2018. "Sci-Hub Provides Access to Nearly All Scholarly Literature." *eLife*, 7: e32822.

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

Inoue, Atsushi, and Gary Solon. 2010. "Two-sample Instrumental Variables Estimators." *Review of Economics and Statistics*, 92(3): 557–561.

International Monetary Fund, IMF. 2011. "World Economic Outlook Database – WEO Update: June 17, 2011." *URL:* `https://www.imf.org/en/Publications/WEO/weo-database/2011/April`, Accessed: 2022-01-03.

Jeon, Doh-Shin, and Domenico Menicucci. 2006. "Bundling Electronic Journals and Competition among Publishers." *Journal of the European Economic Association*, 4(5): 1038–1083.

Jia, Ruixue, Margaret E Roberts, Ye Wang, and Eddie Yang. 2022. "The Impact of US-China Tensions on US Science." *National Bureau of Economic Research*, 29941.

Jones, Charles I. 1995. "R&D-Based Models of Economic Growth." *Journal of Political Economy*, 103(4): 759–784.

Kuchler, Theresa, Dominic Russel, and Johannes Stroebel. 2022. "The Geographic Spread of COVID-19 Correlates with the Structure of Social Networks as Measured by Facebook." *Journal of Urban Economics*, 127: 103314.

Langham-Putrow, Allison, Caitlin Bakker, and Amy Riegelman. 2021. "Is the Open Access Citation Advantage Real? A Systematic Review of the Citation of Open Access and Subscription-Based Articles." *PLoS One*, 16(6): e0253129.

Li, Xuecao, Yuyu Zhou, Min Zhao, and Xia Zhao. 2020. "A Harmonized Global Nighttime Light Dataset 1992–2018." *Scientific Data*, 7(1): 1–9.

Martín-Martín, Alberto, Mike Thelwall, Enrique Orduna-Malea, and Emilio Delgado López-Cózar. 2021. "Google Scholar, Microsoft Academic, Scopus, Dimensions, Web of Science, and OpenCitations' COCI: A Multidisciplinary Comparison of Coverage via Citations." *Scientometrics*, 126(1): 871–906.

McCabe, Mark J. 2002. "Journal Pricing and Mergers: A Portfolio Approach." *American Economic Review*, 92(1): 259–269.

McCabe, Mark J, and Christopher M Snyder. 2005. "Open Access and Academic Journal Quality." *American Economic Review*, 95(2): 453–459.

McCabe, Mark J, and Christopher M Snyder. 2014. "Identifying the Effect of Open Access on Citations Using a Panel of Science Journals." *Economic Inquiry*, 52(4): 1284–1300.

McCabe, Mark J, and Christopher M Snyder. 2021. "Cite Unseen: Theory and Evidence on the Effect of Open Access on Cites to Academic Articles Across the Quality Spectrum." *Managerial and Decision Economics*, 42(8): 1960–1979.

Mokyr, Joel. 2011. *The Gifts of Athena: Historical Origins of the Knowledge Economy.* Princeton University Press.

Mullahy, John, and Edward C Norton. 2022. "Why Transform Y? A Critical Assessment of Dependent-Variable Transformations in Regression Models for Skewed and Sometimes-Zero Outcomes." *NBER Working Paper*, 30735.

Müller, Karsten, and Carlo Schwarz. 2021. "Fanning the Flames of Hate: Social Media and Hate Crime." *Journal of the European Economic Association*, 19(4): 2131–2167.

Müller, Karsten, and Carlo Schwarz. 2023. "From Hashtag to Hate Crime: Twitter and Antiminority Sentiment." *American Economic Journal: Applied Economics*, 15(3): 270–312.

Murray, Fiona, Philippe Aghion, Mathias Dewatripont, Julian Kolev, and Scott Stern. 2016. "Of Mice and Academics: Examining the Effect of Openness on Innovation." *American Economic Journal: Economic Policy*, 8(1): 212–52.

NUTS2, Eurostat. 2018. "Regions in the European Union – Nomenclature of Territorial Units for Statistics NUTS 2016/EU-28." In *Manuals and Guidelines – General and Regional Statistics*. Publications Office of the European Union.

Park, Michael, Erin Leahey, and Russell J Funk. 2023. "Papers and Patents Are Becoming Less Disruptive over Time." *Nature*, 613(7942): 138–144.

Romer, Paul M. 1990. "Endogenous Technological Change." *Journal of Political Economy*, 98(5): S71–S102.

Sample, Ian. 2012. "Harvard University Says It Can't Afford Journal Publishers' Price." *URL:* *https://www.theguardian.com/science/2012/apr/24/harvard-university-journal-publishers-prices*, Accessed: 2021-01-15.

Scheidsteger, Thomas, and Robin Haunschild. 2022. "Comparison of Metadata with Relevance for Bibliometrics between Microsoft Academic Graph and OpenAlex until 2020." *arXiv Pre-print*, 2206.14168.

Seamans, Robert, and Feng Zhu. 2014. "Responses to Entry in Multi-sided Markets: The Impact of Craigslist on Local Newspapers." *Management Science*, 60(2): 476–493.

Segado-Boj, Francisco, Juan Martín-Quevedo, and Juan-José Prieto-Gutiérrez. 2022. "Jumping over the Paywall: Strategies and Motivations for Scholarly Piracy and Other Alternatives." *Information Development*, 02666669221144429.

Stoy, Lennart, Rita Morais, and Lidia Borrell-Damián. 2019. "Decrypting the Big Deal Landscape: Follow-up of the 2019 EUA Big Deals Survey Report." *URL:* *https://eua.eu/downloads/publications/2019%20big%20deals%20report.pdf*, Accessed: 2021-01-17.

Strömberg, David. 2004. "Radio's Impact on Public Spending." *Quarterly Journal of Economics*, 119(1): 189–221.

Teplitskiy, Misha, Eamon Duede, Michael Menietti, and Karim R Lakhani. 2022. "How Status of Research Papers Affects the Way They Are Read and Cited." *Research Policy*, 51(4): 104484.

United Nations, UN. 2011. "The Least Developed Countries Report, 2011." *URL:* *https://unctad.org/system/files/official-document/ldc2011_en.pdf*, Accessed: 2022-01-03.

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

Williams, Heidi L. 2013. "Intellectual Property Rights and Innovation: Evidence from the Human Genome." *Journal of Political Economy*, 121(1): 1–27.

Yanagizawa-Drott, David. 2014. "Propaganda and Conflict: Evidence from the Rwandan Genocide." *Quarterly Journal of Economics*, 129(4): 1947–1994.

Yin, Yian, Yuxiao Dong, Kuansan Wang, Dashun Wang, and Benjamin Jones. 2021. "Science as a Public Good: Public Use and Funding of Science." *NBER Working Paper*, 28748.

2.A Additional Tables

Table 2.A.1: Sci-Hub and Social Connectedness – Summary Statistics

	Mean	SD	Min	Max	N
	(1)	(2)	(3)	(4)	(5)
Panel A: Sci-Hub Downloads					
Total (in 1,000s)	2.76	24.79	0.00	1,169.48	14,537
Total 2011 (in 1,000s)	0.01	0.15	0.00	4.61	2,437
Total 2012 (in 1,000s)	0.39	3.51	0.00	122.83	2,437
Total 2013 (in 1,000s)	0.43	2.69	0.00	62.68	2,437
Total 2015 (in 1,000s)	1.62	8.96	0.00	235.02	2,437
Total 2016 (in 1,000s)	1.06	7.29	0.00	230.98	2,437
Total 2017 (in 1,000s)	12.95	58.21	0.00	1,169.48	2,437
Per Institute	216.86	1334.49	0.00	44,279.00	9,329
Per Researcher	4.13	16.01	0.00	198.73	8,277
Panel B: Social Connectedness Index (in 1,000s)					
Almaty (KAZ)	0.43	5.47	0.00	210.27	2,437
Kazahkstan (KAZ)	0.63	6.85	0.00	114.22	2,437
Kazahkstan excl. Almaty (KAZ)	0.68	7.64	0.00	130.65	2,437
Nur-Sultan (KAZ)	0.65	9.54	0.00	344.26	2,437
Bishkek (KGZ)	1.33	21.45	0.00	491.63	2,437
Ashgabat (TKM)	5.40	129.01	0.00	4,362.62	2,437
Tashkent (UZB)	0.98	12.73	0.00	338.63	2,437
Dushanbe (TJK)	1.95	41.59	0.00	1,324.88	2,437
Kyiv (UKR)	0.47	5.66	0.00	201.34	2,437
Ulaanbaatar (MNG)	5.99	63.32	0.00	869.56	2,437

Note: In Panel A the table provides summary statistics for Sci-Hub downloads across our observation period. Panel B provides summary statistics for the Social Connectedness Index for Almaty, Kazakhstan, and Central Asian capitals.

Table 2.A.2: Publication Measures – Summary Statistics

	Mean	SD	Min	Max	N
	(1)	(2)	(3)	(4)	(5)
Panel A: Research Institutes					
Any	0.64	0.48	0.00	1.00	56,051
Total	18.70	79.69	0.00	2,641.00	56,051
Total $\geq$ 95th Percentile	0.72	5.11	0.00	195.00	56,051
Panel B: Researchers					
Researchers (in 1,000s)	1.25	5.36	0.00	189.97	56,051
Per Institute	50.79	92.34	0.00	2,731.50	36,087
Panel C: Publications					
Total (in 1,000s)	1.98	8.92	0.00	295.30	56,051
Per Institute	73.23	127.90	0.00	2,997.00	36,087
Per Researcher	1.53	0.82	0.00	28.00	30,105
Share Peer-reviewed	0.67	0.24	0.00	1.00	30,103
Share Restricted-access	0.56	0.25	0.00	1.00	30,103
Panel D: References					
Total (in 1,000s)	48.60	242.04	0.00	9,457.02	56,051
Per Institute	1580.89	3411.23	0.00	105,389.00	36,087
Per Researcher	25.85	20.22	0.00	484.00	30,105
Per Publication	16.85	10.08	0.00	228.00	30,103
Share Peer-reviewed	0.85	0.19	0.00	1.00	29,114
Share Restricted-Access	0.68	0.15	0.00	1.00	29,114
Panel E: Citations					
Total (in 1,000s)	40.75	219.04	0.00	6,133.67	56,051
Per Institute	1092.33	2376.46	0.00	39,514.20	36,087
Per Researcher	22.58	32.20	0.00	940.50	30,105
Per Publication	14.47	18.61	0.00	536.50	30,103
Share Peer-reviewed	0.94	0.09	0.00	1.00	53,287
Share Cross-citations	0.29	0.32	0.00	2.48	53,287

Note: The table provides summary statistics for research measures retrieved through OpenAlex and described in Section 2.3. In particular, Panels A and B show summary metrics for the number of research institutes and researchers in sub-national units. Panels C, D, and E summarize various publication, citation, and reference measures. Across all variables, the unit of observation is sub-national units from 2000 to 2022.

Table 2.A.3: Control Variables – Summary Statistics

	Mean	SD	Min	Max	N
	(1)	(2)	(3)	(4)	(5)
Panel A: Education					
Any Research Institute	0.64	0.48	0.00	1.00	2,437
Research Institutes, 2010	18.99	80.23	0.00	2,253.00	2,437
Research Institutes $\geq$ 95th Percentile, 2010	0.88	5.49	0.00	188.00	2,437
Researchers (in 1,000s), 2010	1.31	5.36	0.00	110.82	2,437
Panel B: Geography					
Capital	0.08	0.27	0.00	1.00	2,437
Area (in 10,000 km^2)	8.81	89.73	0.00	3,493.19	2,437
Latitude	17.06	22.76	-53.80	71.78	2,437
Longitude	21.25	67.84	-176.22	177.98	2,437
Distance to Almaty (in 1,000 km)	7.38	3.81	0.00	17.72	2,437
Panel C: Population					
Population (Million), 2010	2.11	7.57	0.00	204.35	2,437
Population Density (per km^2), 2010	0.43	2.05	0.00	41.28	2,437
Panel D: Development					
GDP* (USD Billion), 2010	25.48	68.04	0.00	1,004.07	2,437
GDP* per Capita (USD), 2010	21.76	185.37	0.00	8,910.61	2,437

Note: The table provides summary statistics for all control variables in Section 2.3. Time-varying variables are fixed in 2010.

Table 2.A.4: Extensive Margin Effects of Sci-Hub Downloads

	Dependent Variable: Any Publication					
	(1)	(2)	(3)	(4)	(5)	(6)
Any Download	0.525***	0.342***	0.010	0.009	0.009	0.009
	(0.034)	(0.029)	(0.006)	(0.006)	(0.006)	(0.006)
Observations	2,799	2,735	2,735	2,735	2,735	2,735
Number of Clusters	222	158	158	158	158	158
Fixed Effects						
Country	-	✓	✓	✓	✓	✓
Controls in 2010						
Education	-	-	✓	✓	✓	✓
Geography	-	-	-	✓	✓	✓
Population	-	-	-	-	✓	✓
Development	-	-	-	-	-	✓

Note: The table displays the results from regressing an indicator for having procured any research (until 2022) on an indicator for having any Sci-Hub download (until 2022). Standard errors are clustered at the sub-national level. Significance levels are indicated as follows: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 2.A.5: First Stage – Inverse Hyperbolic Sine Transformation

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
ihs CON Almaty × Post 2010	0.596***	0.613***	0.823***	0.507***	0.323***	0.331***	0.379***	0.378***
	(0.062)	(0.059)	(0.132)	(0.104)	(0.075)	(0.076)	(0.082)	(0.081)
Observations	41,344	41,344	40,444	40,444	40,444	40,444	40,444	40,444
Number of Clusters	195	195	142	142	142	142	142	142
F-statistic	92.375	107.196	38.627	23.964	18.541	18.806	21.228	21.745
Fixed Effects								
Sub-national	-	✓	✓	✓	✓	✓	✓	✓
Year × Country	-	-	✓	✓	✓	✓	✓	✓
CON Neighboring Capitals	-	-	-	✓	✓	✓	✓	✓
Controls in 2010 × Year FE								
Education	-	-	-	-	✓	✓	✓	✓
Geography	-	-	-	-	-	✓	✓	✓
Population	-	-	-	-	-	-	✓	✓
Development	-	-	-	-	-	-	-	✓

Note: The table displays regression results from Equation (IV1) across various specifications using the inverse hyperbolic sine transformation. Standard errors are clustered at the sub-national level. Significance levels are indicated as follows: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 2.A.6: First Stage Estimates by Region

	Dependent Variable: ln Downloads							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
ln CON Almaty × Post 2010 × Developed	1.217***	1.257***	1.624***	1.344***	0.414***	0.473***	0.567***	0.638***
	(0.026)	(0.026)	(0.160)	(0.172)	(0.116)	(0.123)	(0.118)	(0.121)
ln CON Almaty × Post 2010 × Emerging	0.539***	0.575***	0.960***	0.768***	0.522***	0.527***	0.576***	0.569***
	(0.025)	(0.024)	(0.150)	(0.148)	(0.088)	(0.089)	(0.091)	(0.090)
ln CON Almaty × Post 2010 × Developing	0.088***	0.138***	0.073**	-0.061	0.040	0.041	0.059*	0.049
	(0.016)	(0.016)	(0.030)	(0.045)	(0.033)	(0.033)	(0.033)	(0.032)
Observations	41,341	41,341	40,440	40,440	40,440	40,440	40,440	40,440
Number of Clusters	2,437	2,437	2,384	2,384	2,384	2,384	2,384	2,384
F-statistic	846.423	995.093	50.187	32.212	14.160	14.636	18.184	19.712
Fixed Effects								
Sub-national	-	✓	✓	✓	✓	✓	✓	✓
Year × Country	-	-	✓	✓	✓	✓	✓	✓
CON Neighb. Capitals	-	-	-	✓	✓	✓	✓	✓
Controls in 2010 × Year FE								
Education	-	-	-	-	✓	✓	✓	✓
Geography	-	-	-	-	-	✓	✓	✓
Population	-	-	-	-	-	-	✓	✓
Development	-	-	-	-	-	-	-	✓

Note: The table displays regression results from Equation (IV1) by region and across various specifications. Standard errors are clustered at the sub-national level. Significance levels are indicated as follows: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 2.A.7: Home Run Papers – Total Papers in Citation Distribution

	Across Fields (≥ ... pct.)				Within Fields (≥ ... pct.)			
	95th	99th	99.5th	99.9th	95th	99th	99.5th	99.9th
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Reduced Form								
ln CON Almaty × Post 2010	-0.009	-0.004	-0.004	0.001	-0.008	-0.013	-0.010	0.005
	(0.017)	(0.014)	(0.014)	(0.013)	(0.017)	(0.015)	(0.014)	(0.013)
Observations	40,528	40,528	40,528	40,528	40,528	40,528	40,528	40,528
Panel B: 2SLS								
ln Downloads	-0.027	-0.012	-0.011	0.004	-0.025	-0.039	-0.032	0.014
	(0.052)	(0.043)	(0.041)	(0.040)	(0.052)	(0.045)	(0.042)	(0.039)
Observations	40,528	40,528	40,528	40,528	40,528	40,528	40,528	40,528
F-statistic	38.321	38.321	38.321	38.321	38.321	38.321	38.321	38.321
Panel C: OLS								
ln Downloads	0.006	0.009**	0.009*	0.012**	0.006	0.007*	0.007	0.007
	(0.004)	(0.004)	(0.004)	(0.005)	(0.004)	(0.004)	(0.004)	(0.004)
Observations	40,528	40,528	40,528	40,528	40,528	40,528	40,528	40,528
Fixed Effects								
Sub-national	✓	✓	✓	✓	✓	✓	✓	✓
Country × Year	✓	✓	✓	✓	✓	✓	✓	✓
CON Neighb. Capitals	✓	✓	✓	✓	✓	✓	✓	✓
Controls in 2010 × Year FE	✓	✓	✓	✓	✓	✓	✓	✓

Note: The table displays regression results from Equation (IV2) for various measures of producing "home-run" papers – articles that reach the 95th, 99th, 99.5th, or 99.9th percentile citation distribution. In Columns (1)–(4) "home-run" papers are defined across fields, whereas in Columns (5)–(8), they are defined within fields. Across all panels, the sample is limited to years for which download data are available. Standard errors are clustered at the sub-national level. Significance levels are indicated as follows: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 2.A.8: Change in Citation Patterns by Region

	Number of Citations			
	Total	Non-peer-reviewed	Peer-reviewed	Cross-field
	(1)	(2)	(3)	(4)
Panel A: Reduced Form				
ln CON Almaty × Post 2010 × Developed	0.035	0.004	0.035	0.119
	(0.056)	(0.054)	(0.056)	(0.085)
ln CON Almaty × Post 2010 × Emerging	0.069	0.028	0.075	-0.101
	(0.047)	(0.045)	(0.046)	(0.073)
ln CON Almaty × Post 2010 × Developing	-0.010	-0.017	-0.006	0.003
	(0.027)	(0.020)	(0.027)	(0.033)
Observations	40,440	40,440	40,440	40,440
Panel B: 2SLS				
ln Downloads × Developed	0.058	0.003	0.065	0.067
	(0.075)	(0.069)	(0.075)	(0.111)
ln Downloads × Emerging	0.088	0.020	0.102	-0.133
	(0.085)	(0.076)	(0.084)	(0.122)
ln Downloads × Developing	-0.206	-0.294	-0.133	0.071
	(0.483)	(0.371)	(0.479)	(0.563)
Observations	40,440	40,440	40,440	40,440
F-statistic	19.712	19.712	19.712	19.712
Panel C: OLS				
ln Downloads × Developed	-0.007	-0.013	-0.008	-0.011
	(0.009)	(0.009)	(0.009)	(0.015)
ln Downloads × Emerging	-0.007	0.009	-0.005	-0.009
	(0.008)	(0.007)	(0.007)	(0.012)
ln Downloads × Developing	-0.012	-0.016	-0.012	0.004
	(0.028)	(0.023)	(0.027)	(0.030)
Observations	40,440	40,440	40,440	40,440
Fixed Effects				
Sub-national	✓	✓	✓	✓
Country × Year	✓	✓	✓	✓
CON Neighb. Capitals	✓	✓	✓	✓
Controls in 2010 × Year FE	✓	✓	✓	✓

Note: The table displays regression results from Equation (IV2) for various citation measures by region. Across all panels, the sample is limited to years for which download data are available. Standard errors are clustered at the sub-national level. Significance levels are indicated as follows: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 2.A.9: Migration Patterns by Region

	Stock of Researchers	Outflows		
		Subnational	Country	Developed
	(1)	(2)	(3)	(4)
Panel A: Reduced Form				
ln CON Almaty × Post 2010 × Developed	0.057	0.036	-0.008	-0.029
	(0.042)	(0.032)	(0.032)	(0.023)
ln CON Almaty × Post 2010 × Emerging	0.021	0.023	0.002	0.018
	(0.039)	(0.022)	(0.019)	(0.017)
ln CON Almaty × Post 2010 × Developing	-0.018	0.006	0.008	0.008
	(0.020)	(0.010)	(0.010)	(0.010)
Observations	40,440	40,440	40,440	40,440
Panel B: 2SLS				
ln Downloads × Developed	0.049	0.051	0.000	-0.013
	(0.058)	(0.037)	(0.034)	(0.025)
ln Downloads × Emerging	0.011	0.043	0.013	0.034
	(0.068)	(0.036)	(0.030)	(0.028)
ln Downloads × Developing	-0.330	0.094	0.146	0.128
	(0.371)	(0.171)	(0.166)	(0.161)
Observations	40,440	40,440	40,440	40,440
F-statistic	19.712	19.712	19.712	19.712
Panel C: OLS				
ln Downloads × Developed	-0.000	-0.014**	-0.012**	-0.015***
	(0.006)	(0.006)	(0.005)	(0.004)
ln Downloads × Emerging	-0.006	0.008*	0.009**	0.014***
	(0.006)	(0.004)	(0.004)	(0.004)
ln Downloads × Developing	0.026*	0.028***	0.025**	0.027***
	(0.014)	(0.011)	(0.010)	(0.009)
Observations	40,440	40,440	40,440	40,440
Fixed Effects				
Sub-national	✓	✓	✓	✓
Country × Year	✓	✓	✓	✓
CON Neighb. Capitals	✓	✓	✓	✓
Controls in 2010 × Year FE	✓	✓	✓	✓

Note: The table displays regression results from Equation (IV2) for various migration measures by region. Across all panels, the sample is limited to years for which download data are available. Standard errors are clustered at the sub-national level. Significance levels are indicated as follows: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 2.A.10: Patent Measures – Summary Statistics

	Mean	SD	Min	Max	N
	(1)	(2)	(3)	(4)	(5)
Panel A: Patent References (per Patent)					
Total	1.09	2.17	0.00	97.00	10,344
Referenced by Applicant	0.10	0.32	0.00	11.00	10,344
Referenced by Examiner	0.99	2.07	0.00	97.00	10,344
Referenced on Front	1.02	2.12	0.00	97.00	10,344
Referenced on Body	0.13	0.42	0.00	18.50	10,344
Quality Q1	0.01	0.04	0.00	1.00	10,344
Quality Q2	0.03	0.07	0.00	2.00	10,344
Quality Q3	0.06	0.13	0.00	3.00	10,344
Quality Q4	0.19	0.38	0.00	11.00	10,344
Quality Q5	0.62	1.46	0.00	64.00	10,344
Peer-reviewed	0.92	1.91	0.00	79.00	10,344
Restricted-access	0.86	1.79	0.00	71.00	10,344
Open-access	0.05	0.16	0.00	8.00	10,344
Panel B: Patent Citations (per Publication)					
Total	0.69	24.55	0.00	2,501.00	27,170
Cited by Applicant	0.54	19.98	0.00	1,935.00	27,170
Cited by Examiner	0.15	4.71	0.00	566.00	27,170
Cited on Front	0.57	20.39	0.00	2,072.00	27,170
Cited on Body	0.21	8.15	0.00	880.00	27,170
Citing Q1	0.01	0.18	0.00	17.91	27,170
Citing Q2	0.02	0.84	0.00	120.00	27,170
Citing Q3	0.03	0.53	0.00	62.00	27,170
Citing Q4	0.08	1.91	0.00	174.00	27,170
Citing Q5	0.49	20.30	0.00	2,229.00	27,170

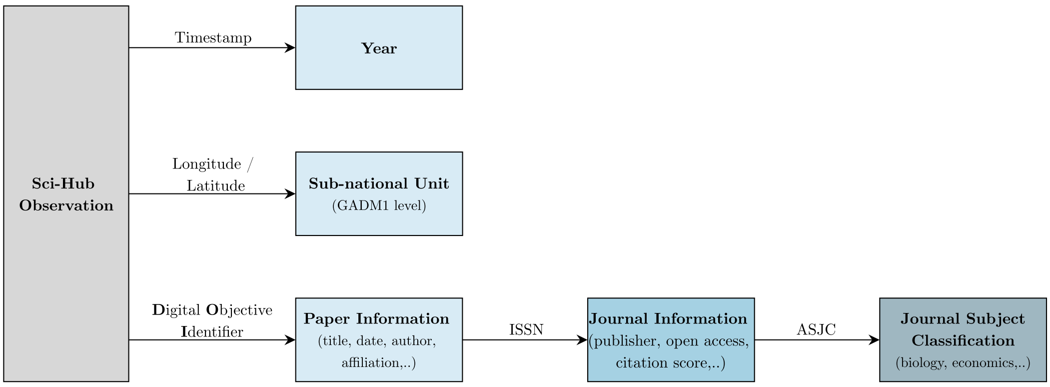
Note: The table provides summary statistics for patent citations and references as described in Section 2.D.1. Across all variables, the unit of observation is sub-national units from 2000 to 2020. In the case of references, observations are limited to OECD countries.

2.B Additional Figures

Figure 2.B.1: Sci-Hub Screenshot

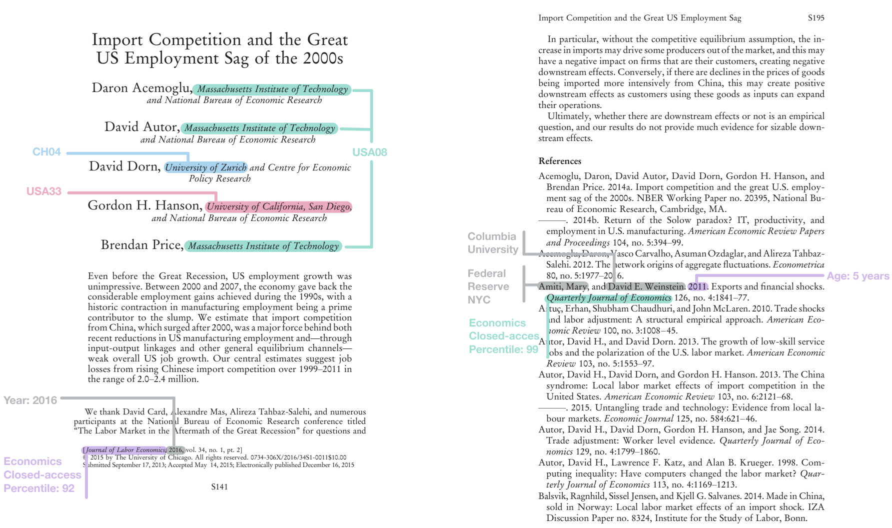
Note: The figure shows a screenshot of Sci-Hub's front page as of November 3, 2022.

Figure 2.B.2: Sci-Hub Data Structure



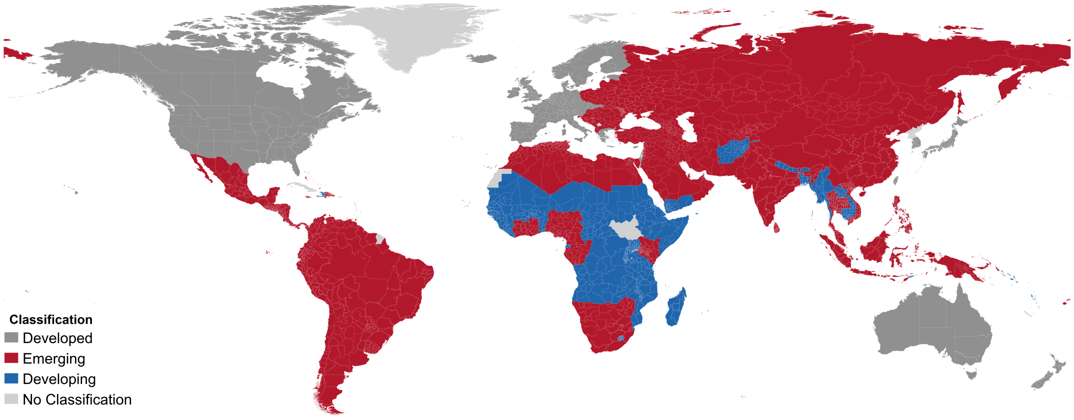
Note: The figure shows the structure of an entry in the Sci-Hub log-file downloads and describes how it is subsequently processed.

Figure 2.B.3: Research Output Classification Example



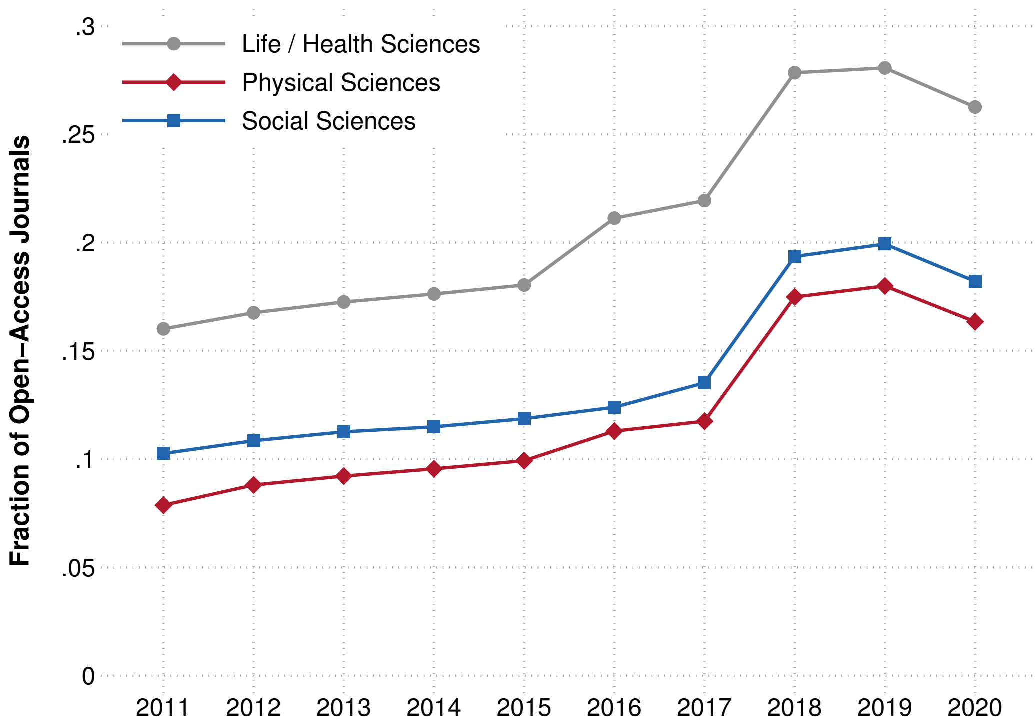
Note: The figure describes the type of characteristics extracted from a publication recorded in OpenAlex.

Figure 2.B.4: Country Classification



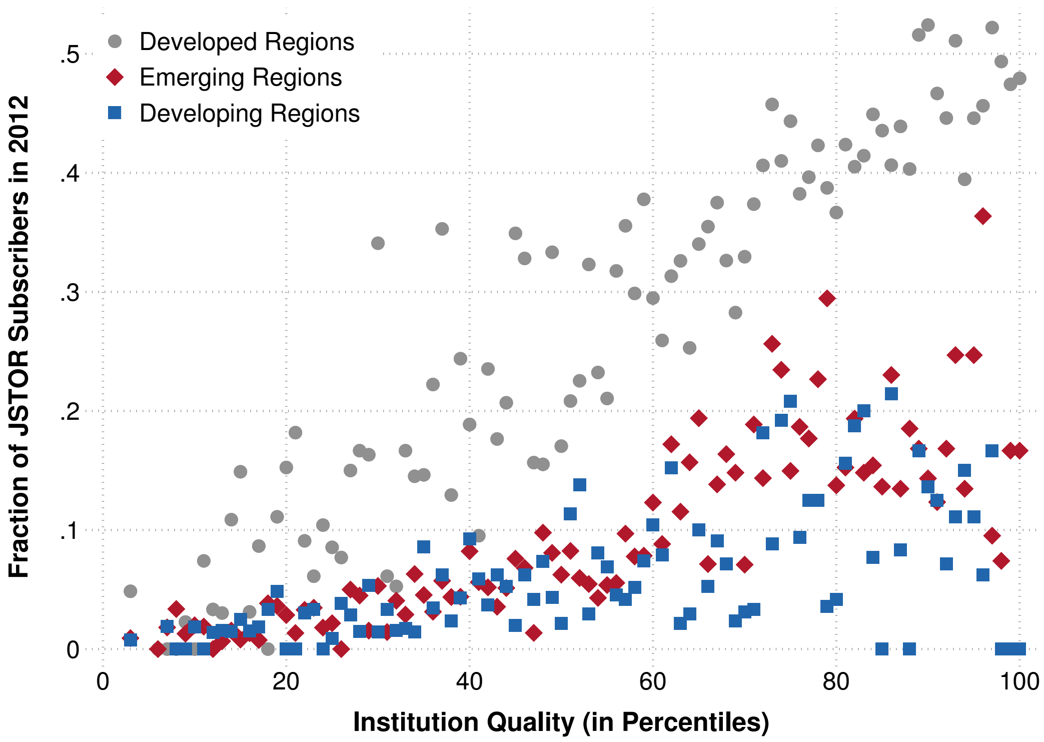
Note: The figure shows the classification of countries into developed, emerging, and developing regions. In particular, developed regions are all countries classified as 'least developed' by the United Nations (2011). All remaining countries are classified as developed or emerging regions based on the distinction of 'advanced' and 'emerging' economies by the International Monetary Fund (2011). Light white lines indicate borders of sub-national units.

Figure 2.B.5: Fraction of Open-Access Journal by Fields across Years



Note: The figure shows the fraction of open-access journals by fields across years.

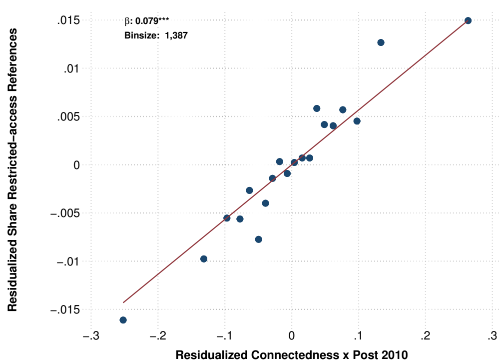
Figure 2.B.6: JSTOR Subscribers by Institution Quality and Region



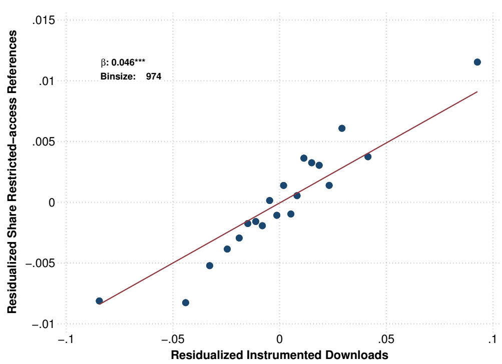
Note: The figure depicts the fraction of JSTOR subscribers by institution quality and region.

Figure 2.B.7: Share of Restricted-access References – Visual Evidence

(a) Reduced Form



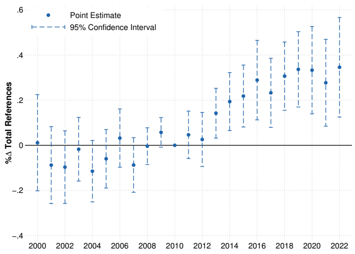
(b) 2SLS



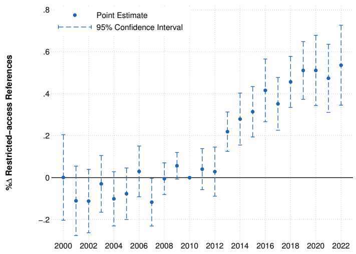
Note: Panel (a) shows point estimates and confidence intervals of the dynamic effects corresponding to the specification in Table 2.1 Panel A column (8). Panel (b) plots the residuals and coefficient estimate of the corresponding static difference-in-differences model. Standard errors are clustered by sub-national region.

Figure 2.B.8: Reduced Form Event Studies by Region (Count Variables)

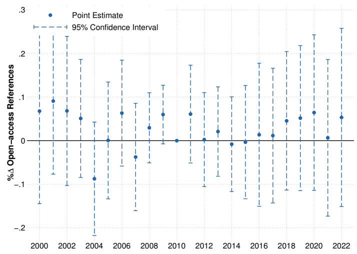
(a) Total References (Developed)



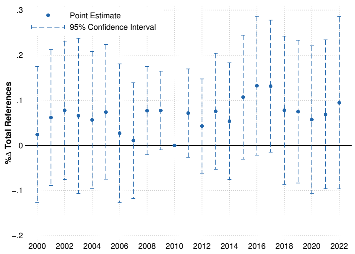
(b) RA References (Developed)



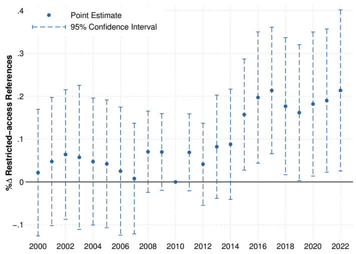
(c) OA References (Developed)



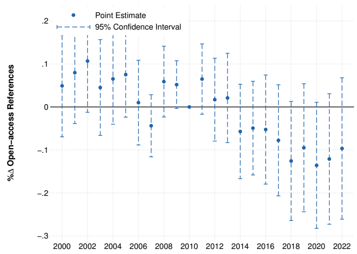
(d) Total References (Emerging)



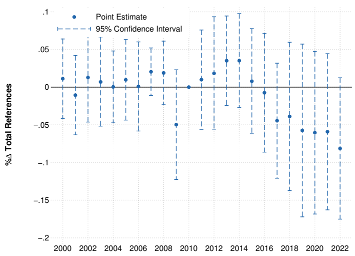
(e) RA References (Emerging)



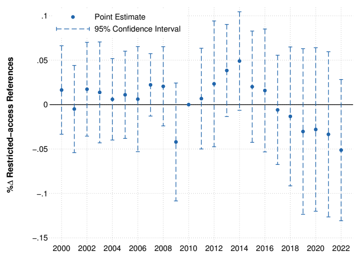
(f) OA References (Emerging)



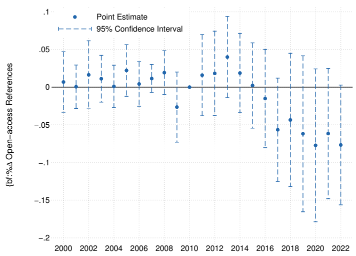
(g) Total References (Developing)



(h) RA References (Developing)



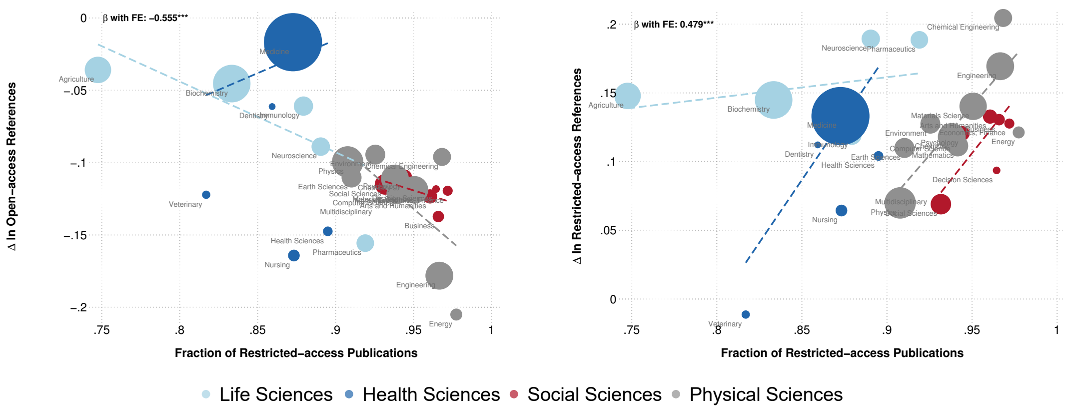
(i) OA References (Developing)



Note: The figure shows reduced form event study estimates with corresponding 95% confidence intervals for the outcomes and specification displayed in Table 2.3 Panel A. The post-2010 indicator is replaced with a full set of annual indicators, omitting 2010, the year before Sci-Hub was established. Standard errors are clustered by sub-national region.

Figure 2.B.9: Change in Reference Dynamics by Field and Sub-field

(a) Open-access (b) Restricted-access



Note: The figure shows disaggregated 2SLS estimates for the number of open-access and restricted-access references according to the specification in Panel B of Table 2.3. Each scatter represents a separate regression in which the dependent variable is the number of open-access or restricted-access references in a field. Effect sizes are indicated on the vertical axis. The share of open-access journals is displayed on the horizontal axis. The size of each scatter indicates the size of a field, measured by the total number of publications in 2010.

2.C Additional Analyses

2.C.1 Weak Instrument Considerations

It is well known that t-ratio tests over-reject when instruments are weak (Bound, Jaeger and Baker, 1995; Staiger and Stock, 1997). The discussion on dealing with potentially weak instruments revolves around two parameters: the first-stage F-statistic and the endogeneity coefficient ρ, measuring the correlation between structural and first-stage residuals. Within this framework, a high degree of endogeneity calls for a strong instrument, i.e., a a high first-stage F-statistic. In contrast, 'low' endogeneity is reconcilable with a low first-stage F-statistic. In particular, conventional (unadjusted) IV standard errors sufficiently account for weak instruments unless endogeneity is 'extraordinarily high', defined as $|\rho| > .565$ (Angrist and Kolesár, 2021). However, because it might be challenging to bound ρ a priori, numerous frequentist methods exist to adjust standard errors and confidence intervals for potential inference distortions (Anderson and Rubin, 1949; Lee et al., 2022).

We address potential weak instrument concerns twofold. First, we report 95-percent confidence intervals $[\hat{\rho}_L, \hat{\rho}_U]$ of the endogeneity parameter ρ. Table 2.C.1 shows that our specification exhibits moderate to high levels of endogeneity, exceeding the threshold of $|\rho| > .565$ when considering our main specification. The high degree of endogeneity might not be surprising given that knowledge creation is a highly endogenous process. At the same time, the high degree of endogeneity justifies our instrumental variable approach and offers an explanation for the stark difference between OLS and 2SLS estimates we see in Tables (2.3)–(2.4).

Complementing the bounding exercise on ρ, Table 2.C.1 reports p-values of the Anderson and Rubin F-test (Anderson and Rubin, 1949) as well as tF-adjusted standard errors (Lee et al., 2022). The procedure by Anderson and Rubin yields confidence intervals with undistorted coverage for any pair of values ρ and F. On the other hand, tF-adjusted standard errors assume a worst-case endogeneity scenario, i.e., $|\rho| = 1$, and accordingly adjust the conventional 2SLS standard errors by an adjustment factor based on the first-stage F-statistic and the considered significance level.[32] Under both procedures, our results remain

[32] Both procedures yield correct coverage under arbitrarily weak instruments; however, the expected length of the Anderson and Rubin confidence interval is infinite, while the corresponding tF interval is finite (Lee et al., 2022).

significant at the 1-percent level even when considering a worst-case endogeneity scenario of $|\rho| = 1$ as assumed when computing tF-adjusted standard errors.

Table 2.C.1: Weak IV – Share of Restricted-access References

	Dependent Variable: Share Restricted-access References							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: 2SLS Estimate								
ln Downloads	-0.023***	-0.024***	0.026***	0.049***	0.049***	0.047***	0.046***	0.046***
	(0.002)	(0.002)	(0.003)	(0.015)	(0.010)	(0.010)	(0.009)	(0.009)
Observations	20,463	20,413	19,420	19,420	19,420	19,420	19,420	19,420
F-statistic	983.565	1760.168	139.658	11.890	22.981	24.268	31.087	30.898
Panel B: Weak IV Considerations								
Endogeneity Parameter ρ								
$\max\{\lvert\hat{\rho}_L\rvert, \lvert\hat{\rho}_U\rvert\}$	0.420	0.500	0.520	0.900	0.760	0.750	0.720	0.720
Anderson-Rubin Inference								
p-value	< .001	< .001	< .001	< .001	< .001	< .001	< .001	< .001
tF-adjusted Standard Errors								
5-percent Significance	(0.002)	(0.002)	(0.031)	(0.031)	(0.013)	(0.012)	(0.010)	(0.010)
1-percent Significance	(0.003)	(0.002)	(0.088)	(0.088)	(0.017)	(0.016)	(0.013)	(0.013)
Fixed Effects								
Sub-national	-	✓	✓	✓	✓	✓	✓	✓
Year × Country	-	-	✓	✓	✓	✓	✓	✓
CON Neighb. Capitals	-	-	-	✓	✓	✓	✓	✓
Controls in 2010 × Year FE								
Education	-	-	-	-	✓	✓	✓	✓
Geography	-	-	-	-	-	✓	✓	✓
Population	-	-	-	-	-	-	✓	✓
Development	-	-	-	-	-	-	-	✓

Note: Panel A displays 2SLS estimates based on Equation (IV2). Panel B reports three measures to discover and account for the presence of weak instruments. First, we report a bound on the endogeneity parameter ρ by following Online Appendix Section A.8.3 of Lee et al. (2022). In particular, we use 95-percent tF confidence interval endpoints $[\hat{\beta}_L, \hat{\beta}_U]$ to compute the endpoints $\rho(\hat{\beta}_L)$ and $\rho(\hat{\beta}_U)$. Second, we report p-values of the Anderson-Rubin F-test of endogenous regressors (Anderson and Rubin, 1949). Third, we construct tF-adjusted standard errors for 5-percent and 1-percent significance levels using first-stage F-statistics and critical values provided in Lee et al. (2022). Standard errors are clustered at the sub-national level. Significance levels are indicated as follows: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

2.C.2 Two-sample IV

One challenge we face is that Sci-Hub downloads are only available for six years – that is 2011 to 2013 and 2015 to 2017 – while missing for all other years post-2010. Although our instrument and dependent variables are observable across the entire sample period, the missing download observations affect the 2SLS estimates. To see this, consider a setting where the endogenous regressor $\boldsymbol{X}$ is only observed for a subset of observations, denoted by $\boldsymbol{X}_{\text{sub}}$. Further assume that the dependent variable, $\boldsymbol{Y}$ and the instrument $\boldsymbol{Z}$ are observed across all observations, denoted by $\boldsymbol{Y}_{\text{all}}$ and $\boldsymbol{Z}_{\text{all}}$. In this setting, even though the outcome and the instrument are observed across all observations, the standard 2SLS estimate utilizes only observations of the sample for which $\boldsymbol{X}$ is non-missing, that is, $\hat{\beta}^{\text{2SLS}}_{\text{sub}} = \left(\hat{\boldsymbol{X}}'_{\text{sub}}\hat{\boldsymbol{X}}_{\text{sub}}\right)^{-1}\hat{\boldsymbol{X}}'_{\text{sub}}\boldsymbol{Y}_{\text{sub}}$ where $\hat{\boldsymbol{X}}_{\text{sub}} = \boldsymbol{Z}_{\text{sub}}(\boldsymbol{Z}'_{\text{sub}}\boldsymbol{Z}_{\text{sub}})^{-1}\boldsymbol{Z}_{\text{sub}}\boldsymbol{X}_{\text{sub}}$.

However, as laid out in Angrist and Krueger (1992), instrumental variables estimation on the entire sample is still possible even when only information on $\boldsymbol{Y}$ and $\boldsymbol{Z}$ but not $\boldsymbol{X}$ is available. The idea is to estimate the first-stage regression on the restricted sample, but perform the subsequent prediction step on the entire sample, that is, $\hat{\boldsymbol{X}}_{\text{all}} = \boldsymbol{Z}_{\text{all}}(\boldsymbol{Z}'_{\text{sub}}\boldsymbol{Z}_{\text{sub}})^{-1}\boldsymbol{Z}_{\text{sub}}\boldsymbol{X}_{\text{sub}}$. The 2SLS estimate then follows from $\hat{\beta}^{\text{2SLS}}_{\text{all}} = \left(\hat{\boldsymbol{X}}'_{\text{all}}\hat{\boldsymbol{X}}_{\text{all}}\right)^{-1}\hat{\boldsymbol{X}}'_{\text{all}}\boldsymbol{Y}_{\text{all}}$[33].

To transfer this idea to our setting, we slightly adjust our empirical model in Equation (IV1) by replacing year fixed-effects with decade fixed effects. In particular, in a setting with any kind of year fixed effects it is not possible to predict the first stage on the entire sample because in years with missing Sci-Hub downloads, the corresponding observations are missing for all observations. Our sample period can roughly be divided into two decades defined by the years before and after $t = 2010$. We estimate the following adjusted first-stage regression:

$$\begin{aligned} \ln \text{Down}_{it} = {} & \alpha_i + \alpha_{c(i)d(t)} \\ & + \beta_1 \ln \text{CON}_i^{\text{Almaty}} \times \mathbb{1}_{t>2010} + \sum_n \delta_2^{(n)} \ln \text{CON}_i^{\text{n}} \times \mathbb{1}_{t>2010} \\ & + \boldsymbol{X}_{i2010}\gamma_{d(t)} + \varepsilon_{it} \end{aligned} \tag{2SIV}$$

where $\alpha_{c(i)d(t)}$ accounts for country-specific factors that change by decade. All other variables, except for $\alpha_{c(i)d(t)}$, are defined as in Equation (IV1). We conduct inference on $\hat{\beta}^{\text{2SLS}}_{\text{all}}$ using a clustered bootstrap with 1,000 replications.

In Tables 2.C.2–2.C.4 we compare estimates using the standard 2SLS with the two-sample 2SLS approach across our main outcome tables. Qualitatively, the results are robust across both approaches with slightly higher point estimates when utilizing the missing observations in the two-sample 2SLS approach.

[33] Inoue and Solon (2010) proposes a slightly modified estimator by introducing a correcting matrix $\boldsymbol{C}$ to adjust for finite sample differences of the covariance matrix of $\boldsymbol{Z}$ between the two samples. However, in our setting $\boldsymbol{Z}$ is identically distributed across both samples since connectedness is constant across sub-national within pre- and post-treatment periods.

Table 2.C.2: Two-sample IV Estimates – References

	Number of References			Share Restricted-access References
	Total	Open-access	Restricted-access	
	(1)	(2)	(3)	(4)
Panel A: 2SLS				
ln Downloads	0.110	-0.083	0.194	0.047***
	(0.120)	(0.093)	(0.120)	(0.012)
Observations	41,341	41,341	41,341	20,408
Number of Clusters	2,437	2,437	2,437	1,461
F-statistic	41.339	41.339	41.339	31.652
Panel B: Two-Sampe 2SLS				
ln Downloads	0.080	-0.177	0.233**	0.080***
	(0.128)	(0.130)	(0.117)	(0.016)
Observations	56,051	56,051	56,051	28,859
Number of Clusters	2,437	2,437	2,437	1,461
F-statistic	41.339	41.339	41.339	31.652
Fixed Effects				
Sub-national	✓	✓	✓	✓
Country × Decade	✓	✓	✓	✓
CON Neighb. Capitals	✓	✓	✓	✓
Controls in 2010 × Decade FE	✓	✓	✓	✓

Note: The table displays regression results from Equation (2SIV) across various specifications using the inverse hyperbolic sine transformation. Standard errors are bootstrapped with 1,000 replications at the sub-national level. Significance levels are indicated as follows: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 2.C.3: Two-sample IV Estimates – Citations

	Number of Citations			
	Total	Non-peer-reviewed	Peer-reviewed	Cross-field
	(1)	(2)	(3)	(4)
Panel A: 2SLS				
ln Downloads	0.108	0.018	0.122	-0.096
	(0.082)	(0.071)	(0.082)	(0.117)
Observations	41,341	41,341	41,341	41,341
Number of Clusters	2,437	2,437	2,437	2,437
F-statistic	41.339	41.339	41.339	41.339
Panel B: Two-Sampe 2SLS				
ln Downloads	0.134	0.046	0.139*	-0.082
	(0.083)	(0.076)	(0.078)	(0.143)
Observations	56,051	56,051	56,051	56,051
Number of Clusters	2,437	2,437	2,437	2,437
F-statistic	41.339	41.339	41.339	41.339
Fixed Effects				
Sub-national	✓	✓	✓	✓
Country × Decade	✓	✓	✓	✓
CON Neighb. Capitals	✓	✓	✓	✓
Controls in 2010 × Decade FE	✓	✓	✓	✓

Note: The table displays regression results from Equation (2SIV) across various specifications using the inverse hyperbolic sine transformation. Standard errors are bootstrapped with 1,000 replications at the sub-national level. Significance levels are indicated as follows: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 2.C.4: Two-sample IV Estimates – Citations by Region

	Number of Citations			
	Total	Non-peer-reviewed	Peer-reviewed	Cross-field
	(1)	(2)	(3)	(4)
Panel A: 2SLS				
ln Downloads × Developed	0.056	0.000	0.064	0.063
	(0.075)	(0.068)	(0.074)	(0.109)
ln Downloads × Emerging	0.087	0.017	0.102	-0.137
	(0.086)	(0.076)	(0.085)	(0.122)
ln Downloads × Developing	-0.229	-0.314	-0.152	0.054
	(0.492)	(0.379)	(0.487)	(0.568)
Observations	41,341	41,341	41,341	41,341
Number of Clusters	2,437	2,437	2,437	2,437
F-statistic	20.418	20.418	20.418	20.418
Panel B: Two-Sampe 2SLS				
ln Downloads × Developed	0.088	0.019	0.091	0.115
	(0.059)	(0.079)	(0.088)	(0.129)
ln Downloads × Emerging	0.120	0.058	0.126**	-0.115
	(0.063)	(0.065)	(0.062)	(0.132)
ln Downloads × Developing	-0.421	-0.252	-0.380	-0.235
	(0.724)	(0.659)	(0.765)	(0.660)
Observations	56,051	56,051	56,051	56,051
Number of Clusters	2,437	2,437	2,437	2,437
F-statistic	20.418	20.418	20.418	20.418
Fixed Effects				
Sub-national	✓	✓	✓	✓
Country × Decade	✓	✓	✓	✓
CON Neighb. Capitals	✓	✓	✓	✓
Controls in 2010 × Decade FE	✓	✓	✓	✓

Note: The table displays regression results from Equation (2SIV) across various specifications using the inverse hyperbolic sine transformation. Standard errors are bootstrapped with 1,000 replications at the sub-national level. Significance levels are indicated as follows: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

2.D Additional Data

2.D.1 Patents

Data Collection We rely on two external data sources. First, we utilize the 'OECD REGPAT' database as of August 2022 (for Economic Co-operation and Development, 2022). The database covers patent applications from 1977 to 2022 from applicants and inventors in OECD countries filed to either (1) the European Patent Office (EPO) or (2) under the Patent Co-operation Treaty (PCT).[34] For all patents 'OECD REGPAT' database contains regional identifiers based on the address provided in the patent application. The regional identifiers correspond to the 2013 version of the Nomenclature of Territorial Units for Statistics (NUTS) for European countries and OECD's Territorial Level 3 (TL3) for other countries. We construct a spatial crosswalk for both identifiers to align the data with the sub-national unit structure described in Section 2.3.

Second, we utilize data on the citations from USPTO and EPO patents to scientific articles since 1836 and 1978, respectively. For a detailed description of the extraction of scientific articles from patents, we refer the interested reader to Marx and Fuegi (2022) and the explanations therein. We perform two steps to utilize the openly accessible dataset in our analysis. First, we match patents to the 'OECD REGPAT' database using their unique patent publication number. Next, we corroborate the referenced scientific articles – identified by their Microsoft Academic Graph identifier and corresponding with the OpenAlex identifier – with the paper and journal characteristics (again fixed in 2011) described in Section 2.3. We restrict references along two dimensions. First, we exclude references included by patent office examiners who are unlikely to face access restrictions and therefore less likely to be Sci-Hub users.[35] Further, mirroring the approach in Bryan and Ozcan (2021), we only consider in-text references and exclude front-page references.

Lastly, we aggregate the combined data into a panel of sub-national units across years. As in the case of publications, patents with multiple inventors residing in the same sub-national unit are only considered once. Patents with numerous inventors residing in different sub-national units are considered separately.

[34] In cases where we observe patents filed to the EPO that are also protected under PCT, we only consider the latter.

[35] There are two ways how references can be assigned to a patent. First, references can be listed by the applicant. Second, when checking the patent's validity, the patent office examiner can include scientific literature related to the patent.

Patent References Mirroring our approach to publications, we evaluate whether inventors in regions with high Sci-Hub usage (induced through high social connectedness to Almaty) adjust their reference dynamics by citing more restricted-access research post-2010. Column (1) of Table 2.8 documents changes in the share of restricted-access references within all scientific references listed in patents. Both the reduced form and the 2SLS estimate indicate that the share of restricted-access references is unchanged. Next, we check in Columns (2)–(6) whether the share of restricted-access references adjusts when disaggregating references across quality quintiles. Again, we find no meaningful change in reference dynamics except for a positive effect for references from inventors in the 1st quality quintile, significant at the 10% level. However, the latter estimate turns insignificant when adjusting for multiple hypothesis testing. These findings might be explained by access restrictions being less binding for inventors (especially those working for large corporations) compared to researchers working in an academic environment.

2.D.2 Distance to Research Frontier

To evaluate the potential impacts of access restrictions on research topics and directions, we develop a text-based measure of similarity to the research frontier. For each academic field and year, we define the research frontier as all papers in the top percentile of the citation distribution; representing the most innovative and influential contributions within their respective fields. To identify the thematic focus of these frontiers (Appendix Figure 2.D.1a), we use a topic modeling approach. Topic modeling is an unsupervised machine learning technique, which aims to identify latent themes in textual data, enabling the representation of each text as a distribution over topics. Thus, unlike binary classifications, topic modeling provides a nuanced representation of research content. Here, we use the BERTopic algorithm introduced by Grootendorst (2022).

The topic modeling process involves two steps. First, we generate embeddings – vector representations of textual data – for each abstract within the research frontier. This is accomplished using the pre-trained multilingual language model *paraphrase-multilingual-MiniLM-L12-v2* (Reimers and Gurevych, 2019), which produces 384-dimensional embeddings that effectively capture contextual relationships between words and supports over 50 languages. Second, embeddings are clustered based on their proximity within vector space, with each cluster representing a specific topic. The number of clusters, or topics, is determined by optimizing the model's hyperparameters through cross-validation, aiming to maximize the coherence score – a metric that evaluates the quality and interpretability of the topics (Mimno et al., 2011). This process ultimately generates a set of topics for each field and year, along with a topic distribution for each paper in the research frontier, denoted by $\vec{q}j$. Collectively, the set of topic distributions for all research frontier papers is represented as $\mathbf{Q}$. The corresponding covariance matrix, $\mathbb{V}\mathbf{Q}$, captures the interrelationships and substitutability among topics within the research frontier.

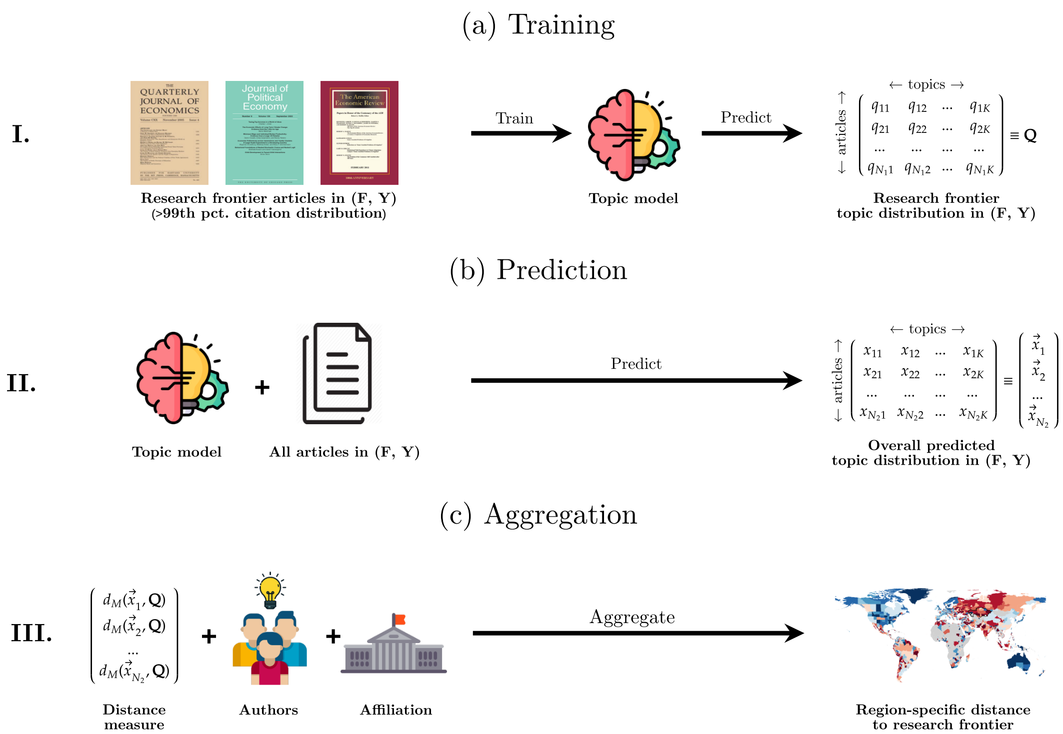
To assess the similarity of other papers to the research frontier $\mathbf{Q}$, we apply the trained topic model to predict the topic distributions of previously unseen papers within the same field and year (Appendix Figure 2.D.1b). We then calculate the similarity between a paper's topic distribution and that of the research frontier using the Mahalanobis distance:

$$d_M \equiv d_M(\vec{x}_i, \mathbf{Q}) = \frac{1}{N_1} \sum_{\vec{q}_j \in \mathbf{Q}} \sqrt{(\vec{x}_i - \vec{q}_j)' \mathbb{V}_{\mathbf{Q}}^{-1} (\vec{x}_i - \vec{q}_j)}$$

Here, $\vec{x}i$ represents the topic distribution of an individual paper, $\mathbf{Q}$ represents the set of topic distributions of all papers in the research frontier, and $\mathbb{V}\mathbf{Q}$ captures the substitutability of topics within the research frontier. For each paper, d_M quantifies its distance from the research frontier, with a one-unit increase in d_M corresponding to a one-standard-deviation divergence from the research frontier. This distance equals zero when a paper's topic distribution perfectly matches that of the research frontier and grows quadratically as the divergence increases.

To analyze the alignment of research topics with the research frontier across geographical units, we aggregate the paper-specific distances, d_M, by field, subnational region, and year (Appendix Figure 2.D.1c). This aggregation provides insights into how closely research topics in specific regions align with the global research frontier. By using this measure as the outcome variable in our regression analysis, we evaluate whether researchers in highly connected regions, compared to those in less connected regions, move closer to the research frontier following the launch of Sci-Hub.

Figure 2.D.1: Construction of Topic Distance from Research Frontier



Note: The figure provides a schematic representation of how field-specific topic distributions are constructed, as outlined in Appendix Section 2.D.2. First, for each field and year, we train separate topic models using papers from the top percentile of the citation distribution, representing the research frontier (Appendix Figure 2.D.1a). These trained models are then applied to predict the topic distributions of all other abstracts published within the same academic discipline and year (Appendix Figure 2.D.1b). Next, the Mahalanobis distance between the topic distribution of each paper and the topic distribution at the research frontier is calculated. Finally, these paper-specific topic distribution distances are averaged across sub-national units and years (Appendix Figure 2.D.1c).

Bibliography

Anderson, Theodore W, and Herman Rubin. 1949. "Estimation of the Parameters of a Single Equation in a Complete System of Stochastic Equations." *Annals of Mathematical Statistics*, 20(1): 46–63.

Angrist, Joshua, and Michal Kolesár. 2021. "One Instrument to Rule Them All: The Bias and Coverage of Just-ID IV." *NBER Working Paper*, 29417.

Angrist, Joshua D, and Alan B Krueger. 1992. "The Effect of Age at School Entry on Educational Attainment: An Application of Instrumental Variables with Moments from Two Samples." *Journal of the American Statistical Association*, 87(418): 328–336.

Bound, John, David A Jaeger, and Regina M Baker. 1995. "Problems with Instrumental Variables Estimation when the Correlation between the Instruments and the Endogenous Explanatory Variable is Weak." *Journal of the American Statistical Association*, 90(430): 443–450.

Bryan, Kevin A, and Yasin Ozcan. 2021. "The Impact of Open Access Mandates on Invention." *Review of Economics and Statistics*, 103(5): 954–967.

for Economic Co-operation, Organisation, and OECD Development. 2022. "OECD REGPAT Database, August 2022." *URL: https://transfer.oecd.org/w/f-12223b5f-f275-4456-9ef2-0a305b8eab37*, Accessed: 2023-03-12.

Grootendorst, Maarten. 2022. "BERTopic: Neural Topic Modeling With a Class-based TF-IDF Procedure." *arXiv Pre-print*, 2203.05794.

Inoue, Atsushi, and Gary Solon. 2010. "Two-sample Instrumental Variables Estimators." *Review of Economics and Statistics*, 92(3): 557–561.

International Monetary Fund, IMF. 2011. "World Economic Outlook Database – WEO Update: June 17, 2011." *URL: https://www.imf.org/en/Publications/WEO/weo-database/2011/April*, Accessed: 2022-01-03.

Lee, David S, Justin McCrary, Marcelo J Moreira, and Jack Porter. 2022. "Valid t-ratio Inference for IV." *American Economic Review*, 112(10): 3260–90.

Marx, Matt, and Aaron Fuegi. 2022. "Reliance on Science by Inventors: Hybrid Extraction of In-text Patent-to-article Citations." *Journal of Economics & Management Strategy*, 31(2): 369–392.

Mimno, David, Hanna Wallach, Edmund Talley, Miriam Leenders, and Andrew McCallum. 2011. "Optimizing Semantic Coherence in Topic Models." 262–272.

Reimers, Nils, and Iryna Gurevych. 2019. "Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks." Association for Computational Linguistics.

Staiger, Douglas, and James H Stock. 1997. "Instrumental Variables Regression with Weak Instruments." *Econometrica*, 557–586.

United Nations, UN. 2011. "The Least Developed Countries Report, 2011." *URL:* *https://unctad.org/system/files/official-document/ldc2011_en.pdf*, Accessed: 2022-01-03.

Chapter 3

Retrieving Organs, Losing Motivation? The Response of Medical Staff to Corruption News

(joint with Alida Sangrigoli, Giuseppe Sorrenti, and Gilberto Turati)

3.1 Introduction

Workers' motivation and morale are essential for public sector performance, especially where incentives are primarily intrinsic rather than monetary. Motivation and morale can be enhanced through opportunities for promotion (Finan, Olken and Pande, 2017) or by highlighting the inherent value of public service missions (Besley and Ghatak, 2005). Conversely, they can be negatively affected by issues like non-meritocratic promotions, which can reduce productivity (Deserranno and León Ciliotta, 2021), or the disclosure of workplace corruption, which can demoralize workers.

This paper investigates how media coverage of corruption influences the behavior of public sector workers. In particular, we focus on the behavior of non-medical staff in Intensive Care Units (ICUs) involved in the organ procurement process. Corruption is a global issue affecting various sectors, including healthcare (Vian, 2008). Examples of corrupt practices in healthcare include physicians accepting

* We thank Maddalena Totarelli and seminar participants at the ZEW Hospital Summer Course 2021, the University of Pavia, and the Italian Health Economics Association (AIES) 2015 Conference for helpful suggestions on earlier drafts of this paper.

bribes to favor specific drugs or equipment and manipulating waiting lists (Vigdor, 2020; Scepanovic, 2006).[1]

Corruption can have several *direct* effects, ranging from inefficient spending to either bypassing bureaucratic obstacles (greasing the wheels) or increasing inefficiencies and costs (sanding the wheels).[2] However, once exposed and publicized, it can also generate *indirect* effects, such as lowering workers' motivation and morale or eroding trust among patients and citizens.[3]

We focus on two central research questions: Does media disclosure of corruption scandals affect medical staff behavior in organ procurement? Does the response vary based on the professional roles of those involved in the scandal?

We study the case of the Italian National Health Service (NHS), focusing on two high-profile corruption scandals involving a hospital manager and a surgeon. Within this framework, we analyze how media-driven perceptions of corruption influence the behavior of medical staff working in Intensive Care Units (ICUs). Following Bottan and Perez-Truglia (2015), our interest lies not in corruption itself but in the impact of media coverage of corruption scandals.

The organ procurement process is an ideal setting to study these indirect effects for several reasons. First, organ transplants – particularly kidney and liver transplants – are cost-effective compared to alternative treatments (Mendeloff et al., 2004; Jarl and Gerdtham, 2012). However, like many other countries, Italy faces a significant shortage of organ donors (Becker and Elías, 2007).[4] A decline in the number of reported donors could be a negative indirect consequence of corruption, with serious welfare implications for patients and their families.[5] Second, organ procurement relies heavily on trust. A shock to perceived corruption

[1] Bergman, Grennan and Swanson (2021) highlight the important role of physician-firm interactions in hospitals' procurement of medical devices.

[2] For instance, Mironov and Zhuravskaya (2016) evaluate the efficiency of corruption in public procurement, finding no evidence to support the "efficient grease" hypothesis.

[3] A growing literature studies the impact of media and the behavior of individuals affected by corrupt practices. Ferraz and Finan (2008) and Strömberg (2015) provide evidence of significant effects on the electoral outcomes of incumbent politicians following disclosure of their corrupt practices. Daniele, Aassve and Le Moglie (2023) suggest a long-term impact of corruption on trust in institutions, which also affects voting behavior. Bottan and Perez-Truglia (2015) show a significant decline in religious participation and charitable contributions following a clergy abuse scandal in the United States. Enikolopov, Petrova and Sonin (2018) present evidence that blog posts about corruption in state-controlled firms cause a significant decline in stock performance of targeted companies in Russia.

[4] Italy ranks among the top fifteen countries globally in terms of actual donors.

[5] For instance, Jensen, Sørensen and Petersen (2014) find that kidney transplantation, by eliminating the need for dialysis, yields an additional 2.8 quality-adjusted life years and saves approximately EUR 30,000 per patient.

could influence medical staff behavior by eroding trust and motivation. Third, Italy has long been perceived as a country with high levels of corruption. In 2015, it ranked 61st worldwide and second to last within the European Union (EU) (Transparency International, 2015).

Our identification strategy builds upon the specific structure of Italy's organ procurement process. The initial stage of this process depends solely on the effort of medical staff, enabling us to isolate their reaction to the disclosure of corruption scandals. Nevertheless, the effort required in this process stage is considerable and individual motivation is key to maximizing the probability of converting *potential* donors into *reported* donors. Specifically, ICU medical staff must first identify *potential* donors – patients at risk of evolving toward brain death – and maintain them in a state between brain and clinical death to preserve the possibility of organ transplantation. A legally deceased patient with no medical contraindications and a confirmed brain death assessment by the medical staff is classified as a *reported* donor. At this stage, no external factors – such as family decisions or legal barriers – interfere with the medical staff's role. By contrast, the final stage of the process, which determines the number of *actual* donors, i.e., deceased individuals whose organs are successfully retrieved, is influenced by family consent and other external constraints. The absence of such interferences in the earlier stage allows us to isolate the response of ICU medical staff, measured through changes in the number of reported donors following shocks to perceived corruption.

We focus on one of the most active regional Organ Procurement Centers in Italy, located at Hospital 'Molinette' in Turin, which serves the Northwestern regions of Piedmont and the Aosta Valley. Our choice is motivated by three key factors. First, organ donation rates in these regions are among the highest in Italy. In 2014, Piedmont and the Aosta Valley reported 49.6 potential donors per million people (pmp) and 28.8 actual donors pmp, well above the national averages of 39.4 potential and 23.2 actual donors pmp (Centro Nazionale Trapianti, 2014). Second, in 2001–2002, 'Molinette' was at the center of two high-profile corruption scandals that received extensive media coverage. One scandal involved the hospital's management, while the other directly implicated surgeons in the transplant center. The distinct roles of those involved – management versus surgeons – allow us to analyze potential differences in how medical staff respond to different types of corruption. Third, Italy's constitutional framework assigns healthcare policy management to the regional level. As a result, the Organ Procurement Center at 'Molinette' operates independently within its regional jurisdiction.

Our analysis relies on three newly compiled datasets. The first dataset includes monthly information on potential, reported, and actual donors between 2001 and 2005 for hospitals in Piedmont and the Aosta Valley, where the scandals occurred. Using hospitals as the unit of observation, we collect donor data at the hospital-month level. To construct a counterfactual scenario, we assemble a second dataset containing organ donation data from hospitals in two bordering regions not affected by the scandals, namely Lombardy and Liguria. The third dataset tracks media coverage of the corruption scandals. Since social media was largely unavailable at the time, we measure public exposure to corruption news by the number of newspaper articles covering the local healthcare sector. To ensure robustness, we compile data from multiple newspaper outlets and television broadcasts.

We evaluate the impact of corruption scandals on medical staff behavior using a difference-in-differences (DiD) approach. Specifically, we compare regions exposed to corruption scandals (Piedmont and the Aosta Valley) with unexposed bordering regions (Lombardy and Liguria). Importantly, hospitals in control regions fall under a different organ procurement jurisdiction, ensuring that their medical staff should remain unaffected by the scandals. To measure media exposure intensity, we use the number of newspaper articles covering the scandals, allowing us to assess whether effects are greater in regions with more media coverage. The validity of our DiD design rests on the parallel trends assumption that in the absence of the scandals, organ procurement trends in affected and unaffected regions would have evolved in parallel. We verify this assumption through an event study analysis, confirming no pre-existing differences between affected and unaffected hospitals.

Our baseline analysis finds that joint media coverage of the two corruption scandals negatively affects the number of reported donors, though the overall effect is small and statistically insignificant. However, the nature of the scandals plays a crucial role in shaping medical staff behavior. The hospital management case has no significant impact on reported donors, as ICU staff likely view it as less relevant to their work. In contrast, the surgeon's scandal has a significant and substantial negative effect.

Although the impact is short-lived – lasting approximately five to ten months – it is both sizable and robust across various empirical specifications. An increase of ten monthly news articles on the surgeon scandal leads to a reduction of four reported donors per month across affected hospitals. Over ten months, this amounts to approximately 35% fewer reported donors than would have been expected in the absence of the scandal.

Furthermore, while media coverage had national reach, our analysis shows that its effects were confined to hospitals within the Organ Procurement Center's administrative boundaries, though they extended beyond Turin, where the scandals occurred. We hypothesize that the surgeon's case elicits a stronger response because ICU staff work closely with surgeons and share a professional mission, making revelations of corruption within their ranks particularly demotivating.

We conclude our analysis by examining the potential impact of media coverage of corruption scandals on opposition to donation and the number of actual donors – defined as reported donors net of oppositions. Once medical staff identify a reported donor, the final decision rests with the donor's relatives, who must either consent to donation or present an opposition unless the potential donor had previously registered their willingness (or refusal) to donate organs. The analysis reveals no significant effect of media coverage of corruption scandals on opposition to donation.[6] This absence of an effect suggests that decisions regarding organ donation are primarily driven by moral and ethical considerations, which remain largely unaffected by news of corruption scandals. This implies that the attitudes and choices of potential donors and their families are informed decisions shaped over time, making it unlikely – at least in the short term – to be influenced by corruption-related media coverage. Our analysis of actual donors supports this conclusion. Media coverage has a negative and statistically significant effect on the number of actual donors when medical staff behavior is not accounted for. However, once we control for the medical staff's response – for instance, by including predicted reported donors as a covariate – the estimated effect of media coverage turns to a precisely estimated zero. This result confirms that while medical staff behavior is sensitive to corruption scandals, non-medical opposition to organ donation remains unaffected by such disclosures.

To rationalize the behavioral changes of medical staff, we outline a conceptual framework considering three potential mechanisms: (i) medical staff motivation; (ii) expected response by families of the deceased; (iii) peer pressure. A shock to perceived corruption induced by media reporting of corruption scandals is likely to undermine intrinsic motivation and morale and raise expected opposition by families, all of which will imply a reduction in the number of reported donors

[6] The analysis of oppositions accounts for the endogenous response of medical staff to media coverage of corruption scandals.

(via a reduction in the optimal effort by ICU staff). On the other hand, the same shock will reduce the cost of resisting peer pressure, thereby increasing optimal effort and the number of reported donors. While we cannot empirically disentangle the contribution of each factor, our findings show that motivation and expected opposition produce stronger impacts, at least in the short run, than the reduction in peer pressure.

Relationship to the Literature Our work contributes to several strands of literature. First, we contribute to the literature investigating physician behavior, emphasizing preferences, productivity, and behavioral responses to different incentives. As for preferences, one fundamental issue has been the study of intrinsic (or public service or pro-social) motivation in addition to private self-interest. The observation that physicians may have goals beyond maximizing their own utility dates back to Arrow (1963). However, while several theoretical contributions have adopted this assumption, it is only in the last decade that scholars have started providing empirical evidence on altruistic motivation. A general result of this literature is that intrinsic and extrinsic motivations characterize physicians. However, they also exhibit considerable heterogeneity, making them no more altruistic than the general population (Godager and Wiesen, 2013; Li, Dow and Kariv, 2017; Li, 2018; Crea et al., 2019).

Second, Silver (2021) studies the impact of peer pressure on the productivity of physicians in the context of Emergency Department care and shows that the behavior of individual physicians is heavily affected by pressure from peers: physicians are induced to work faster in faster peer-group environments; they cut back on time they spend per each case to keep up with their colleagues. As for incentives, given the importance of extrinsic rewards, it is unsurprising that financial incentives improve physicians' output, making them work harder (Clemens and Gottlieb, 2014). Our paper provides evidence on the behavior of medical staff working in Intensive Care Units, a setting similar to the Emergency Department, where the productivity and effort of physicians are crucial for outcomes, as is their intrinsic motivation. Furthermore, the Italian framework analyzed in this study is characterized by a fixed wage and the absence of financial incentives, which makes intrinsic rewards and peer pressure even more critical in driving optimal effort.

Third, we contribute to the economic literature on organ donations. Despite the importance of medical staff effort in terms of identifying and supporting reported donors emphasized by the medical literature (Thompson et al., 1995), the economic literature has focused almost exclusively on how institutional changes in the default choice or the allocation rules can improve the supply of organs (Johnson and Goldstein, 2003; Abadie and Gay, 2006; Kessler and Roth, 2012; Li, Hawley and Schnier, 2013). In discussing allocation rules, economists have also argued in favor of using incentives (especially financial incentives) to reduce, or even eliminate, the supply shortage (Becker and Elías, 2007; Howard, 2007; Thorne, 2006; Byrne and Thompson, 2001; Kaserman and Barnett, 2002). Roth (2007) explores how moral constraints can prevent the use of market incentives in the allocation of organs, while Elías, Lacetera and Macis (2015, 2019) provide experimental evidence on how information can modify moral attitudes. Recently, Akbarpour et al. (2020) have proposed and discussed a new matching algorithm to eliminate timing constraints in kidney exchange. Our paper contributes to the economic literature on organ donations by emphasizing the role of medical staff effort in influencing the supply of organs. We show that salient information on colleagues' behavior within the health care system substantially affects the supply of organs.

Fourth, we contribute to the literature by studying the impact of news on health outcomes and patients' perceptions. The role of media in shaping health-related behavior and outcomes is getting more and more attention nowadays. For instance, Ash et al. (2021) show that skeptical media narratives during the COVID-19 pandemic negatively affected individuals' health-related behavior. The role of media on organ donations has been discussed so far only by a strand of the medical literature focusing on how media reporting affects people's perception of transplants (Feeley and Vincent III, 2007; Morgan et al., 2007; Harbaugh et al., 2011). This literature provides evidence of the potential role of newspapers and entertainment television in shaping negative public attitudes and beliefs about organ donation through sensationalist and exaggerated reporting.[7] Corruption and other unethical behaviors disclosed by media can also undermine citizens' trust in public health care (Radin, 2013; Alsan and Wanamaker, 2018). Concerning this literature, the key innovation of our work is to provide heterogeneous estimates on the reduction of workers' motivation when they discover a colleague or a hospital manager is corrupt. Additionally, we find that individual and family opposition

[7] This literature labels the negative influence of media on organ donations as the "Panorama effect." The expression originates from a 1980 BBC TV broadcast titled "Panorama" meant to question the validity of "brain death" criteria. Public outrage over this show in the United Kingdom was widespread, and sudden drops in organ donations followed. More than one year was needed for the adverse effects to be completely reabsorbed (Matesanz, 1996).

to organ donations do not respond to corruption scandals, suggesting that organ donation is an informed choice that has matured over time.

3.2 Background

We investigate whether media coverage of corruption influences the behavior of ICU medical staff. In this section, we provide essential background information on two main issues: the organ procurement process involving medical staff working in ICU to help identify the outcome of their work; the corruption scandals generating important waves of media reporting, which may affect perceived corruption by medical staff.

3.2.1 The Process of Organ Procurement in Italy

The Italian system of organ procurement is part of the NHS. The NHS is a universal public scheme for providing health care to citizens. As in other countries, such as Spain, the NHS gives regional governments a key role in managing resources at the local level (Turati, 2013). Although the framework legislation is defined at the country level, the organ procurement system is highly decentralized. It largely relies on regional Organ Procurement Centers to manage organ donation, retrieval, and transplant. This complex procedure involves delicate technical steps where a single mistake can jeopardize the outcome.

Figure 3.1 describes the main stages of turning a critically-ill patient into an organ donor.[8] The first step of the organ procurement process is the identification and constant monitoring by ICU medical staff of all patients who are irreversibly losing their brain functions; we refer to these individuals as *potential* donors (Step 1).[9] When brain death occurs, the anesthetist in charge of the patient must inform the hospital's health manager of the diagnosis and initiate an observation procedure of at least six hours, normally followed by a confirmation given by a medical board of the death status (Step 2). This process, independent of the

[8] The reconstruction of the organ procurement process is based on Venettoni (2007) and Regione Lazio (2008).

[9] Brain or encephalic death is the "irreversible cessation of cerebral and brain stem function characterized by absence of electrical activity in the brain, blood flow to the brain, and brain function as determined by clinical assessment of responses. A brain-dead person is dead, although his or her cardiopulmonary functioning may be artificially maintained for some time (World Health Organization, 2009)."

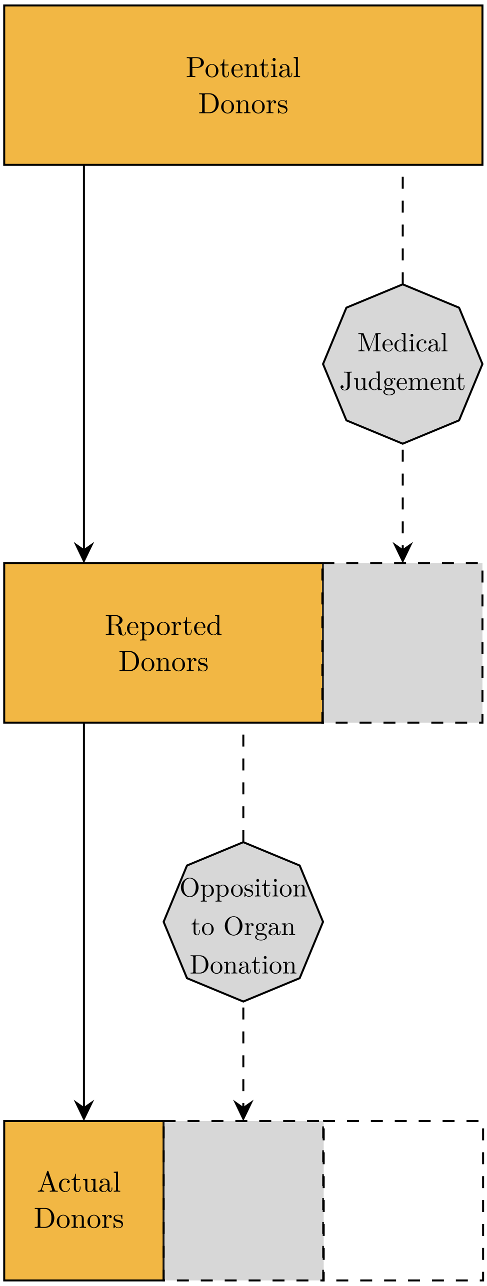
donation, is intended to certify a patient's death from a *legal* point of view. A deceased person with no medical contraindications to donation and whose brain death has been assessed becomes a *reported* donor (World Health Organization, 2009). The ICU specialist and/or the local coordinator of the hospital must report to the regional Organ Procurement Center of reference all the information available on the reported donor. This step should be done as soon as possible to avoid deadlocks in the process, to minimize the risk of both organ deterioration and *clinical* death (which occurs when the heart stops beating), and to allow the Organ Procurement Center to promptly consult with the allocation lists (Step 3). During this stage, it is essential to carefully monitor patients via mechanical ventilation and other life support measures to prevent organ deterioration that might originate from cardiac arrest. The maintenance of the reported donor requires an active role and an intensive effort from all involved staff, especially the anesthetists working in the ICU. Staff motivation is crucial during the period between *brain* death and legal certification, leading up to the potential transplant. To convert a *reported* donor into an *actual* donor, medical personnel must conduct two different evaluations. The first is a legal evaluation that aims to ascertain consent to donation (Step 4). In the absence of the patient's expressed will, it is up to the family to give or deny consent to donation. If the deceased is a ward of the state, judicial permission must be obtained. In addition, a clinical evaluation must be carried out to assess the suitability of the potential donor to donate and the functionality of organs (Step 5). If there is no opposition by the patient or the relatives and the patient is clinically suitable for donation, the medical staff proceeds with organ retrieval. The patient becomes an actual donor.[10] Finally, a transplant can involve more than one organ; hence an actual donor usually benefits more than one patient on the waiting list.

3.2.2 The Corruption Scandals

To measure the impact of media reporting on perceived corruption and its effect on reported donors, we start from data and results in Le Moglie and Turati (2019). The authors provide a map of corruption episodes related to the health care sector reported in the news at the regional level during the last twenty

[10] The precise definition of an actual donor is: "A deceased or living person from whom at least one solid organ or part of it has been recovered for transplantation (World Health Organization, 2009)." Here, we are considering just cadaver donations. Notice that, despite being allowed by Law 458/1967, kidney donations from living persons are pretty rare in Italy (Frascà et al., 2009).

Figure 3.1: The Process of Organ Procurement in Italy



1. **Donors identification and monitoring**

   Health personnel in ICUs identify and monitor patients in potential evolution toward brain death.

2. **Brain death diagnosis, assessment, and certification**

   Brain death is always diagnosed when there is an irreversible cessation of all brain function, independently of organ donation. It follows an observation period of at least 6 hours to assess the persistence of the state of the death, which ends with death certification.

3. **Report to the reference coordination center**

   The intensive care specialist and/or the local coordinator promptly report potential donors to the regional or interregional coordination center, which will proceed to consult the allocation lists.

4. **Legal evaluation of the donor: detection of consent**

   In the absence of the patient's expressed will, consent to donation is asked for from the family. Permission of the judicial authority is needed in case the potential donor is a ward of the state.

5. **Clinical evaluation of the donor**

   A clinical evaluation has to be carried out in order to determine the suitability of the potential donor to donate and the functionality of the single organs.

Note: This figure depicts the sequential stages of the organ procurement process, beginning with the identification of potential donors in the intensive care unit (ICU) and concluding with the clinical evaluation of donor suitability. Critical decision points include the assessment of medical eligibility, the determination of legal consent, and potential opposition to donation. Dashed lines indicate stages where procedural delays or complications may occur, whereas solid arrows indicate the standard progression through the process.

years. According to this study, most episodes refer to petty corruption cases, with one article at most reported in the news. However, two cases stand out as those with the highest coverage. Both hit the regional health care system in Piedmont and the Aosta Valley, two regions located in the North West of

Italy (see Figure 3.B.1).[11] The two cases involved the Hospital 'Molinette' in Turin, the biggest in Piedmont and the third-largest nationwide. The Hospital 'Molinette' hosts the Regional Procurement Center serving Piedmont and the Aosta Valley. The scandals involved high-level personnel within the hospital and allow a comparison between two different episodes: an episode of corruption involving a hospital manager and another episode directly involving surgeons.

News about the scandals made media headlines at local and national levels. The first scandal broke out in December 2001 (Figure 3.2), when the hospital CEO Luigi Odasso was caught accepting bribes in his office. In January 2002, Odasso was accused of favoring a patient on a waiting list for a kidney in exchange for bribes. Although the corruption targeted transplant activities, the events did not directly involve the specific units responsible for organ donations or any hospital medical staff. In November 2002, Hospital 'Molinette' again came under scrutiny due to a new corruption scandal. Michele Di Summa, a well-known heart surgeon in charge of the regional Heart Transplant Center, was accused of accepting large sums in exchange for using cardiac valves supplied by *For.Med* Padova (which sold Brazilian valves produced by the company *Tri Technologies*) and *Ingegneria Biomedica* (which sold locally produced devices). Also, in this case, the media highly covered the news. Figure 3.3 displays the front page of 'La Stampa' on November 5, 2002; the main article is about the corruption scandal involving Di Summa.[12] When defects were found in some of these valves, particularly those produced by the Brazilian company, Di Summa and his colleague Poletti were arrested. Charges leveled against Di Summa accused him of being well aware of the Brazilian valves' defects and failing to immediately deliver to the Ministry of Health the list of receiving patients following the ministerial decision to withdraw the valves from the market. The Brazilian valves were believed to cause the death of patients with the valve implanted. However, later judges ascertained that this was not the case. Nevertheless, the scandal had strong resonance at the regional and national levels and, as reported in Section 3.4, was intensively covered by local and national media.

[11] Despite being separated since 1948, when the Aosta Valley gained its special statute, the two regions were and still are closely connected. For instance, from an administrative point of view, municipalities in the Aosta Valley were part of the province of Turin before the special statute endorsed regional autonomy. Still, nowadays, many people move from and to Piedmont for daily commuting. Moreover, many patients from the Aosta Valley seek care in Piedmont, especially in Turin, as the city is known for its highly specialized hospitals and is only 120 kilometers from Aosta.

[12] 'La Stampa' is the most-read newspaper in Piedmont and Aosta Valley. Section 3.3 will provide more details on newspaper diffusion across the Italian territory.

Figure 3.2: Newspaper Coverage of CEO Scandal

LA STAMPA GIOVEDÌ 20 DICEMBRE 2001 5 CRONACHE

REAZIONI GIUNTA REGIONALE STRAORDINARIA E I COMMENTI DELLE FORMAZIONI POLITICHE

DOMANI UN SOSTITUTO

Breve riunione straordinaria ieri sera della Giunta piemontese. L'esecutivo non ha nominato né un sostituto né un commissario straordinario per l'ospedale: dalla Procura non è giunto ancora l'atto ufficiale sull'arresto del manager. Forse verrà nominato domani

«NON FERMIAMO TUTTO»

An esprime «profonda amarezza» per l'arresto del direttore generale delle Molinette. E sottolinea come ora sia importante che «l'ospedale non abbia a subire alcun rallentamento nelle prestazioni». L'ha detto il capogruppo in Consiglio regionale, Ennio Galasso

SUBITO UN COMMISSARIO

«La situazione è gravissima, e in particolare è gravissimo che la giunta non abbia ancora provveduto a nominare un commissario». Così il consigliere regionale Ds, Pietro Marcenaro, è intervenuto all'assemblea piemontese sulla vicenda Odasso

COMMISSIONE D'INCHIESTA

«Una commissione d'inchiesta»: l'ha chiesto in Consiglio regionale il gruppo di Rifondazione. «Episodi di tale gravità - ha detto Mario Contu - sollevano dubbi sull'efficacia dei controlli che la Regione dovrebbe esercitare sull'operato dei manager»

Arrestato mentre incassa la tangente da 15 milioni

Torino, in carcere il direttore generale dell'ospedale «Molinette»

Alberto Gaino

TORINO

Di buste che passavano dalle mani di imprenditori a quelle di Luigi Odasso, 52 anni, direttore generale delle Molinette, dietro la porta chiusa dell'ufficio, le microtelecamere della Guardia di finanza ne avevano «immortalate» un certo numero. Ma non si aveva la certezza che, dentro, vi fossero banconote. La mazzetta si è improvvisamente materializzata ieri mattina, poco dopo le 11, quando il pm Giuseppe Ferrando ha dato il via alle fiamme gialle: la busta era già stata riposta nel cassetto della scrivania e gli uomini del capitano Ermanno Palombini sono andati a colpo sicuro. Una volta aperta, vi hanno trovato 15 milioni in contanti in banconote di piccolo taglio. La rata mensile dell'appalto di apparecchiature per la dialisi che l'imprenditrice cuneese Renata Prati aveva appena consegnato a Odasso. Costui ha avuto la prontezza di spirito di dire: «Sono soldi miei». In ogni caso si trovavano ormai in suo possesso.

I due sono stati arrestati in flagranza di reato, come nove anni fa un certo Mario Chiesa. A quel tempo la mazzetta valeva il 3 per cento, o giù di lì, dell'ammontare degli appalti. Ora siamo saliti al 10 per cento. Sembra trascorsa una vita dalla corruzione in prima pagina e colpisce tanto di più che un manager pubblico, così esposto come il direttore generale della prima azienda ospedaliera del Piemonte e terza d'Italia, potesse trattare e incassare tangenti nel proprio ufficio. Senso dell'impunità? Eppure il pm Ferrando gli stava alle calcagna da dopo l'estate, da quando, a settembre, aveva ricevuto una denuncia piuttosto dettagliata sui maneggi del manager pubblico. Fu così che mise al lavoro gli uomini della compagnia torinese della Finanza, gli stessi che avevano lavorato con lui in un'altra storia di mazzette, quella volta ai piani bassi, anzi sotterranei dell'immenso ospedale: le tangenti sulle salme.

Odasso tuonò contro i corrotti, ora tocca a lui rispondere di corruzione continuata. In concorso con Renata Prati e altre 8 persone (sette imprenditori e un collaboratore del direttore generale). Si conosce l'identità di quest'ultimo: Aldo Rosso, dirigente dell'ufficio tecnico dell'ospedale e uomo ombra del manager. Per l'intera giornata la Finanza ha effettuato perquisizioni e, in almeno in un caso, a Nizza, dove Odasso è nato e ha mantenuto radici e prima casa, è arrivata in tempo per sequestrare un'auto in fuga «piena di documenti». Nel più classico dei copioni di Tangentopoli che si ripropone.

E' stato ripreso da una telecamera nascosta mentre riceve la somma da un'imprenditrice La donna ha confessato la corruzione Altri otto indagati

Vuol dire che Odasso o chi per lui - mentre i due uomini in borghese delle «fiamme gialle» gli chiedevano conto della tangente con in mano un decreto di perquisizione del suo ufficio - ha avuto il tempo di dare l'allarme ai suoi. Da Torino partivano continui ordini per i sottufficiali impegnati nei vari sopralluoghi: «Fatevi aprire il caveau», è stato l'input dato agli uomini nella villa di via Verdi, immersa nel verde di una grande proprietà su una collina che domina Nizza Monferrato. E nel centro di Torino, all'ultimo piano dell'elegante appartamento-studio di piazza Carignano, la ditta che arrivava dal comando della Guardia di Finanza era la stessa: «Guardate dentro la cassaforte».

A Cuneo, in frazione Madonna dell'Olmo, si perquisiva la sede della Selefar e, non lontano, l'abitazione della cinquantatreenne Prati. Nel frattempo la signora è stata accompagnata a palazzo di giustizia, al sesto piano, nel piccolo ufficio del pm dove è stata sentita per oltre due ore dal magistrato: il suo avvocato, Ezio Audisio, se n'è andato nel più totale mutismo mentre la donna veniva riaccompagnata al comando torinese della Guardia di Finanza, in attesa degli sviluppi delle prossime ore e giorni suggeriti dalle sue ammissioni. Ufficialmente la donna avrebbe ammesso alcuni regali a Odasso, in realtà ha collaborato più attivamente alle indagini e conta di evitare il carcere.

Il manager al centro dell'indagine viene interrogato stamane. Il suo legale, Andrea Galasso, ha per le mani soltanto il decreto di sequestro, alcune laconiche righe di testo, oltre alle notizie di fonte giornalistica per valutare la posizione di Odasso. Sa quanto basta, però, per escludere che l'arresto fosse necessario. E dichiara: «Di fronte a un'ipotesi di reato che prevede l'arresto facoltativo, la misura restrittiva della libertà, sia pure temporanea, mi pare eccessivamente afflittiva».

Luigi Odasso mentre viene portato via in auto A destra l'imprenditrice che ha confessato

L' INFORMATRICE FARMACEUTICA

CUNEO. Renata Prati è titolare di una delle più importanti società di fornitura di materiale sanitario del Piemonte. Cinquantatrè anni, sempre elegantissima, ambiziosa, preparata, ha percorso a grandi passi le tappe - partendo da quella di informatrice farmaceutica - di una brillante carriera di imprenditrice. L'azienda che ha fondato ha sede a Madonna dell'Olmo di Cuneo, in via Maestri del Lavoro. Fornisce a ospedali e cliniche impianti e apparecchi medicali ed elettromedicali. I prodotti «top» forniti dalla «Selefar» sono le pompe da infusione per farmaci da somministrare endovena, apparecchi di alta qualità, di fabbricazione giapponese, delle quali la società cuneese ha conquistato un'importante fetta di mercato in regione: sono numerosissime le Aziende sanitarie locali e gli ospedali piemontesi che hanno rapporti commerciali con la Selefar, il cui giro d'affari annuo è di diversi miliardi.

Ieri la guardia di Finanza ha perquisito gli uffici della società a Madonna dell'Olmo (dove sarebbe stata sequestrata anche una folta documentazione) e l'abitazione dell'imprenditrice, in frazione Spinetta di Cuneo. [m. bo.]

I PARTICOLARI DELL'INCHIESTA CONDOTTA DALLA GUARDIA DI FINANZA NEL CAPOLUOGO PIEMONTESE

Odasso ieri mattina dietro al giudice Giuseppe Ferrando mentre viene arrestato nel suo ufficio alle Molinette In basso finanzieri con gli scatoloni di documenti sequestrati nell'ospedale

Mesi di microspie e di filmati «Così lo abbiamo incastrato»

Pedinato da settembre dopo la denuncia fatta dal parente di un fornitore corrotto

Le voci registrate: «Le ho portato il contenuto» La mamma di Odasso «Solo una bolla di sapone»

le indagini

TORINO

SIAMO intervenuti oggi perché c'era una buona visuale»: è stato il caustico commento del procuratore aggiunto Mario Griffey, alle sette della sera, in chiusura della giornata di arresti in flagranza di reato, perquisizioni, interrogatori. E' più di un'allusione al set di microtelecamere che da due mesi riprendeva Luigi Odasso mentre si muoveva nel suo ufficio, riceveva dipendenti e collaboratori e almeno 6 imprenditori con la busta nella borsa per lui, prima del fatale ingresso per quella porta della signora Prati da Cuneo.

Le immagini finivano su un pulmino attrezzato di monitor che la Guardia di Finanza aveva parcheggiato nei paraggi dell'ospedale. Le voci integravano i gesti che, pur al riparo di porte e finestre sempre chiuse, restavano furtivi. «Le ho portato il convenuto». Le trattative c'erano già state. Come la busta è passata di mano, due finanzieri in borghese si sono precipitati dalla strada all'ingresso della direzione generale, un metro oltre la grande cancellata dell'ospedale, e sono saliti con il cuore in gola al primo piano. «Guardia di Finanza, siamo qui per una perquisizione», hanno stoppato sul nascere ogni richiesta delle due commesse in divisa blu sedute sotto un grande quadro di San Giovanni Battista (patrono dell'ospedale) e si sono precipitati nell'ufficio di Odasso, interrompendo i saluti fra il direttore generale e l'imprenditrice. La mazzetta era già stata riposta nel cassetto della scrivania: è stata prelevata con estrema attenzione per essere sottoposta al controllo delle impronte digitali.

L'inchiesta nasce con una cadenza da déjà vu: la denuncia di un parente di un imprenditore corrotto cui si aggiunge, strada facendo, l'esposto del titolare di un'azienda di manutenzione del verde, piante e fiori che aveva presentato un'offerta con un ribasso del 40 per cento per un appaltino e si era visto chiedere la tangente. Si montano due telecamere a circuito chiuso, con regia nelle immediatezze dell'ospedale. Si ascoltano le telefonate. E' un tripudio di microspie nello studio del direttore generale che continua imperterrito i suoi affari.

Gli appalti sotto inchiesta sono meno di 10, da interventi di ristrutturazione edilizia a forniture ospedaliere: su un bilancio di circa 700 miliardi le spese per acquisti coprono il 30 per cento. Alla Guardia di Finanza assicurano che le indagini sono appena iniziate: ora saranno controllati tutti gli appalti. E a breve si ventila un nuovo «round» di arresti.

Aggiungono in procura: «Alla luce di questi fatti sarà rivisitata anche un'apparente piccola storia giudiziaria», arenatasi sull'accusa di millantato credito a un legale valsusino che aveva promesso un rapido trapianto di reni a un dializzato.

Per non mettere sul chi va là i diretti interessati, il pm Ferrando aveva sinora evitato di inviare i collaboratori all'Ufficio del Registro, per acquisire la «collezione» degli appalti delle Molinette dal '98, quando Odasso ne diventa il direttore generale, in poi. Né a Palazzo di Giustizia si è fatto un calcolo delle tangenti che il manager potrebbe avere incassato. «Ora non abbiamo questa preoccupazione», rimbalza dal comando delle Fiamme Gialle. Il primo impegno di stamane, per il pm e la squadra del capitano Palombini, è di cominciare a vagliare i documenti sequestrati nella villa di Nizza Monferrato. Prima delle 10,30, l'ora in cui è fissato l'interrogatorio di Odasso.

Il direttore generale è stato rinchiuso per la notte in una cella del carcere torinese delle Vallette, l'imprenditrice ha ricevuto un altro trattamento (altra conferma della sua buona disposizione a collaborare con l'inchiesta): le hanno approntato un letto in una stanzetta del comando della Guardia di Finanza in attesa dell'udienza di oggi dal gip di turno. C'è tempo di registrare l'ultima polemica, questa volta di un consigliere regionale dei Ds, Roberto Placido: «Perché il vecchio revisore dei conti dell'ospedale se n'era andato e non è stato sostituito?». Il sipario della diretta tv sulla mazzetta cala e le ultime parole sono di mamma Odasso, sicura dell'innocenza del figlio: «E' solo una bolla di sapone». [al.ga.]

Note: The figure shows an article covering the CEO corruption scandal. Source: 'La Stampa' – December 20, 2001. Title: Arrested While Cashing In a 15 Billion Bribe.

Figure 3.3: Newspaper Coverage of Surgeon Scandal

IMPRENDO
Soluzioni pensate per la piccola impresa.

LA STAMPA

MARTEDÌ 5 NOVEMBRE 2002. ANNO 136. N. 303. € 0,90 IN ITALIA (PREZZI TANDEM ED ESTERO IN ULTIMA) • SPEDIZIONE ABB. POSTALE 45% ART. 2 COMMA 20/B LEGGE 662/96 - TO • www.lastampa.it

TROVI IMPRENDO PRESSO:
UniCredit Banca Banca CRT
UniCredit Banca Cariverona
UniCredit Banca Credito Italiano
UniCredit Banca Rolo Banca 1473
UniCredit Banca Cassamarca
UniCredit Banca CRTrieste
O Local

ISLAMICI AL POTERE

L'Europa: giudicheremo la Turchia in base ai fatti

L'EUROPA ENTRA NELLA COMUNITÀ TURCA

BRUXELLES. «L'Europa sbaglierebbe a giudicare il prossimo governo turco per il fatto che è espressione di una base islamica».

SONO ACCUSATI DI CONCUSSIONE. I PRIMARI ARRESTATI IERI POMERIGGIO, ALL'USCITA DALLA SALA OPERATORIA

Torino, due cardiochirurghi in cella per tangenti

Avrebbero preso mazzette per la fornitura di valvole cardiache alle Molinette

TORINO. Sono stati arrestati ieri pomeriggio i due cardiochirurghi dell'ospedale Molinette al centro di un'inchiesta per tangenti. Il professor Michele Di Summa, tra l'altro responsabile del centro trapianti della Regione Piemonte, e il suo collega Giuseppe Poletti, responsabile della seconda camera operatoria di Cardiochirurgia, sarebbero accusati di concussione. Entrambi sono anche docenti universitari. I due primari, che ieri hanno regolarmente lavorato in sala operatoria, avrebbero preteso tangenti (si parla di 750 milioni di vecchie lire) per assegnare, in quanto componenti della commissione aggiudicatrice, la fornitura di valvole cardiache a due diverse società.

Accossato, Ballario, Gaino, Longo, Pietropinto e Tropeano A PAGINA 7 E IN CRONACA

PERSONAGGI

LO STAKANOVISTA E IL TESSITORE

Una media di 500 interventi l'anno Avevano creato un reparto d'avanguardia punto di riferimento e di speranza per centinaia di malati

Angelo Conti A PAGINA 7

Lorenzo Mondo

LA CADUTA DEI FALSI EROI

E' tanto grossa che vien da dire: non è vero. Lascia sbigottiti che le Molinette si trovino un'altra volta a fare rima con manette. Quelle strette ai polsi di due cardiochirurghi, accusati di avere preteso tangenti da fornitori di valvole cardiache, oltretutto di dubbia qualità. Sarebbero tangentari come Odasso, scivolato tra le braccia della giustizia meno di un anno fa. Come se la Sanità contemplasse, insieme alle trasfusioni, speciali pratiche vampiriche.

Odasso faceva il manager, il contabile, non aveva un rapporto diretto con i pazienti e sconvolge meno l'idea che pasticciasse con i soldi. Ma un medico, uno che le viscere dell'uomo le vede palpitare?

Nella mattinata avevano ancora eseguito due delicati interventi. E sono tanti i malati che, dopo esserci passati o in attesa di farlo, si prodigavano in benedizioni, li salutavano come angeli salvatori. Questi specialisti di gran nome, capaci di aggiustarti il cuore rotto, avrebbero il loro cuore irrimediabilmente guasto? Erano così certi dell'impunità da trascurare la premonitrice caduta di Odasso? E non sentivano l'orgoglio di lavorare in un ospedale di eccellenza, apprezzato in Italia e all'estero per le sue prestazioni, immeritevole di colpi bassi? Si affaccendavano ad arrotondare i loro non ignobili guadagni senza abbassare gli occhi, per pudore, davanti a tanta gente doppiamente povera, di soldi e di salute, che gli passava davanti ogni giorno?

Non è vero. Si tratterà, come dicono, di maldicenza o congiura. Non hanno contribuito minimamente a sfigurare il volto di una città che ha già troppe afflizioni. A smentire un costume antico, forse remoto, di severa correttezza. Vale per loro come per ogni cittadino la normale, civile presunzione di innocenza. Ma al di là di questo ci sono troppe altre ragioni per augurarsi che la nuvola di maleficio di dissolva. Per loro, per le Molinette, per Torino.

Note: The figure shows an article covering the surgeon corruption scandal. Source: Front page of 'La Stampa' – November 5, 2002. Title: Turin, Two Heart Surgeons Behind Bars for Receiving Bribes.

3.3 Data

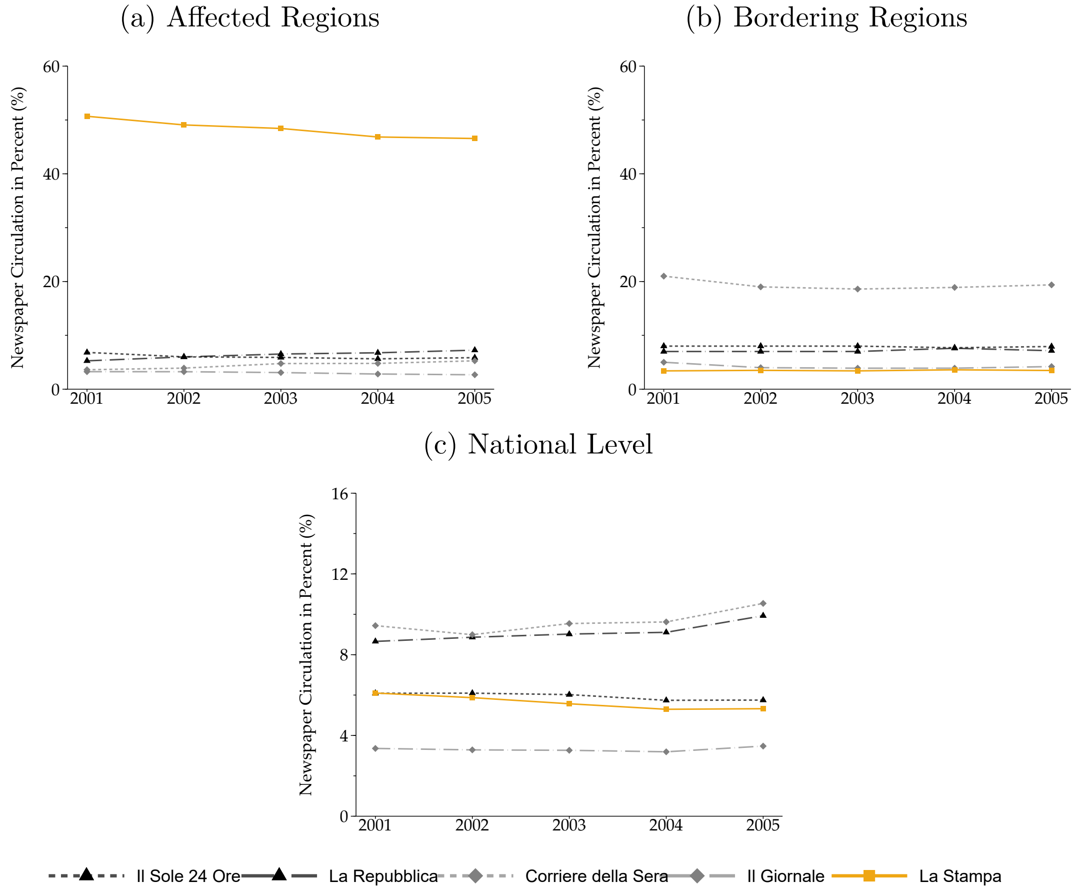
We construct a new data set based on several sources of information. First, to conduct counterfactual analyses, we focus on the geographical area where the scandals took place – i.e., the two regions of Piedmont and Aosta Valley – and the two bordering regions of Lombardy and Liguria. Figure 3.B.1 shows a map of Italy with the location of the regions analyzed in this study.

We start by collecting information on the number of reported and actual donors provided by the three relevant regional Organ Procurement Centers for Piedmont and the Aosta Valley, Lombardy, and Liguria. Data are collected at the hospital level and monthly to cover the period from 2001 to 2005 spanning both the pre- and post-scandals periods.

To capture media coverage of corruption scandals, we retrieve the number of articles published by the two daily newspapers 'La Stampa' and 'Il Corriere della Sera.'[13] Figure 3.4 motivates the choice of these two newspapers by representing market shares for the five leading Italian daily newspapers. Figure 3.4-a shows newspaper circulation in the two regions directly affected by the scandals. Figure 3.4-b focuses on bordering regions. Finally, Figure 3.4-c depicts the national newspaper circulation. Two facts are striking. At the national level, the five newspapers report coverage rates between 3 and 11 percent, with the highest share observed for 'Il Corriere della Sera.' In Piedmont and the Aosta Valley, 'La Stampa' is the absolute market leader with a share of around 50 percent. It represents one out of two newspapers distributed and read across the regions hit by the corruption scandals of interest for this study. The newspaper 'Il Corriere della Sera' is the market leader in the bordering regions of Lombardy and Liguria. This evidence suggests using 'Il Corriere della Sera' and 'La Stampa' to proxy media coverage. We use different keywords to browse the online archives. In particular, we consider the words 'Odasso', 'Di Summa', and 'heart valves' to obtain the monthly number of articles about the two corruption cases. We deal with possible concerns underlying the choice of the two specific newspapers. First, to consider differences in the circulation of the two newspapers, we construct a province-specific weighted average of the number of articles in 'La Stampa' and 'Il Corriere della Sera.' The coverage rate for each newspaper at the local level is

[13] We also consider the provincial editions of 'La Stampa' to ensure provincial variation in the number of news. The newspaper includes a national section throughout Italy, a regional section common to Piedmont and the Aosta Valley, and a specific provincial section for each province.

Figure 3.4: Newspaper Coverage in Italy



Note: The figure shows newspaper circulation in Italy disaggregated across the five largest newspaper outlets. Sports newspapers are excluded from the analysis. Affected regions include Piedmont and the Aosta Valley. Bordering regions include Lombardy and Liguria. Data for bordering regions for 2005 are missing. In constructing the weighted number of newspaper articles we impute the 2005 data by the mean of the preceding four years.

used as a weight.[14] Second, we test the robustness of our findings with the use of alternative sources of information. In particular, we collect news coverage of both scandals broadcasted by the *Telegiornale Regionale* (TGR), the regional daily TV news.[15] Table 3.1 reports the correlation between the number of articles or news concerning the surgeon at the Heart Transplant Center (*Surgeon Scandal*) and the number of articles or news on the case of corruption involving the Hospital 'Molinette' CEO (*CEO Scandal*). The high level of correlation between different

[14] All results in the following analyses are robust to using the raw number of articles instead of the weighted number.

[15] Information about corruption cases with a strong geographical connotation is usually discussed in the regional daily TV news. The regional daily TV news is transmitted by the national television channel *Rai 3*, which turns into a regionally differentiated broadcast for the TGR three times per day. Notice that our focus on two traditional media, like newspapers and TV news, is justified by the fact that the share of people using the internet was between 27 percent in 2001 and 35 percent in 2005, and social media were largely unknown.

sources of information displayed in the table reinforces the idea that the choice of the source of information is likely to be free of concerns in our framework. Table 3.2 reports summary statistics for all the variables in our sample.[16] The

Table 3.1: Correlation Matrix for Different Sources of News

	La Stampa	Corriere	Weighted	TV News
	(1)	(2)	(3)	(4)
Panel A: Coverage of Surgeon Scandal				
La Stampa	1			
Corriere	0.851***	1		
Weighted	0.955***	0.848***	1	
TV News	0.675***	0.899***	0.692***	1
Panel B: Coverage of CEO Scandal				
La Stampa	1			
Corriere	0.851***	1		
Weighted	0.955***	0.848***	1	
TV News	0.675***	0.899***	0.692***	1

Note: This table shows the pairwise correlation of coverage of corruption scandals by different newspapers and media outlets. The weighted number of newspaper articles is constructed by weighting newspaper articles related to the scandals according to the market share shown in Figure 3.4. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

information is collected monthly from 2001 to 2005 for all the 22 spoke hospitals of the Organ Procurement Center, which is the hub of the system in Piedmont and the Aosta Valley (affected regions). In addition, the same data are collected for the 50 spoke hospitals of the Organ Procurement Centers in Lombardy and Liguria (bordering regions). The average number of reported donors for each hospital in the affected regions is 0.73 per month, higher than the value of 0.55 observed for the bordering regions. The average number of articles on the surgeon scandal is 1.53 per month over the whole period, with a standard deviation of 5.5 underlying the variation over time in the coverage of the scandal. Similarly, in the period covered by our analysis, we observe about 1.64 articles per month on the CEO scandal, with a standard deviation of 6.4. As for newspaper circulation, there are more than 50 copies of 'La Stampa' per 1,000 inhabitants in the affected regions compared to about six copies in bordering regions. The picture is different for 'Il Corriere della Sera,' with about 4.7 copies per 1,000 inhabitants in affected regions and approximately 37.3 copies per 1,000 inhabitants in bordering regions.

[16] In Appendix Tables 3.A.1 and 3.A.2 we present descriptive statistics differentiating regions exposed to corruption scandals and unexposed bordering regions before and after the corruption scandals.

Table 3.2: Descriptive Statistics

	Affected Regions		Bordering Regions		
	Mean (1)	SD (2)	Mean (3)	SD (4)	Difference (5)
Organ Donations					
Reported Donations	0.733	1.125	0.546	1.121	0.267***
Opposed Donations	0.214	0.500	0.110	0.389	0.200***
Actual Donations	0.470	0.816	0.379	0.841	0.050***
Newspaper Coverage (# Articles)					
Surgeon Scandal	1.529	5.532	0.994	3.331	2.052***
Surgeon Scandal (3-Month Avg.)	1.581	4.458	1.027	2.676	2.124***
CEO Scandal	1.637	6.406	1.261	3.979	1.774***
CEO Scandal (3-Month Avg.)	1.686	5.618	1.302	3.439	1.822***
Newspaper Circulation (per 1,000 inhabitants)					
La Stampa	51.389	11.552	6.676	12.967	61.167***
Il Corriere Della Sera	4.710	1.040	37.329	73.353	-5.875***
Observations	1,320		3,000		
Hospitals	22		50		

Note: This table shows descriptive statistics for the sample analyzed in this study. The difference reported in Column (5) is the coefficient obtained by regressing an indicator for affected regions on each variable, controlling for year, month, and hospital fixed effects. Significance levels are constructed from bootstrapped (1,000 replications) and wild clustered standard errors at the hospital level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

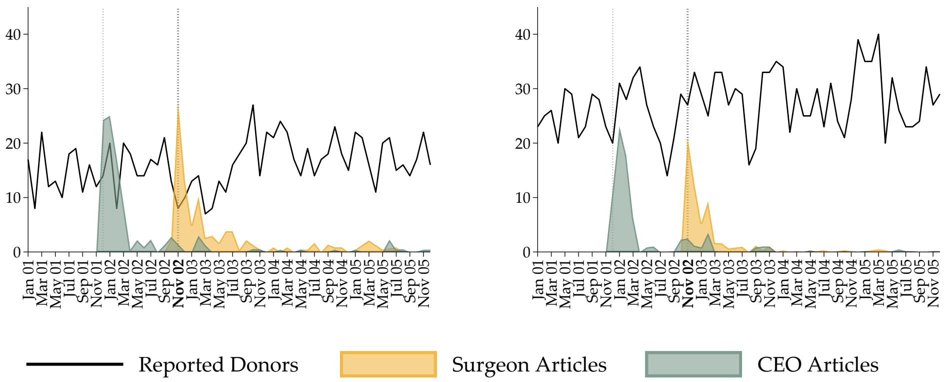
Figure 3.5 provides a first descriptive analysis of the possible relation between the total number of reported donors (black line) and the media coverage of the two corruption cases for both the affected (Figure 3.5-a) and bordering (Figure 3.5-b) regions. Despite the erratic trend, the analysis of affected regions displays three different periods.[17] A stable average level of reported donors characterizes a first period starting from January 2001. The onset of the case of corruption involving the Hospital 'Molinette' CEO (first vertical gray line) does not imply changes in reported donors. The second period started in November 2002 (second vertical gray line), at the onset of the scandal involving the surgeon Di Summa. Contemporaneously with the media coverage of this case (yellow line), the average number of reported donors decreases. Moreover, peaks in the number of articles are usually coupled with peaks of opposite signs in the number of reported donors – e.g., November 2002 and, vice versa, May 2003. The third period, starting from August 2003, is characterized by levels of reported donors closer to (or even higher than) those registered in the period before November 2002. Figure 3.5-b seems to support the intuition of our identification strategy as it displays a stable (erratic) trend over the analyzed period for regions not directly affected by the

[17] The erratic trend reflects the low frequency of reported donor events.

scandals. This stability suggests a negligible impact of corruption news when this news involves other regional healthcare systems.

Figure 3.5: Media Coverage and Reported Donors

(a) Affected Regions (b) Bordering Regions



Note: The figure shows the number of reported donors (solid black line) and the weighted number of newspaper articles related to the surgeon (yellow) and CEO (green) corruption scandals in affected and bordering regions. The weighted number of newspaper articles is constructed by weighting the provincial number of newspaper articles on both scandals by the respective yearly newspapers market shares shown in Figure 3.4. The first vertical dashed line refers to the occurrence of the CEO scandal, while the second vertical dashed line marks the occurrence of the surgeon scandal.

3.4 Empirical Strategy

As explained in Section 3.2.1, the organ procurement process depends on ICU staff to identify and monitor potential donors. Perceived corruption may affect this task by reducing intrinsic motivation or dampening morale. To assess its impact on staff behavior, we employ a difference-in-differences (DiD) design. We assume that media reports on corruption scandals influence perceived corruption. Hospitals in Piedmont and the Aosta Valley are designated as “affected regions”, while those in Lombardy and Liguria serve as the control group. Based on this logic, we implement a DiD model of the following form:

$$\begin{aligned} \text{Donors}_{ht} = {} & \alpha_h + \alpha_t + \\ & + \beta_1 \,\text{Art. CEO}_{p(h)t} + \beta_2 \,\text{Art. Surgeon}_{p(h)t} \\ & + \beta_3 \,\text{Art. CEO}_{p(h)t} \cdot \text{Affected}_h + \beta_4 \,\text{Art. Surgeon}_{p(h)t} \cdot \text{Affected}_h \\ & + \boldsymbol{X}_{p(h)t}\boldsymbol{\gamma} + \varepsilon_{ht}, \end{aligned} \tag{3.1}$$

Donors_{ht} denotes the monthly number of reported organ donors identified by ICU staff at hospital h in period (month-year) t. We proxy perceived corruption using the number of newspaper articles covering two scandals at Hospital 'Molinette' between 2001 and 2005. Art. $\text{CEO}_{p(h)t}$ represents the weighted number of articles in province $p(h)$ at time t related to the CEO scandal, while Art. $\text{Surgeon}_{p(h)t}$ captures the coverage of the surgeon scandal. We consider the number of articles on either scandal as a shifter for the effort of ICU medical staff.

Affected_h is a binary variable equal to one if hospital h is located in Piedmont or the Aosta Valley. $\boldsymbol{X}_{p(h)t}$ is a vector of control variables, including population, newspaper circulation at the provincial level, and hospital-specific time trends. To account for unobserved factors, the model includes hospital, α_h, and time fixed effects, α_t.[18] Standard errors are clustered at the hospital level.

Our primary focus is on β_3 and β_4, which capture the differential effect of weighted monthly media coverage of the CEO and surgeon scandals on ICU staff behavior in affected regions relative to bordering regions. Section 3.5.2 discusses the estimators' vulnerability to staggered treatment adoption and introduces adjusted estimators to address this issue.

Identifying Assumption The validity of our DiD specification relies on the assumption that in the absence of any scandal, organ procurement in affected and bordering regions would have evolved in parallel. To empirically assess this "parallel trend" assumption, we complement our model with an event-study design. Specifically, for each scandal, we estimate:

$$\begin{aligned}\text{Donors}_{ht} = \alpha_h &+ \alpha_t \\ &+ \sum_{\substack{\tau=a \\ \tau\neq 0}}^{b} \beta_\tau \,\text{Affected}_h \cdot \mathbb{1}\{t = \text{Start of Scandal} + \tau\} \\ &+ \boldsymbol{X}_{p(h)t}\boldsymbol{\gamma} + \varepsilon_{ht},\end{aligned} \tag{3.2}$$

where $\mathbb{1}\{t = \text{Start of Scandal} + \tau\}$ is an indicator equal to one if period t is τ months from the onset of the respective scandal. The parameters a and b define the event window around the scandal, representing the number of months before

[18] No within-hospital staff changes were observed during the sample period.

and after its occurrence.[19] All other variables remain as previously defined. As before, we cluster standard errors at the hospital level.

Our main variables of interest are the set of coefficients $\{\beta_\tau\}_{\tau=a}^{b}$ that allows to dynamically evaluate how organ procurement in affected regions evolved relative to bordering regions.

We estimate this specification separately for each scandal. First, Figure 3.6a presents event-study estimates for the CEO scandal. The results show no significant pre-treatment differences between treated and control regions. A formal F-test on the joint significance of all pre-treatment coefficients yields a p-value of 0.19, suggesting that organ procurement trends were similar across groups prior to the CEO scandal. Second, Figure 3.6b presents the event-study estimates for the surgeon scandal. As with the CEO scandal, the figure suggests that hospitals in both treated and control regions followed similar organ procurement trends before the onset of the surgeon scandal. To formally test this, we conducted a joint F-test on the hypothesis that all pre-treatment coefficients are zero. The test returned a p-value of 0.91, indicating no evidence of pre-treatment differences between treated and control regions.

In contrast, the post-treatment coefficients confirm the graphical evidence in Figure 3.5. Following the surgeon scandal, the number of reported donors experienced a statistically significant decline, lasting approximately ten months.

Additionally, Appendix Figure 3.B.3 presents a joint event study using the CEO scandal as a reference point. The results further reinforce our findings.[20]

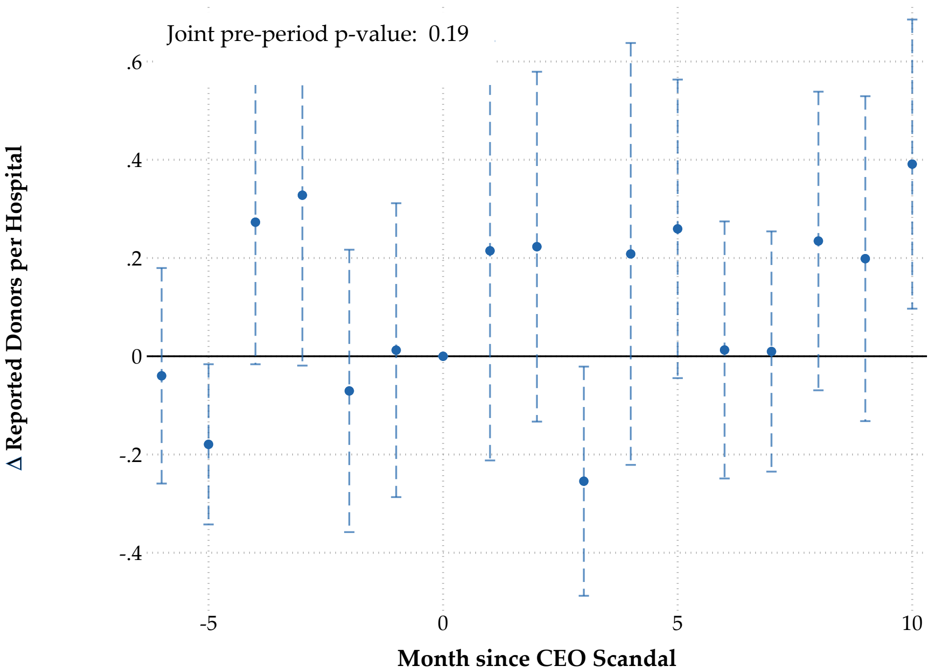
---

[19] For both scandals, we set $a = -6$ to ensure a sufficiently long pre-period for assessing parallel trends. For the CEO scandal, we choose $b = 10$ to prevent overlap with the event window of the Surgeon scandal, while for the latter, we set $b = 15$.

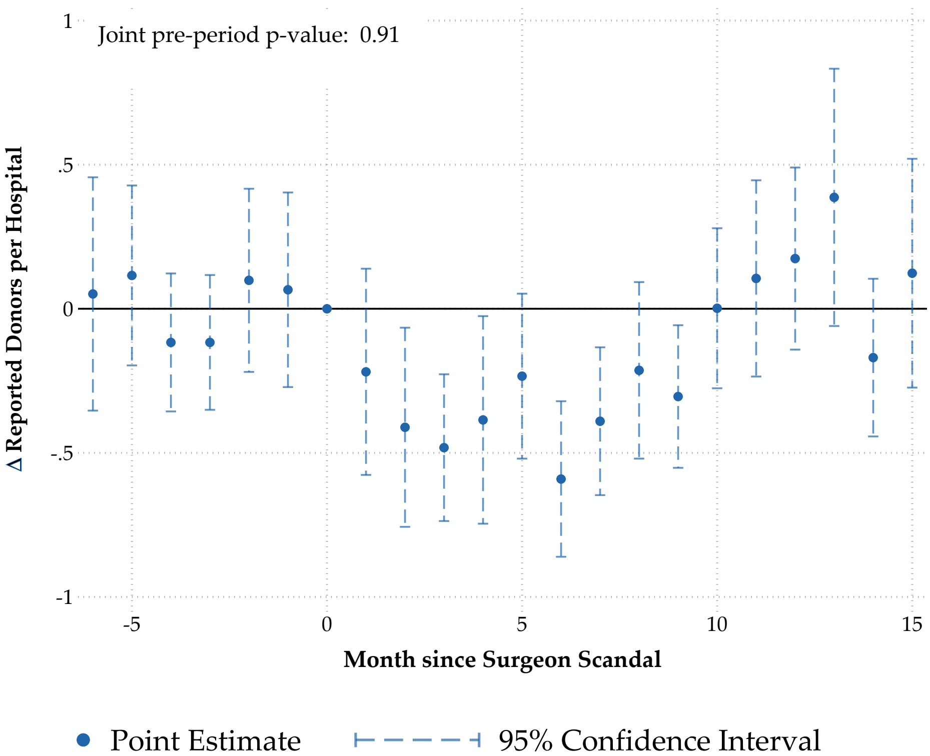
[20] The presence of two distinct treatments diverges from standard event-study frameworks, which typically assume an absorbing treatment status and do not account for multiple treatments. However, we are still likely to recover consistent estimates as our previous analysis finds no significant effect of the CEO scandal and only short-lived changes following the surgeon scandal.

Figure 3.6: Event-study Estimates

(a) CEO Scandal



(b) Surgeon Scandal



Note: The figure displays point estimates and 95 percent confidence intervals on the effect of the Surgeon scandal on the number of reported donors. All estimates are based on the regression model in Equation (3.2). Standard errors are clustered at the hospital level.

3.5 Results

In this section, we present the key findings on the relationship between media coverage of organ scandals and subsequent organ donation rates.

3.5.1 Baseline Estimates

Table 3.3 displays estimates of Equation (3.1). Columns (1)–(3) report OLS estimates, and Columns (4)–(6) report Poisson regression estimates to take into consideration the count data nature of our dependent variable. Columns (1) and (4) include population and newspaper circulation control variables. Columns (2) and (5) are augmented with hospital, month, and year fixed effects. Columns (3) and (6) also include month-hospital fixed effects. Standard errors are bootstrapped with 1,000 replications and wild clustered at the hospital level. Our primary interest lies in the interaction terms between the number of articles on each specific corruption case and the indicator variable for affected regions versus bordering regions. Panel A reports estimates for the effect of the aggregate number of articles about the two corruption scandals on the number of reported donors. OLS estimates in Columns (1)–(3) show that the effect of the total number of articles is always negative. However, the negative effect is small (between 0.002 and 0.004) and statistically non-significant. Interestingly, despite using very different sets of control variables, the magnitude of the coefficients is remarkably similar across specifications. Poisson estimates in Columns (4)–(6) yield similar results.

Panel B reports estimates considering the news on the two corruption cases separately. OLS estimates highlight at least three important results. First, different scandals display different effects on reported donors. On the one hand, the interaction coefficient capturing the effect of the number of articles on the case involving the surgeon is always negative, sizeable, and statistically significant. In affected regions, an additional article covering the specific case of corruption induces a drop in the number of reported donors by 0.015-0.018 units with respect to unaffected (bordering) regions. On the other hand, there is no effect on reported donors of the number of articles on the case involving the CEO. The interaction coefficient is close to zero and statistically non-significant. Second, results do not change with Poisson regression models in Columns (4)–(6). The estimates suggest the same negative sign for the interaction coefficient regarding the number of articles about the surgeon corruption case. The number of articles about the CEO case does not play any role in affecting reported donors. Third,

Table 3.3: Reported Donors: Difference-in-Differences Estimates

	Dependent Variable: Number of Reported Donors					
	OLS Model			Poisson Model		
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Total Number of Articles						
Total Articles	-0.002	-0.004	-0.004	-0.003	-0.006	-0.005
	(0.003)	(0.003)	(0.003)	(0.004)	(0.004)	(0.004)
Observations	4,320	4,320	4,320	4,320	4,320	4,320
Control Mean	0.546	0.546	0.546	0.546	0.546	0.546
Panel B: Disaggregated Number of Articles						
Surgeon	0.002	0.001	0.000	0.004	0.002	0.000
	(0.004)	(0.005)	(0.005)	(0.006)	(0.008)	(0.009)
Surgeon × Affected	-0.015**	-0.018**	-0.016**	-0.028**	-0.034***	-0.031***
	(0.006)	(0.007)	(0.007)	(0.012)	(0.011)	(0.010)
CEO	0.001	-0.001	-0.002	0.002	-0.001	-0.002
	(0.004)	(0.004)	(0.005)	(0.007)	(0.008)	(0.008)
CEO × Affected	0.002	0.002	0.003	0.001	0.001	0.003
	(0.006)	(0.005)	(0.006)	(0.008)	(0.007)	(0.008)
Observations	4,320	4,320	4,320	4,320	4,320	4,320
Control Mean	0.546	0.546	0.546	0.546	0.546	0.546
Controls	✓	✓	✓	✓	✓	✓
Fixed Effects						
Hospital	-	✓	✓	-	✓	✓
Month	-	✓	✓	-	✓	✓
Year	-	✓	✓	-	✓	✓
Month × Hospital	-	-	✓	-	-	✓

Note: This table shows the DiD estimates of the effect of the media coverage of the CEO and surgeon scandals on the number of reported donors. The dependent variable is the number of reported donors at the hospital-month-year level. The weighted number of newspaper articles is constructed by weighting the provincial number of newspaper articles on both scandals by the respective yearly newspapers market shares shown in Figure 3.4. Affected regions include Piedmont and Aosta Valley. Specifications in Columns (1)–(3) are estimated using OLS, and Columns (4)–(6) are estimated as Poisson regression models. In Columns (1)–(6) we include controls for population and newspapers circulation. In Columns (2) and (5), we additionally include year, month, and hospital fixed effects. In Columns (3) and (6), we additionally interact month and hospital fixed effects. Bootstrapped (1,000 replications) and wild clustered standard errors at the hospital level are in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

as suggested by the similarity of point estimates across Columns, results are insensitive to including different sets of control variables in the model.[21]

[21] Appendix 3.C reports results of a very simple introductory analysis on our data based only on affected regions, Piedmont and Aosta Valley. All the results discussed in this section are

The evidence is further bolstered by the two event studies. When considering the CEO case, Figure 3.6a demonstrates the absence of relevant effects during the post-treatment period. Conversely, for the surgeon scandal, Figure 3.6b showcases an average decrease of 0.32 in reported donors over a ten-month timeframe compared to hospitals in the control regions during the post-treatment period.

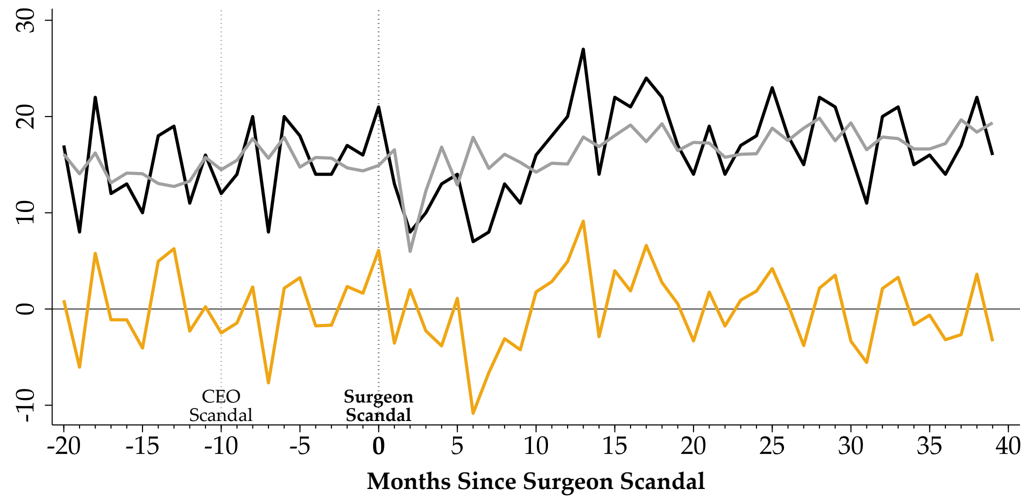
To better discuss the effect size underlying our analysis, we graphically visualize the results in Figure 3.7. Specifically, to evaluate the total loss in reported donors implied by media coverage of corruption news and the subsequent behavioral responses by the medical staff, we define a counterfactual scenario in which the scandal would not have happened. The counterfactual scenario is defined by setting estimates for the effect of the number of articles on each corruption scandal to zero.[22] In the second step, we deduct the actual number of reported donors from the predicted number of reported donors under the counterfactual scenario. Figure 3.7 plots the corresponding aggregated time series. In Figure 3.7-a we analyze the case of the CEO scandal in affected regions. In Figure 3.7-b, we analyze the case of the surgeon scandal in affected regions. The graphical analysis displays several important results. First, the CEO scandal (first vertical line) does not imply reported donors' responses. This is true for both affected regions (Figure 3.7-a) and, as expected, for bordering regions (Appendix Figure 3.B.2-a). On the contrary, considering Figure 3.7-b, the second vertical line corresponding to the surgeon scandal anticipates a sizeable drop in reported donors in affected regions. The drop lasts 5 to 10 months from the onset of the surgeon's case of corruption. The yellow-shaded area makes up the aggregate loss in reported donors implied by the corruption scandal involving the surgeon. Compared to the counterfactual scenario with no scandal, the medical staff reported roughly 50 fewer donors (about 22 in the first five months). This number corresponds to about 35 percent of the number of reported donors that Piedmont and Aosta Valley registered in the same 10-month time period right before the outbreak of the scandal. Bordering regions do not display any particular effect following the outbreak of the surgeon scandal (Appendix Figure 3.B.2-b).

substantially confirmed. Remarkably, the magnitude of the effect is similar. OLS coefficients imply a drop of 0.015-0.016 in the number of reported donors for one more article on the surgeon case. In contrast, the coefficient for the CEO case is never statistically significant. Table 3.A.4 re-estimates baseline models using TV news to measure perceived corruption. Results are insensitive to the source of news.

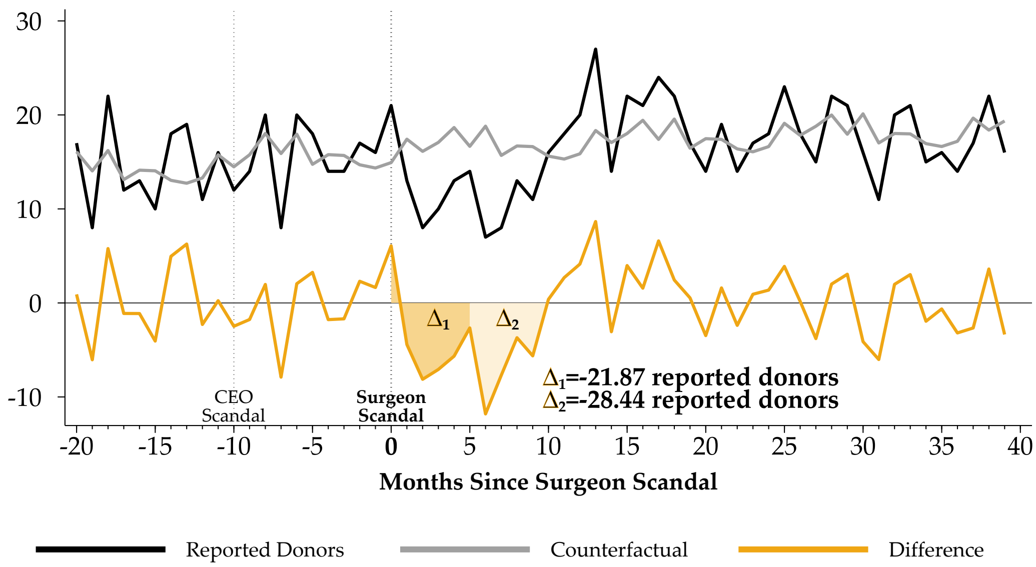
[22] For instance, to simulate the counterfactual scenario in the absence of the surgeon case, we set the coefficients β_2 and β_4 in Equation (3.1) to zero.

Figure 3.7: Counterfactual Effects by Time: Affected Regions

(a) Affected Regions: CEO Scandal



(b) Affected Regions: Surgeon Scandal



Note: The figure depicts actual and counterfactual reported donors across time for the case of affected regions. The counterfactual is constructed by predicting reported donors using the estimates from Table 3.3, Column (2) and by setting to zero the coefficient for the number of articles on the surgeon case. The time series on reported donors results from averaging the hospital level data in the affected regions by year-month and subsequently centering it around the onset of the surgeon scandal.

Overall, these results support the interpretation of the scandal as an exogenous shock affecting only those hospitals operating within the administrative boundaries of the two regions directly involved in the scandals and rule out the possibility of a common – e.g., at the national level – perception shock influencing the number of reported donors in the two analyzed regions.

3.5.2 Robustness Checks

We now conduct several robustness checks to support our main findings.

Distance to Border We test our estimates' robustness to different sample restrictions. Up to now, all the hospitals in affected and bordering regions have been included in the analysis. As a first robustness test, we restrict the sample used in the DiD analysis to "neighboring" hospitals located on each side of the regional borders separating affected and unaffected regions. This test sheds additional light on the spatial dimension of the effects of corruption news. Table 3.4 reports the estimates of Equation (3.1) with the sample restricted to those hospitals within a certain distance from the border. Column (1) is based on a 150 km cutoff and replicates the full sample of the baseline analysis, making clear the relatively small size of the geographical area under study. Column (2) restricts the sample to hospitals within 120 km from the border, Column (3) restricts this distance to 90 km, and Column (4) further restricts the distance to 60 km.[23] The analysis confirms that, in affected regions, CEO-related articles have no impact on reported donor numbers. Conversely, results reveal that the point estimate for the impact of the surgeon corruption scandal remains similar in magnitude and significance when shrinking the sample only to include hospitals close to the border. The drop in sample size in Column (4) makes the coefficient smaller and less precise. Overall, the stability of the coefficients suggests that the administrative border of the treated regions bounds the response of medical staff. This likely reflects the institutional setup of Italy's NHS: with the health care systems managed by regional governments, corruption news about cases that occurred in a specific region are only informative about that specific region and do not spill over to other regions or jurisdictions.

Scandals Epicenter Table 3.5 provides a second test for the spatial dimension of our findings. Specifically, we verify whether the effect implied by corruption scandals is mostly driven by the hospital directly hit by the scandal (Hospital 'Molinette') or it is also visible for other hospitals operating as spokes of the same regional coordination center. The Hospital 'Molinette' is the largest center for organ donations in the Piedmont and Aosta Valley area, with a monthly average of reported donors of 2.7 units; it also hosts the regional coordination center. The

[23] Sample size gets considerably smaller for further restrictions in terms of the distance to the border.

Table 3.4: Estimates by Distance to Border

	Dependent Variable: Number of Reported Donors			
	(1)	(2)	(3)	(4)
Surgeon	0.001	0.001	0.001	0.003
	(0.005)	(0.005)	(0.005)	(0.006)
Surgeon × Affected	-0.018**	-0.017***	-0.018***	-0.009
	(0.007)	(0.007)	(0.007)	(0.019)
CEO	-0.001	-0.001	-0.001	-0.005
	(0.004)	(0.004)	(0.005)	(0.006)
CEO × Affected	0.002	0.002	0.002	-0.008
	(0.005)	(0.005)	(0.005)	(0.009)
Observations	4,320	4,140	4,020	2,880
Distance to border in km	< 150	< 120	< 90	< 60
Number of Hospitals				
Treatment	22	22	21	12
Control	50	47	46	36
Controls	✓	✓	✓	✓
Fixed Effects				
Hospital	✓	✓	✓	✓
Month	✓	✓	✓	✓
Year	✓	✓	✓	✓

Note: This table replicates DiD estimates by restricting the sample to only include hospitals from affected (bordering) regions within a certain range to the closest border of a bordering (affected) region. The dependent variable is the number of reported donors at the hospital-month-year level. The weighted number of newspaper articles is constructed by weighting the provincial number of newspaper articles on both scandals by the respective yearly newspapers market shares shown in Figure 3.4. Distance between each hospital and the border is measured as the linear distance to the closest border of a bordering (affected) region. All specifications are estimated using OLS. Specifications in Columns (1)–(3) are estimated using OLS, and Columns (4)–(6) are estimated as Poisson regression models. In Columns (1)–(6) we include controls for population and newspapers circulation. In Columns (2) and (5), we additionally include year, month, and hospital fixed effects. In Columns (3) and (6), we additionally interact month and hospital fixed effects. Bootstrapped (1,000 replications) and wild clustered standard errors at the hospital level are in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

monthly average for the other hospital centers is around 0.64 units. Columns (1) and (2) of Table 3.5 replicate the analysis by excluding either the Hospital 'Molinette' or all the hospitals in the city of Turin, the city where the Hospital 'Molinette' is located and the regional capital city in Piedmont. Complementing this exercise, Column (3) sheds light on the dynamics at Hospital 'Molinette', by excluding all other hospitals in affected regions. The estimates confirm that our results are not solely driven by the involvement of some of the medical staff at Hospital 'Molinette', though the strongest effect is still observed at Hospital 'Molinette'. A similar picture emerges in Columns (4) and (5) that apply an inverse-distance weighting approach, i.e., assigning more weight to hospitals closer to the scandal epicenter. Weights are constructed as $w_i^* = \frac{w_i}{\sum_{i=1}^{N} w_i}$ where $w_i = \frac{1}{d(h^*,h_i)^\alpha}$. Here $d(h^*, h_i)^\alpha$ represents the linear distance of hospital i, located at coordinates h_i, to the hospital hit by the scandals, located at h^*. Column (4) weights observations by a power parameter of $\alpha = 1$, while Column (5) assigns more weight to hospitals closer to the scandal hospital by taking $\alpha = 2$. All the results remain similar to those of the baseline analysis. If anything, the use of weights slightly decreases coefficient magnitude.

Time Window around Corruption Scandals The time window around the corruption scandals selected for the analysis might also shape and confound the empirical results. In principle, we should expect the reaction by the medical staff to occur in the proximity of corruption scandals. Accordingly, we test how the effect size for the impact of corruption news changes when different time windows are analyzed. Figure 3.8 reports the analysis for four different time windows around each case, namely 12, 9, 6, and 3 months.[24] All the four time windows are symmetrical to the onset of each case of corruption. The figure reports the results for the Poisson regression model (Table 3.3, Column (2)) and shows the coefficients for the number of articles on the two corruption cases, surgeon and CEO, with the respective 95 percent confidence intervals.[25] The analysis highlights a clear pattern. The point estimates obtained through the four sub-samples are similar in size to those obtained in the baseline estimates for the whole 2001-2005 period (reported in the left part of the figure). The similarity holds for both the CEO scandal and the surgeon scandal. The fact that also in the 3-month window, the effect size, although slightly less precisely estimated, is similar to that for larger time windows suggests an immediate and possibly emotional response by medical staff to the onset of the scandal.

Staggered and Multiple Treatment Considerations This section discusses potential vulnerabilities of the two-way fixed effects estimator presented in Equation (3.1) to heterogeneous effects. Importantly, our empirical setting does not involve staggered treatment adoption, a concern raised in recent econometric literature (see, for example, Goodman-Bacon (2021) and de Chaisemartin and D'Haultfœuille (2020)). Specifically, Equation (3.1) does not involve staggered

[24] This approach involves a trade-off, as reduced sample size may affect estimate precision.
[25] Confidence intervals are constructed through bootstrapped (1,000 replications) and wild clustered standard errors at the hospital level.

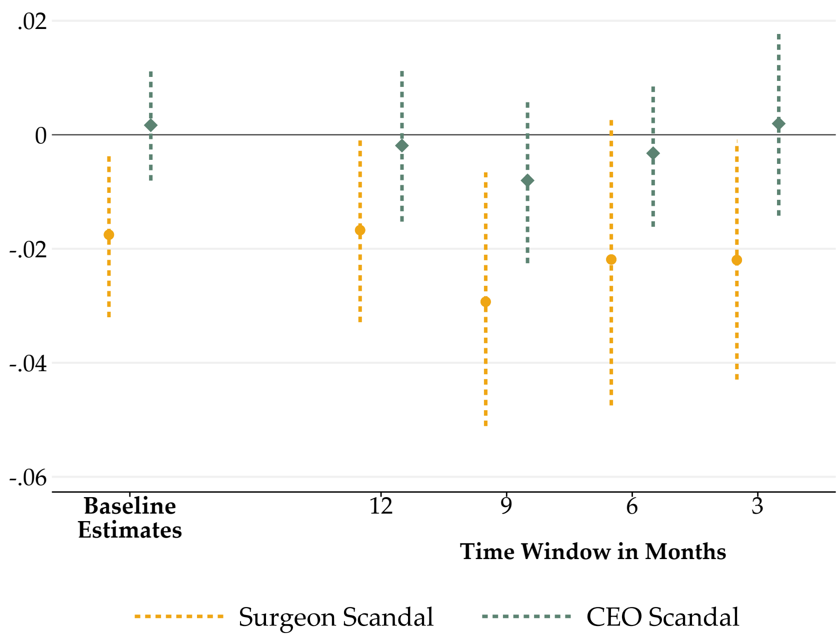
Table 3.5: Sensitivity Analysis with respect to Scandal Epicenter

	Dependent Variable: Number of Reported Donors				
	OLS Model			Weighted OLS Model	
	(1)	(2)	(3)	(4)	(5)
Surgeon	0.001	0.002	-0.001	-0.003	-0.006
	(0.005)	(0.005)	(0.006)	(0.007)	(0.008)
Surgeon × Affected	-0.016***	-0.018**	-0.054***	-0.016**	-0.013*
	(0.006)	(0.007)	(0.005)	(0.008)	(0.007)
CEO	-0.000	-0.000	-0.002	0.001	0.000
	(0.004)	(0.004)	(0.004)	(0.012)	(0.017)
CEO × Affected	0.002	0.000	-0.005	0.009	0.014
	(0.005)	(0.006)	(0.176)	(0.008)	(0.011)
Observations	4,260	4,020	3,060	4,260	4,260
Controls	✓	✓	✓	✓	
Fixed Effects					
Hospital	✓	✓	✓	✓	✓
Month	✓	✓	✓	✓	✓
Year	✓	✓	✓	✓	✓
Treatment Exclusion Set					
Molinette Hospital	✓	✓	-	✓	✓
Torino Hospitals	-	✓	✓	-	-
Other Hospitals	-	-	✓	-	-
Weighting Parameters					
ID-Weighting	-	-	-	✓	✓
Weighting Exponent	-	-	-	1	2

Note: This table replicates the DiD analysis by considering the possible attenuation of the effects of corruption for hospitals distant from the epicenter of the scandals. The dependent variable is the number of reported donors at the hospital-month-year level. The weighted number of newspaper articles is constructed by weighting the provincial number of newspaper articles on both scandals by the yearly newspaper market shares shown in Figure 3.4. Column (1) excludes the scandal epicenter (Hospital 'Molinette') from the sample. Column (2) additionally excludes the remaining four hospitals located in Turin. Column (3) excludes all treatment hospitals except the scandal epicenter (Hospital 'Molinette'). Columns (4) and (5) apply, in addition to excluding the scandal epicenter, weighted linear probability models by assigning hospitals closer to the scandal epicenter inverse-distance weights. Weights are calculated as $w_i^* = \frac{w_i}{\sum_{i=1}^{N} w_i}$, where $w_i = \frac{1}{d(h^*, h_i)^\alpha}$. $d(\cdot,\cdot)^\alpha$ represents the linear distance of hospital i, located at h_i, to the scandal epicenter, located at h^*. The exponent α represents the power parameter to assign more weight to hospitals closer to the scandal epicenter. In Column (3) a power parameter $\alpha = 1$ is assumed, and in Column (4) a power parameter $\alpha = 2$ is assumed. All specifications include population and newspaper circulation controls and year, month, and hospital fixed effects. Bootstrapped (1,000 replications) and wild clustered standard errors at the hospital level are in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

treatment adoption across units but rather incorporates multiple treatments within units. By explicitly modeling both scandals and including a substantial number of never-treated control hospitals, the two-way fixed effects in Equation (3.1) avoids leveraging 'forbidden comparisons' (Borusyak, Jaravel and Spiess, 2021) that might lead to inconsistent estimators in the case of staggered

Figure 3.8: Time Windows Around the Corruption Scandal



Note: The figure shows point estimates and 95 percent confidence intervals for the effect of the number of newspaper articles about the corruption scandals on the number of reported donors. All estimates coefficients and standard errors are based on the regression model in Table 3.3, Column (2). Coefficients are obtained by using four different symmetrical time windows (12-,9-, 6-, and 3-months) around the onset of the corruption scandal.

treatment adoption. Our results, therefore, maintain robustness under these considerations.

However, as discussed in de Chaisemartin and D'Haultfœuille (2023), contamination bias can also occur in settings with multiple treatments within units when treatment effects vary across units. In our setting, estimates based on specification 3.1 may, therefore, be contaminated if health care workers' response to the surgeon scandal depends on their exposure to the CEO scandal. Following de Chaisemartin and D'Haultfœuille (2023), Table 3.6 provides evidence that the effect of the surgeon scandal is not driven by prior exposure to the CEO scandal. For reference, Column (1) reproduces our main estimates corresponding to Column (2) in Panel B of Table 3.3. In Column (2), we interact the number of articles on the surgeon scandal with the average coverage of the CEO scandal. To this end, we first construct a measure to proxy the province-specific exposure to the CEO scandal. In particular, for each province p, we compute the average number of articles on the CEO scandal in the 20 months after the CEO scandal, $\overline{\text{Art. CEO}}_p$. Next, we interact this measure with Art. Surgeon_{pt} and re-estimate

Table 3.6: Sensitivity Analysis of Surgeon Coefficients

	Dependent Variable: Number of Reported Donors			
	Baseline	Interaction	Multiplicative	Cumulative
	(1)	(2)	(3)	(4)
Surgeon	0.001	0.012	0.001	0.021
	(0.005)	(0.013)	(0.006)	(0.108)
Surgeon × Affected	-0.020***	-0.037**	-0.010*	-0.044
	(0.007)	(0.015)	(0.005)	(0.100)
CEO	-0.001	-0.000	-0.001	0.001
	(0.005)	(0.004)	(0.005)	(0.004)
CEO × Affected	0.002	0.001	0.001	0.001
	(0.005)	(0.005)	(0.005)	(0.005)
Surgeon × CEO	-	-0.005	-	-
	-	(0.005)	-	-
Surgeon × CEO × Affected	-	0.007	-	-
	-	(0.005)	-	-
Observations	4,320	4,320	4,320	4,320
Test for Equality of Surgeon Coefficients				
Chi-Square Statistic	-	-	10.644	0.061
Chi-Square p-value	-	-	0.001	0.805
Controls	✓	✓	✓	✓
Fixed Effects				
Hospital	✓	✓	✓	✓
Month	✓	✓	✓	✓
Year	✓	✓	✓	✓

Note: This table evaluates the sensitivity of our baseline estimates when adjusting for multiple treatments. For comparison Column (1) replicates the baseline estimates. In Column (2) additionally controls for the interaction of surgeon articles and CEO articles as proposed in de Chaisemartin and D'Haultfœuille (2023). In Column (3) Surgeon articles are multiplied by the total number of CEO articles published in the month before the occurrence of the Surgeon scandal. Column (4) considers the cumulative article count. All specifications include population and newspaper circulation controls and year, month, and hospital fixed effects. Bootstrapped (1,000 replications) and wild clustered standard errors at the hospital level are in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

our main regression model. If part of the effect is indeed driven by prior exposure to the CEO scandal, then the coefficient of the interacted measure should increase with respect to our main result. However, as shown in Column (2), the effect gets smaller, suggesting that provinces with less exposure to the CEO scandal drive the effect. Similarly, in Column (3), we add the interaction of the two treatments to the regression. Reassuringly, the estimated coefficients are insignificant, while the effect for the Surgeon scandal increases in magnitude, implying that if anything, our baseline regression estimate is biased towards zero. Column (4) further supports this finding by considering the cumulative number of articles on the surgeon and CEO scandals.

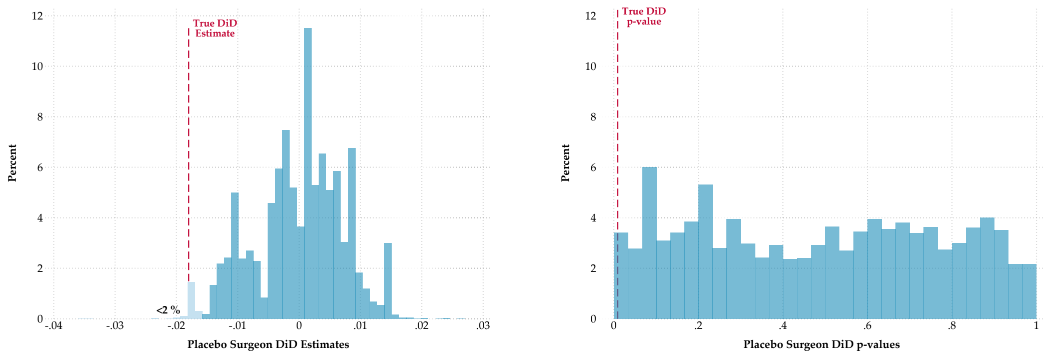
Placebo Estimates As a final robustness test, we perform a placebo test in which we randomly shift the timing of the two corruption scandals. We construct placebo time series for the number of Surgeon and CEO articles using a two-step procedure. First, we shift each regional time series on the surgeon's case back or forward. Second, given a particular shift of the surgeon scandal articles, we then compute all feasible back- and forward-shifts for the CEO article series, conditional on each Surgeon shift. We repeat this exercise for all possible back- and forward shifts of the surgeon time series.[26] In total, we perform 6,012 placebo regressions based on the regression model in Table 3.3, Column (2).

Figure 3.9a reports the distribution of the DiD point estimates on the surgeon case. The placebo estimates are normally distributed and centered around zero. Reassuringly, the estimate we document is located far in the lower tail, suggesting it is unlikely to result from random variation. Figure 3.9b shows the corresponding distribution of the placebo p-values. As expected under random assignment, the estimated p-values are (approximately) uniformly distributed.

[26] The number of possible shifts is limited by the number of times we observe each hospital, i.e., 60 months. Further, we omit shift combinations in which the DiD coefficient on the surgeon or CEO is not identified due to multicollinearity.

Figure 3.9: Placebo Difference-in-Differences Estimates

(a) Placebo Coefficients (b) Placebo p-values



Note: The figure shows the distribution of placebo point estimates and p-values for the effect of the number of newspaper articles about the surgeon scandal on the number of reported donors. Placebo time series on the number of surgeon and CEO articles result from a two-step procedure. First, each regional time series on the surgeon case is shifted back or forward. Second, given a particular shift of the surgeon scandal articles, all possible back- and forward-shifts for the regional CEO article time series are computed. The exercise is repeated for all possible back- and forward-shifts of the surgeon time series. The figure represents 6,012 placebo time series based on the regression model in Table 3.3, Column (2). The red dashed line highlights the true point estimate and p-value as in Table 3.3, Column (2).

3.6 Discussion and Mechanisms

The aim of this section is twofold: first, we explore the potential impact of corruption news on the behavior of other actors involved in the procurement process, looking in particular at oppositions; second, we sketch a simple conceptual framework to discuss the mechanisms behind our results.

3.6.1 Oppositions to Organ Donation

So far, we have documented medical staff behavioral responses: our analysis shows a decline in the number of *reported* donors induced by media coverage of corruption scandals. However, other actors in the organ donation process are affected by corruption news. Whenever a reported donor is identified, the medical authority must verify whether the deceased patient registered her consent to organ donation. Without an expressed will, it is up to patients' relatives to consent or deny organ donation. While we can exclude that the decision of the deceased patient is affected by the corruption scandals, it is plausible to think of possible reactions by family members. To test this hypothesis, we investigate the effect of media coverage of corruption scandals on *oppositions* to donation and on the number of *actual* donors, namely the number of reported donors net of oppositions (see Figure 3.1).

Panel A of Table 3.7 reports the analysis of the impact of media coverage of corruption scandals on the number of oppositions based on the DiD setting in Equation (3.1). For the sake of brevity, we only report results obtained through OLS specifications. Given the impact of media coverage of corruption scandals on reported donors, which is the first step in the organ donation process, we propose alternative ways to isolate the effect of interest. In Column (1), we estimate Equation (3.1) by considering the number of oppositions as the new outcome variable. These estimates are likely biased by the endogeneity underlying the number of reported donors. As a first attempt to deal with endogeneity concerns, Column (2) replicates the analysis by including the actual number of reported donors as control variables. In Column (3), instead of using the observed number of reported donors, we augment the specification with the number of reported donors as predicted by the model in Column (2) of Table 3.3. To further mitigate endogeneity concerns, Columns (4) and (5) focus on different definitions of the number of oppositions to capture behavioral responses by reported donors and their families to media coverage of corruption scandals. The specification in Column (4), instead of using the raw number of oppositions, focuses on the three-month moving average. Column (5) considers the share of opposition over the number of reported donors as an outcome variable. Both specifications in Columns (4) and (5) include reported donors as control variables. The analysis of oppositions in Table 3.7-A displays three main results. First, the CEO case does not show any impact on opposition independently of the estimated specification. Post-CEO scandal, affected and bordering regions perform similarly: as for the medical staff, patients and patients' relatives are not affected by hospital management scandals. Second, the surgeon's case does not shape behavioral responses by reported donors and their families. Indeed, the coefficient for the interaction term between the number of articles on the surgeon case and the indicator for affected regions is always close to zero and statistically non-significant. Third, the point estimates are remarkably similar across different empirical specifications. The stability suggests that results on oppositions are robust to different possible ways to empirically deal with the endogeneity underlying the number of reported donors.

The analysis of oppositions provides a comprehensive overview of the impact of media coverage of corruption scandals on organ donation. Furthermore, it clarifies which actors in the process are most heavily affected by the scandals.

While medical staff behavior responds to the scandals, non-medical opposition to organ donation seems unaffected. The latter result likely derives from two different reasons. On the one hand, corruption news has no impact in the presence of previously registered consent to a donation by the reported donor. Starting from 1999, each Italian citizen can express her consent to a donation in different forms, e.g., by explicitly mentioning it in the individual Identity Card. Of course, relatives cannot change this individual choice. On the other hand, missing this explicit choice by the donor, the choice of and attitude towards organ donation by relatives constitute an informed choice carefully pondered over time and solidly based on moral values that are unlikely to be affected by the negative (and likely transitory) emotions caused by corruption scandals like those analyzed in this study.

We complete the analysis by considering the number of actual donors, namely the number of reported donors net of oppositions. Table 3.7-B replicates the analysis in Panel A by considering the number of actual donors as the outcome of interest. Results confirm that media coverage of corruption scandals mainly shapes behavioral responses by the medical staff, while the effect on patients and their relatives is negligible. In Column (1), the media coverage of the surgeon's case is found to decrease the number of actual donors in affected regions relative to unaffected regions. Each additional article on the surgeon case reduces actual donors by 0.013 units. Columns (2)–(5) deal with the endogeneity of reported donors and confirm that the negative effect is only driven by medical staff behavioral responses to corruption scandals. Indeed, in all specifications taking into account the impact of media coverage on medical staff behavior, the effect of the surgeon case turns to a precisely-estimated zero. Once the response of medical staff behavior is taken into account, the impact of media reporting of corruption scandals on the number of actual donors disappears.

3.6.2 Text Analysis

Alternatively, the differential response of medical staff might be influenced by varied reporting on the two corruption scandals. To explore this possibility, we conduct a text analysis on all newspaper articles covering the scandals. Articles are gathered by searching major Italian newspaper archives for either the CEO or Surgeon scandal, resulting in a total sample of 247 articles: 143 covering the CEO scandal, 83 covering the Surgeon scandal, and 20 covering both scandals.

The analysis proceeds in two steps. First, we evaluate semantic similarity of CEO and Surgeon articles. In particular, we evaluate if the tone of articles differs between the two cases or if articles use different language to describe the scandals. We measure semantic similarity through three measures. Subjectivity measures the degree to which a piece of text expresses personal opinions, feelings, or judgments, rather than factual information. It ranges from 0 to 1, where 0 indicates an objective, factual statement and 1 indicates a highly subjective, opinionated statement. Polarity is a measure of the sentiment expressed in a piece of text. It ranges from -1 to 1, where negative values indicate negative sentiment and positive values indicate positive sentiment. Lastly, we provide a measure of language similarity. To this end, we represent each article as a word embedding. Embeddings are vector representations of a text body in continuous space. Articles with similar word embeddings are also likely to use similar language. To test for differences in embeddings between CEO and surgeon articles we first retrieve the word embedding of each article using a pre-trained language model[27]. Next, we, extract the first principal component across all article embeddings. We standardize all three measures to mean zero and standard deviation one, such that a one unit increase corresponds to a one standard deviation increase of the respective measure.

To test for statistical differences along these measures, we estimate the following regression equation

$$Y_{it} = \alpha_{n(i)} + \alpha_t + \beta \text{Surgeon}_i + \varepsilon_{it} \tag{3.3}$$

where Y_i indicates some text metric of article i published at time t. $Surgeon_i$ is an indicator equaling one if the article covers the Surgeon scandal. Using $\alpha_{n(i)}$ and α_t, we account for unobserved newspaper-specific and time-specific effects.

Estimates in Columns (1)–(3) of Table 3.8 show no significant differences between the articles on either dimension. Across all three measures, we document a small and statistically insignificant effect, indicating that Surgeon and CEO articles use similar semantics and language. Next, moving beyond semantics, we aim to analyze if the content structure of articles differs by scandal. The articles naturally differ in the type of scandal reported; however, we investigate whether the scandals are also framed or contextualized differently. For example, one scandal might be more focused on individual misconduct, while the other focuses

[27] In particular, we use the 'paraphrase-multilingual-MiniLM-L12-v2' language model, which paraphrases multilingual sentences and paragraphs as a 384-dimensional dense vector space.

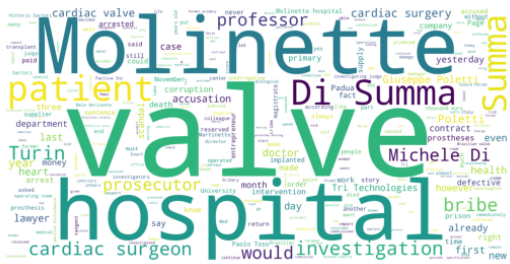
on the institutional implications of the scandal. In a first step, we display the most frequently used words in the newspaper articles in Figure 3.10. The size of each word is proportional to its relative frequency within the articles. Panels (b) and (c) depict wordclouds separately for the CEO and Surgeon scandal. Both scandals took place at the Molinette hospital in Turin, which prominently features in all wordclouds. Unsurprisingly, the CEO articles mention the hospital manager Luigi Odasso who took bribes to manipulate organ waiting lists. The surgeon articles center around surgeon Michel Di Summa who accepted money to use defective heart valves. Besides, there is no evidence that articles are contextualized differently. To analyze the content structure in more detail, we proceed by training

Figure 3.10: Wordcloud of Newspaper Articles

(a) All Articles

(b) CEO Articles

(c) Surgeon Articles



Note: The figure displays wordclouds depicting the most frequently used words in the newspaper articles discussed in Section 3.6.2. The articles have been translated into English, and common stopwords have been excluded. The size of each word is proportional to its relative frequency within the articles.

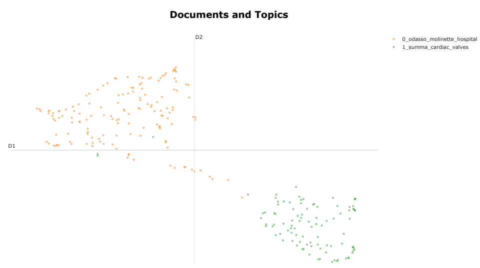
a topic model on the article corpus. A topic model is a statistical model designed to discover abstract topics within a collection of documents or texts. It is employed in natural language processing and machine learning to identify the underlying themes or topics prevalent in a set of documents. The goal is to automatically extract meaningful patterns and associations among words for categorizing and understanding the content of text documents. Intuitively, a topic model algorithm computes a word embedding for each article and then clusters articles close in vector space.

We train a topic model[28] on all newspaper articles covering either the CEO or Surgeon scandal. After training, we extract a topic distribution for each article along the identified topics. To compare the topic distribution of articles on the CEO scandal, we extract the first two principal components of the topic distribution and use them as outcome variables in Equation (3.3). The topic model identifies two topics, represented by the keywords “odasso, molinette, hospital” and “summa, cardiac, valves”. We visualize the topic distribution in Figure 3.11 (a), with each dot representing an article along the first two principal components of the topic distribution. The color of the dots indicates attribution to one of the two identified topics. The topics are clearly separated in space with little to no overlap. Accordingly, as shown in Column (4) of Table 3.8, we observe a strong statistical difference between CEO and surgeon articles along the first principal component of the topic distribution. However, we do not find support for differences along the second principal component, as shown in Column (5) of Table 3.8. In this regard, the model yields no further distinction or contextualization beyond the categorization into CEO and surgeon scandals. The reason the model only retrieves two topics could be that the default hyperparameter selection may yield a global topic perspective, ignoring subtle differences within topics. Therefore, we allow for a less global view by decreasing the ’number of neighboring sample points’ parameter of the topic model, used during the manifold approximation process. A higher value of this parameter tends to yield a more comprehensive representation of the embedding structure, offering a global perspective. Conversely, a smaller value provides a more localized view. However, as shown in panels (b) to (c) of Figure 3.11, even when reducing the parameter, the model only recovers two topics and a similar topic distribution. Only after further decreasing the parameter in panel (d) does the model identify five additional topics. However, the newly added topics are rather unstructured subsets of the previously identified topics. Also, the newly identified topics are represented by very similar keywords that cannot be interpreted as different contextualizations of either scandal. In summary, the text analysis provides no evidence that the reporting of the scandals differed along either semantic or thematic dimensions – besides the categorization into CEO and Surgeon scandal. We cautiously interpret this as evidence that the effect is driven by the shared professional mission of medical staff and surgeons.

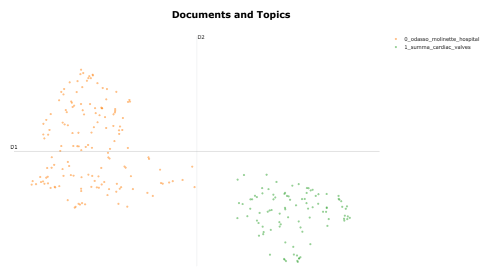
[28] We use the ’BERTopic’ Python module with default settings.

Figure 3.11: Identified Topics and Topic Distribution

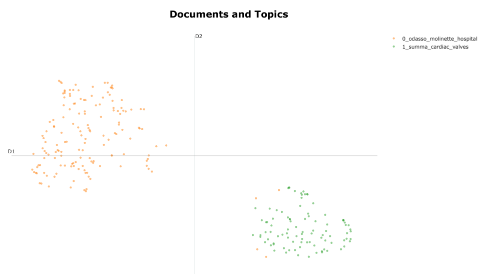
(a) Neighboring Sample Points (=15)



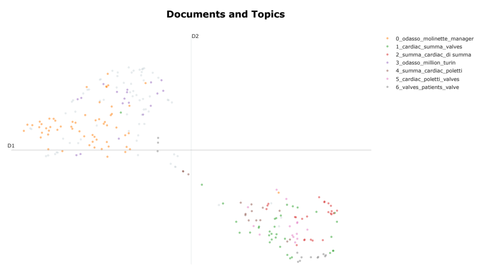
(b) Neighboring Sample Points (=10)



(c) Neighboring Sample Points (=5)



(d) Neighboring Sample Points (=2)



Note: This figure shows topic distributions across newspaper articles, employing various topic model parametrizations. Each dot represents an article with colors indicating attribution to topics along the first two principal components of the topic distribution. The topic distributions results from first training the topic model on the article text corpus and then computing article-specific distribution across topics. The graphs in panels (a)–(d) differ by the number of neighboring sample points, which internally is used during the manifold approximation process. A higher value tends to yield a more comprehensive representation of the embedding structure, offering a global perspective. Conversely, smaller values provide a more localized view. Typically, elevating neighbors leads to the formation of larger clusters in the resulting embedding.

3.6.3 Conceptual Framework

Our empirical findings suggest a change in the behavior of the medical staff (as a result of media reporting of corruption scandals) as the key driver for the drop in the number of reported donors. We now provide a conceptual framework to interpret medical staff's behavioral response to corruption-related news.

We build upon the work of Kolstad (2013) and Cornelissen, Dustmann and Schönberg (2017). We consider an individual model of effort choice by the medical staff working in ICUs and account for the institutional characteristics of the Italian NHS. We characterize health care workers as both extrinsically and intrinsically motivated. All workers in public hospitals are public employees subject to fixed-wage contracts in the Italian NHS, including physicians and nurses.[29] Total benefit B for medical staff depends on the fixed salary w and an intrinsic reward Γ. A higher salary and a higher intrinsic reward improve benefit B. We assume intrinsic reward to be related to the number of reported donors D, which are a direct function of individual effort e. In turn, individual effort is influenced by perceived corruption Ω, both directly and indirectly, via *expected* oppositions O. More effort improves the intrinsic utility since it increases the number of reported donors. However, perceived corruption reduces e directly and also increases expected oppositions. This second effect is driven by the *anticipation* of a potential change in the behavior of family members as a result of a change in perceived corruption. Total cost C depends on individual effort e and peer pressure p. We define peer pressure as the difference between the *expected* effort of peers and individual effort. We consider expected effort since each worker does not observe colleagues' efforts, but she has noisy beliefs about colleagues' efforts within the hospital. The higher the expected effort of peers, the higher the cost to keep up with peers. Individual i's net utility in each period t can then be represented as follows:

$$U_{it} = B_{it}(w_{it}, \Gamma_{it}[D_{it}(e_{it}(\Omega_{it}, \mathbb{E}(O_{it}|\Omega_{it})))]) - C_{it}(e_{it}, p(\mathbb{E}(e_{-it}|\Omega_{it}) - e_{it})) \quad (3.4)$$

For each worker i in each period t, the optimal effort e^*_{it} stems from equating marginal benefit and cost. To ensure concavity of the problem, we assume individual effort: (i) to increase intrinsic motivation Γ at a decreasing rate; (ii) to directly increase the cost at an increasing rate; and (iii) to reduce peer pressure p at a decreasing rate. Optimal effort e^*_{it} is implicitly defined by the following equation:

$$\frac{\partial B_{it}}{\partial \Gamma_{it}} \frac{\partial \Gamma_{it}}{\partial D_{it}} \frac{\partial D_{it}}{\partial e_{it}} = \frac{\partial C_{it}}{\partial e_{it}} + \frac{\partial C_{it}}{\partial p_{it}} \frac{\partial p_{it}}{\partial e_{it}} \quad . \quad (3.5)$$

For each individual, media reporting about corruption episodes in each period t shifts the level of individual perceived corruption in the public health care sector Ω. More articles published by newspapers on corruption episodes increase Ω or, in other words, increase perceived corruption. This, in turn, implies an update of beliefs about the effort of peers: a higher perceived corruption within the health care sector reduces the expected effort of peers.

Thus, perceived corruption Ω influences optimal individual effort e^*_{it} via three channels. The first channel is the reduction of intrinsic motivation, which un-

[29] The current rules also allow for merit-related pay based on results measured by many performance indicators, which are chosen and differ across hospitals. However, the performance-related component is minimal compared to the fixed salary that considering a fixed-wage contract is a reasonable assumption.

dermines intrinsic reward: $\frac{\partial \Gamma_{it}}{\partial D_{it}} \frac{\partial D_{it}}{\partial e_{it}} \frac{\partial e_{it}}{\partial \Omega_{it}} < 0$. It is a direct effect stemming from the change in perceived corruption driven by media reports. The second channel is an indirect effect of media reporting via the impact on expected oppositions. Medical staff anticipates an increase in oppositions by families, induced by a lower level of trust induced by corruption news, hence decreasing effort in response: $\frac{\partial \Gamma_{it}}{\partial D_{it}} \frac{\partial D_{it}}{\partial e_{it}} \frac{\partial e_{it}}{\partial \mathbb{E}(O)_{it}} \frac{\partial \mathbb{E}(O)_{it}}{\partial \Omega_{it}} < 0$.[30] The third channel works on the cost side: since increases in perceived corruption update information on the effort of other workers, it also reduces the cost of peer pressure: $\frac{\partial \mathbb{E}(e_{-it})}{\partial \Omega_{it}} < 0$. Information about corruption in the health care sector makes staff aware that people working within the system are exercising 'bad' effort to gain rents instead of 'good' effort to use resources efficiently and effectively.

This framework helps rationalize our results. For instance, it makes clear that the articles published by newspapers to report on corruption within the health care sector would produce different effects on both Γ and $\mathbb{E}(e_{-it})$ depending on their informative content for medical staff working in ICU. This means that media reporting of a corrupt surgeon is likely more informative than news about a corrupt hospital manager, as our empirical findings show.

Individual effort is crucial, together with idiosyncratic individual ability a, to keep brain-dead patients – still not clinically dead – in the condition necessary for their organs to be transplanted and to identify potential donors D to be reported to the reference Organ Procurement Center. Clearly, higher ability a and effort e^* increase D. The number of reported donors D in the hospital h in each period t can then be written as

$$D_{ht} = A_h + E^*_{ht} + \epsilon_{ht}, \tag{3.6}$$

where A and E^* are the average ability, and the average optimal effort of ICU staff in each hospital h and ϵ reflects output variability beyond workers' control.

Consider now the impact on D of media reporting on corruption scandals in period t. With an increase in the coverage of corruption scandals, Ω will increase, therefore reducing both the marginal benefit of effort (via intrinsic motivation and expected oppositions) and the marginal cost (via the information on the effort of peers). Then, the optimal e^* at the individual level can either decrease or increase, depending on the relative strength of these two effects. This implies that also the sign of $\frac{\partial D_{ht}}{\partial \Omega}$ is not a priori clear: it depends on how the effects on marginal cost and benefit combine in the medical staff behavior, and how these

[30] Notice that this would imply reporting fewer donors directly by the ICU staff, ruling out effects of perceived corruption on oppositions.

effects add up at the hospital level. Our findings suggest that the impact on marginal benefit is larger than the impact on marginal cost. Hence, the reduction of intrinsic motivation (both as a direct effect of perceived corruption, and as an indirect effect working through expected oppositions) explains the reduction in the number of reported donors.

We can use our framework also to analyze effect duration. We find that the impact of Ω on the intrinsic reward is stronger than the impact on peer pressure, but it is short-lived, likely reflecting an emotional reaction on medical staff morale. In this case, the model predicts a decrease in the number of reported donors in the very short run due to the decline in motivation and the increase in potential oppositions. However, once the reducing effect on Γ vanishes, D will increase at a level *higher* than before the scandal. This occurs because the cost of effort is lower now, since beliefs about peers' efforts have been updated downward in response to the scandal. This discussion helps rationalize Figure 3.5.

Table 3.7: Change in Opposed and Actual Donations

	Raw Number			Average	Share
	(1)	(2)	(3)	(4)	(5)
Panel A	Dependent Variable: Oppositions to Organ Donation				
Surgeon	-0.002	-0.003	-0.002	-0.001	-0.001
	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)
Surgeon × Affected	-0.005	-0.001	-0.005	-0.001	-0.001
	(0.003)	(0.002)	(0.003)	(0.002)	(0.001)
CEO	-0.002	-0.002	-0.002	-0.002	-0.000
	(0.002)	(0.002)	(0.003)	(0.002)	(0.001)
CEO × Affected	0.002	0.002	0.002	0.002	0.000
	(0.003)	(0.003)	(0.003)	(0.002)	(0.002)
Observations	4,320	4,320	4,320	4,176	4,320
Panel B	Dependent Variable: Actual Donors				
Surgeon	0.004	0.003*	0.003	0.000	0.002
	(0.004)	(0.002)	(0.004)	(0.002)	(0.002)
Surgeon × Affected	-0.013***	-0.001	0.004	-0.004	-0.002
	(0.005)	(0.002)	(0.005)	(0.002)	(0.002)
CEO	0.004	0.004*	0.005	-0.000	0.003*
	(0.003)	(0.002)	(0.004)	(0.003)	(0.002)
CEO × Affected	-0.002	-0.003	-0.004	0.002	-0.004**
	(0.004)	(0.002)	(0.004)	(0.003)	(0.002)
Observations	4,320	4,320	4,320	4,176	4,320
Controls	✓	✓	✓	✓	✓
Fixed Effects					
Month	✓	✓	✓	✓	✓
Year	✓	✓	✓	✓	✓
Hospital	✓	✓	✓	✓	✓
Control for Reported Donors					
Reported Donors	-	✓	-	✓	✓
Predicted Reported Donors	-	-	✓	-	-

Note: This table shows the DiD estimates of the effect of the media coverage of the CEO and surgeon scandals on oppositions to organ donation (Panel A) and on the number of actual donors (Panel B). The dependent variable is the number of oppositions to organ donation (Panel A) and the number of actual donors (Panel B) at the hospital-month-year level. The weighted number of newspaper articles is constructed by weighting the provincial number of newspaper articles on both scandals by the respective yearly newspapers market shares shown in Figure 3.4. Affected regions include Piedmont and Aosta Valley. All specifications are estimated using OLS. The specification in Column (2) controls for the number of reported donors. The specification in Column (3) controls for the predicted number of reported donors as predicted by the estimates in Table 3.3, Column (2). In Column (4), the dependent variable is the three-month moving average of oppositions to organ donation (Panel A) and actual donors (Panel B). In Column (5), the dependent variable is the share of oppositions to organ donation (Panel A) or actual donors (Panel B) over the number of reported donors. All specifications include population and newspaper circulation controls and year, month, and hospital fixed effects. Bootstrapped (1,000 replications) and wild clustered standard errors at the hospital level are in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 3.8: Text Analysis of Newspaper Articles

	Semantics			Topic Distribution	
	1st Principal Component	Subjectivity	Polarity	1st Principal Component	2nd Principal Component
	(1)	(2)	(3)	(4)	(5)
Surgeon Articles	-0.027	0.067	0.024	5.631***	-0.252
	(0.056)	(0.084)	(0.278)	(0.637)	(0.206)
Observations	245	245	245	245	245
Fixed Effects					
Year	✓	✓	✓	✓	✓
Month	✓	✓	✓	✓	✓
Newspaper	✓	✓	✓	✓	✓
Controls	✓	✓	✓	✓	✓

Note: This table shows estimates from regressing various text-based metrics on an indicator of Surgeon Articles. The sample includes newspaper articles covering the Surgeon or CEO scandal published in the major Italian newspaper. Columns (1)–(3) consider semantic metrics as described in Section 3.6.2. Columns (4)–(5) consider the first two principal components of the topic model trained on the article corpus. All specifications include year, month, and newspaper fixed effects. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

3.7 Conclusion

Corruption is pervasive, and the health care sector is not immune. In this paper, we show that the welfare costs of corruption go beyond inefficiencies in public spending or inequitable care access. Instead, we need to consider also additional indirect costs implied by the impact of media reporting of corruption news on medical staff behavior, motivation, and morale.

Our analysis focuses on the organ procurement process in Italy. We study the behavioral responses of the medical staff to the disclosure of two corruption scandals at one of the largest Italian hospitals. We differentiate between a case of corruption involving a famous heart surgeon's use of defective heart valves and a generic case involving the hospital CEO, who favored prospective patients on the waiting list in exchange for bribes.

We quantify the response of medical staff to the scandals using a difference-in-differences (DiD) design. We compare hospitals in affected and control regions before and after the scandals measuring the perceived level of corruption via the number of newspaper articles on the corruption scandals. The identification strategy is favored by the unique role of the medical staff at intensive care units in keeping reported donors in a physical state useful for their organs to be transplanted and the exogeneity of corruption scandals with respect to reported donors. The empirical approach based on a DiD model is supported by an event study allowing to test formally for the parallel trend assumption.

Our findings show remarkable effects of the surgeon's case of corruption on organ donations, driven by behavioral changes among health professionals. On the contrary, we do not find any significant effect stemming from the corruption case involving hospital management. Findings are robust to several robustness checks. We provide evidence that results are not driven by the choice of the media outlet used to measure the shock on perceived corruption or by the choice of the time window around the scandals. The estimated impact is short-lived and limited within the administrative boundaries of the procurement center hit by the scandals. Even if media resonance was national, we do not find any evidence of the procurement process in the regions bordering those hit by the scandals.

Our results have important policy implications. First, since we have shown that the negative welfare consequences stemming from corruption in the healthcare sector extend to medical staff motivation and morale, there is an additional reason to fight corruption. Welfare effects from one corruption case would imply several patients unable to obtain transplants, resulting in more economic resources spent to provide continued care to patients still on the waiting list and significant losses in terms of their health. For instance, considering the long waiting lists for a kidney transplant, Becker, Elias and Ye (2022) suggest a total expenditure in present value of about 650,000$ per person on dialysis, to which one should add the value of 15 additional years of life to those transplanted. Second, our results underscore the importance of medical staff motivation and expectations. The standard narrative focuses on patients' and their relatives' trust when discussing organ donations. However, workers' motivation is likely equally (or even more) important, especially in large-scale organizations like modern hospitals. Implementing effective auditing policies or encouraging whistle-blowing might be a way to promote patients' health in healthy hospitals.

Bibliography

Abadie, Alberto, and Sebastien Gay. 2006. "The Impact of Presumed Consent Legislation on Cadaveric Organ Donation: A Cross-Country Study." *Journal of Health Economics*, 25(4): 599–620.

Akbarpour, Mohammad, Julien Combe, Yinghua He, Victor Hiller, Robert Shimer, and Olivier Tercieux. 2020. "Unpaired Kidney Exchange: Overcoming Double Coincidence of Wants Without Money." *Proceedings of the 21st ACM Conference on Economics and Computation*, 465–466.

Alsan, Marcella, and Marianne Wanamaker. 2018. "Tuskegee and the Health of Black Men." *Quarterly Journal of Economics*, 133(1): 407–455.

Arrow, Kenneth J. 1963. "Uncertainty and the Welfare Economics of Medical Care." *American Economic Review*, 53(5): 941–973.

Ash, Elliott, Sergio Galletta, Dominik Hangartner, Yotam Margalit, and Matteo Pinna. 2021. "The Effect of Fox News on Health Behavior during COVID-19." *SSRN Working Paper*, 3636762.

Becker, Gary S., and Julio J. Elías. 2007. "Introducing Incentives in the Market for Live and Cadaveric Organ Donations." *Journal of Economic Perspectives*, 21(3): 3–24.

Becker, Gary S., Julio Jorge Elias, and Karen J. Ye. 2022. "The Shortage of Kidneys for Transplant: Altruism, Exchanges, Opt In vs. Opt Out, and the Market for Kidneys." *Journal of Economic Behavior & Organization*, 202: 211–226.

Bergman, Alon, Matthew Grennan, and Ashley Swanson. 2021. "Lobbying Physicians: Payments from Industry and Hospital Procurement of Medical Devices." *NBER Working Paper*, 29583.

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

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

Bottan, Nicolas L., and Ricardo Perez-Truglia. 2015. "Losing My Religion: The Effects of Religious Scandals on Religious Participation and Charitable Giving." *Journal of Public Economics*, 129: 106–119.

Byrne, Margaret M., and Peter Thompson. 2001. "A Positive Analysis of Financial Incentives for Cadaveric Organ Donation." *Journal of Health Economics*, 20(1): 69–83.

Centro Nazionale Trapianti, Ministero della Salute. 2014. "Regione Piemonte Valle d'Aosta - Attivita di Donazione e Trapianto di Cornee e Membrane Amniotiche Anno 2014." *Policy Report*.

Clemens, Jeffrey, and Joshua D. Gottlieb. 2014. "Do Physicians' Financial Incentives Affect Medical Treatment and Patient Health?" *American Economic Review*, 104(4): 1320–49.

Cornelissen, Thomas, Christian Dustmann, and Uta Schönberg. 2017. "Peer Effects in the Workplace." *American Economic Review*, 107(2): 425–56.

Crea, Giovanni, Matteo M. Galizzi, Ismo Linnosmaa, and Marisa Miraldo. 2019. "Physician Altruism and Moral Hazard: (No) Evidence from Finnish National Prescriptions Data." *Journal of Health Economics*, 65: 153–169.

Daniele, Gianmarco, Arnstein Aassve, and Marco Le Moglie. 2023. "Never Forget the First Time: The Persistent Effects of Corruption and the Rise of Populism in Italy." *The Journal of Politics*, 85(2): 468–483.

de Chaisemartin, Clément, and Xavier D'Haultfœuille. 2020. "Two-way Fixed effects Estimators with Heterogeneous Treatment Effects." *American Economic Review*, 110(9): 2964–2996.

de Chaisemartin, Clément, and Xavier D'Haultfœuille. 2023. "Two-way Fixed Effects and Differences-in-Differences Estimators with Several Treatments." *Journal of Econometrics*, 236(2): 105480.

Deserranno, Erika, and Gianmarco León Ciliotta. 2021. "Promotions and Productivity: The Role of Meritocracy and Pay Progression in the Public Sector." *CEPR Discussion Paper*, 15837.

Elías, Julio J., Nicola Lacetera, and Mario Macis. 2015. "Sacred Values? The Effect of Information on Attitudes toward Payments for Human Organs." *American Economic Review*, 105(5): 361–65.

Elías, Julio J., Nicola Lacetera, and Mario Macis. 2019. "Paying for Kidneys? A Randomized Survey and Choice Experiment." *American Economic Review*, 109(8): 2855–88.

Enikolopov, Ruben, Maria Petrova, and Konstantin Sonin. 2018. "Social Media and Corruption." *American Economic Journal: Applied Economics*, 10(1): 150–74.

Feeley, Thomas Hugh, and Donald Vincent III. 2007. "How Organ Donation Is Represented in Newspaper Articles in the United States." *Health Communication*, 21(2): 125–131.

Ferraz, Claudio, and Frederico Finan. 2008. "Exposing Corrupt Politicians: The Effects of Brazil's Publicly Released Audits on Electoral Outcomes." *Quarterly Journal of Economics*, 123(2): 703–745.

Finan, Frederico, Benjamin A. Olken, and Rohini Pande. 2017. "The Personnel Economics of the Developing State." *Handbook of Economic Field Experiments*, 2: 1043–1171.

Frascà, Giovanni M., Giovanni Gaffi, Domenica Taruscia, Mario D'Arezzo, Luigi Benozzi, and Stefano Sagripanti. 2009. "Renal Transplantation From Living Donor in Italy and Europe." *Giornale Italiano di Nefrologia*, 26(4): 443–451.

Godager, Geir, and Daniel Wiesen. 2013. "Profit or Patients' Health Benefit? Exploring the Heterogeneity in Physician Altruism." *Journal of Health Economics*, 32(6): 1105–1116.

Goodman-Bacon, Andrew. 2021. "Difference-in-Differences with Variation in Treatment Timing." *Journal of Econometrics*, 225(2): 254–277.

Harbaugh, Calista, Majed Afana, Stephanie Burdick, Joseph East, Sindhura Kodali, Jay Lee, Shaun Patel, Govind Rangrass, David Ranney, Vikram Sood, Raymond Lynch, Christopher J. Sonnenday, Michael J. Englesbe, and Amit K. Mathur. 2011. "Portrayal of Organ Donation and Transplantation on American Primetime Television." *Clinical Transplantation*, 25(4): E375–E380.

Howard, David H. 2007. "Producing Organ Donors." *Journal of Economic Perspectives*, 21(3): 25–36.

Jarl, Johan, and Ulf Gerdtham. 2012. "Economic Evaluations of Organ Transplantations." *Nordic Journal of Health Economics*, 1(1): 61–82.

Jensen, Cathrine Elgaard, Preben Sørensen, and Karin Dam Petersen. 2014. "In Denmark Kidney Transplantation Is More Cost-Effective Than Dialysis." *Danish Medical Journal*, 61(3): 1–5.

Johnson, Eric J., and Daniel Goldstein. 2003. "Do Defaults Save Lives?" *Science*, 302(5649): 1338–1339.

Kaserman, David L., and Andy H. Barnett. 2002. *The U.S. Organ Procurement System. A Prescription for Reform.* The AEI Press, Washington D.C.

Kessler, Judd B., and Alvin E. Roth. 2012. "Organ Allocation Policy and the Decision to Donate." *American Economic Review*, 102(5): 2018–2047.

Kolstad, Jonathan T. 2013. "Information and Quality When Motivation Is Intrinsic: Evidence from Surgeon Report Cards." *American Economic Review*, 103(7): 2875–2910.

Le Moglie, Marco, and Gilberto Turati. 2019. "Electoral Cycle Bias in the Media Coverage of Corruption News." *Journal of Economic Behavior & Organization*, 163: 140–157.

Li, Danyang, Zackary Hawley, and Kurt Schnier. 2013. "Increasing Organ Donation Via Changes in the Default Choice or Allocation Rule." *Journal of Health Economics*, 32(6): 1117–1129.

Li, Jing. 2018. "Plastic Surgery or Primary Care? Altruistic Preferences and Expected Specialty Choice of US Medical Students." *Journal of Health Economics*, 62: 45–59.

Li, Jing, William H. Dow, and Shachar Kariv. 2017. "Social Preferences of Future Physicians." *Proceedings of the National Academy of Sciences*, 114(48): E10291–E10300.

Matesanz, Rafael. 1996. "The Panorama Effect on Altruistic Organ Donation." *Transplantation*, 62(11): 1700–1701.

Mendeloff, John, Kilkon Ko, Mark S. Roberts, Margaret Byrne, and Mary Amanda Dew. 2004. "Procuring Organ Donors as a Health Investment:

How Much Should We Be Willing to Spend?" *Transplantation*, 78(12): 1704–1710.

Mironov, Maxim, and Ekaterina Zhuravskaya. 2016. "Corruption in Procurement and the Political Cycle in Tunneling: Evidence from Financial Transactions Data." *American Economic Journal: Economic Policy*, 8(2): 287–321.

Morgan, Susan E., Tyler R. Harrison, Lisa Chewning, LaShara Davis, and Mark DiCorcia. 2007. "Entertainment (Mis)Education: The Framing of Organ Donation in Entertainment Television." *Health Communication*, 22(2): 143–151.

Radin, Dagmar. 2013. "Does Corruption Undermine Trust in Health Care? Results From Public Opinion Polls in Croatia." *Social Science & Medicine*, 98: 46–53.

Regione Lazio. 2008. *Procedure per la Diagnosi di Morte Cerebrale e la Donazione d'Organi ai Fini di Trapianto.* Regione Lazio AQ.3 Rev.04/2008.

Roth, Alvin E. 2007. "Repugnance as a Constraint on Markets." *Journal of Economic Perspectives*, 21(3): 37–58.

Scepanovic, Ivo. 2006. "Croatia Cracks Down on Corruption." *The Lancet*, 368(9545): 1410.

Silver, David. 2021. "Haste or Waste? Peer Pressure and Productivity in the Emergency Department." *Review of Economic Studies*, 88(3): 1385–1417.

Strömberg, David. 2015. "Media Coverage and Political Accountability: Theory and Evidence." In *Handbook of Media Economics, 1, 595-622.* , ed. Joel Waldfogel Simon P. Anderson and David Strömberg. North-Holland.

Thompson, John F., Joe C. McCosker, Adrian D. Hibberd, Jeremy R. Chapman, John S. Compton, John F. Mahony, Paul J. Mohacsi, George J. MacDonald, and Paul M. Spratt. 1995. "The Identification of Potential Cadaveric Organ Donors." *Anaesthesia and Intensive Care*, 23(1): 75–80.

Thorne, Emanuel D. 2006. "The Economics of Organ Transplantation." In *Handbook of the Economics of Giving, Altruism and Reciprocity, 2, 1335-1370.* , ed. Serge-Christophe Kolm and Jean Mercier Ythier. Elsevier.

Transparency International, NGO. 2015. "Corruption Perceptions Index 2015." *URL:* *https://www.transparency.org/en/cpi/2015/results/table*, Accessed: 2020-07-14.

Turati, Gilberto. 2013. "The Italian Servizio Sanitario Nazionale: A Renewing Tale of Lost Promises." In *Federalism and Decentralization in European Health and Social Care: Competition, Innovation, and Cohesion, 47-66.* , ed. J. Costa-Font and S. C. Greerg. London: Palgrave MacMillan.

Venettoni, Sante. 2007. *Il Processo di Donazione-Prelievo-Trapianto: Analisi delle Procedure e Ccriticità.* Italian Ministry of Health.

Vian, Taryn. 2008. "Review of Corruption in the Health Sector: Theory, Methods and Interventions." *Health Policy and Planning*, 23(2): 83–94.

Vigdor, Neil. 2020. "It Paid Doctors Kickbacks. Now, Novartis Will Pay a $678 Million Settlement." *URL:* *https://www.nytimes.com/2020/07/01/business/Novartis-kickbacks-diabetes-heart-drugs.html*, Accessed: 2022-05-11.

World Health Organization. 2009. "Global Glossary of Terms and Definitions on Donation and Transplantation." *Geneva: World Health Organization.*

3.A Additional Tables

Table 3.A.1: Descriptive Statistics – Pre- and Post-CEO Scandal

	Affected Regions		Bordering Regions		
	Mean (1)	SD (2)	Mean (3)	SD (4)	Difference (5)
Panel A: Pre-CEO Scandal					
Organ Donations					
Reported Donations	0.653	1.087	0.504	1.074	0.091***
Opposed Donations	0.169	0.437	0.095	0.365	0.091***
Actual Donations	0.463	0.840	0.342	0.777	0.091***
Newspaper Coverage (# Articles)					
Surgeon Scandal	0.000	0.000	0.000	0.000	0.000
Surgeon Scandal (3-Month Avg.)	0.000	0.000	0.000	0.000	0.000
CEO Scandal	0.000	0.000	0.000	0.000	0.000
CEO Scandal (3-Month Avg.)	0.000	0.000	0.000	0.000	0.000
Newspaper Circulation (per 1,000 inhabitants)					
La Stampa	58.641	12.778	7.013	14.632	69.353***
Il Corriere Della Sera	4.178	0.910	43.318	90.376	-7.916***
Observations	242		550		
Hospitals	22		50		
Panel B: Post-CEO Scandal					
Organ Donations					
Reported Donations	0.745	1.129	0.556	1.131	0.306***
Opposed Donations	0.224	0.513	0.113	0.394	0.224***
Actual Donations	0.472	0.811	0.388	0.854	0.041***
Newspaper Coverage (# Articles)					
Surgeon Scandal	1.872	6.069	1.217	3.649	2.513***
Surgeon Scandal (3-Month Avg.)	1.872	4.794	1.216	2.872	2.514***
CEO Scandal	2.004	7.037	1.544	4.353	2.172***
CEO Scandal (3-Month Avg.)	1.995	6.062	1.541	3.692	2.156***
Newspaper Circulation (per 1,000 inhabitants)					
La Stampa	49.761	10.603	6.576	12.431	58.909***
Il Corriere Della Sera	4.829	1.031	35.549	67.396	-5.270***
Observations	1,078		2,450		
Hospitals	22		50		

Note: This table shows descriptive statistics for the sample before and after the CEO scandal. The difference reported in column (5) is the coefficient obtained by regressing an indicator for affected regions on the respective variable controlling for year, month, and hospital fixed effects. Significance levels are constructed from bootstrapped (1,000 replications) and wild clustered standard errors at the hospital level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 3.A.2: Descriptive Statistics – Pre- and Post-Surgeon Scandal

	Affected Regions		Bordering Regions		
	Mean (1)	SD (2)	Mean (3)	SD (4)	Difference (5)
Panel A: Pre-CEO Surgeon					
Organ Donations					
Reported Donations	0.693	1.126	0.502	1.065	0.095***
Opposed Donations	0.223	0.515	0.092	0.357	0.143***
Actual Donations	0.444	0.815	0.349	0.796	-0.000
Newspaper Coverage (# Articles)					
Surgeon Scandal	0.000	0.000	0.000	0.000	0.000
Surgeon Scandal (3-Month Avg.)	0.000	0.000	0.000	0.000	0.000
CEO Scandal	3.930	10.362	2.880	6.334	4.379***
CEO Scandal (3-Month Avg.)	4.299	9.229	3.178	5.497	4.748***
Newspaper Circulation (per 1,000 inhabitants)					
La Stampa	56.515	12.462	6.905	13.928	66.792***
Il Corriere Della Sera	4.222	0.912	40.481	82.389	-6.313***
Observations	462		1,050		
Hospitals	22		50		
Panel B: Post-Surgeon Scandal					
Organ Donations					
Reported Donations	0.747	1.120	0.570	1.149	0.359***
Opposed Donations	0.210	0.493	0.119	0.404	0.231***
Actual Donations	0.485	0.817	0.396	0.864	0.077***
Newspaper Coverage (# Articles)					
Surgeon Scandal	2.352	6.720	1.529	4.032	3.157***
Surgeon Scandal (3-Month Avg.)	2.352	5.268	1.528	3.144	3.159***
CEO Scandal	0.402	1.024	0.390	0.770	0.371***
CEO Scandal (3-Month Avg.)	0.413	0.746	0.388	0.576	0.396***
Newspaper Circulation (per 1,000 inhabitants)					
La Stampa	48.629	10.003	6.498	12.169	57.228***
Il Corriere Della Sera	4.972	1.010	34.878	65.396	-5.541***
Observations	858		1,950		
Hospitals	22		50		

Note: This table shows descriptive statistics for the sample before and after the Surgeon scandal. The difference reported in column (5) is the coefficient obtained by regressing an indicator for affected regions on the respective variable controlling for year, month, and hospital fixed effects. Significance levels are constructed from bootstrapped (1,000 replications) and wild clustered standard errors at the hospital level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 3.A.3: Effect of Corruption News on Reported Donors

	Dependent Variable: Number of Reported Donors					
	OLS Model			Poisson Model		
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Total Number of Articles						
Total Articles	-0.005*	-0.006	-0.006	-0.006	-0.008*	-0.008
	(0.003)	(0.004)	(0.004)	(0.003)	(0.005)	(0.005)
Observations	1,320	1,320	1,320	1,320	1,320	1,320
Panel B: Disaggregated Number of Articles						
Surgeon	-0.016***	-0.016***	-0.015***	-0.029***	-0.029***	-0.029***
	(0.005)	(0.006)	(0.005)	(0.009)	(0.007)	(0.007)
CEO	0.003	0.001	0.001	0.004	0.000	0.000
	(0.004)	(0.005)	(0.005)	(0.004)	(0.005)	(0.006)
Observations	1,320	1,320	1,320	1,320	1,320	1,320
Controls	✓	✓	✓	✓	✓	✓
Fixed Effects						
Hospital	-	✓	✓	-	✓	✓
Month	-	✓	✓	-	✓	✓
Year	-	✓	✓	-	✓	✓
Month × Hospital	-	-	✓	-	-	✓

Note: This table shows the effect of the media coverage of the CEO and surgeon scandals on the number of reported donors. The dependent variable is the number of reported donors at the hospital-month-year level. The weighted number of newspaper articles is constructed by weighting the provincial number of newspaper articles on both scandals by the yearly newspaper market shares shown in Figure 3.4. Specifications in columns (1)–(3) are estimated using OLS, and columns (4)–(6) are estimated as Poisson regression models. In columns (1)–(6) we include controls for population and newspaper circulation. In columns (2) and (5), we include year, month, and hospital fixed effects. We interact with month and hospital fixed effects in columns (3) and (6). Bootstrapped (1,000 replications) and wild clustered standard errors at the hospital level are in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

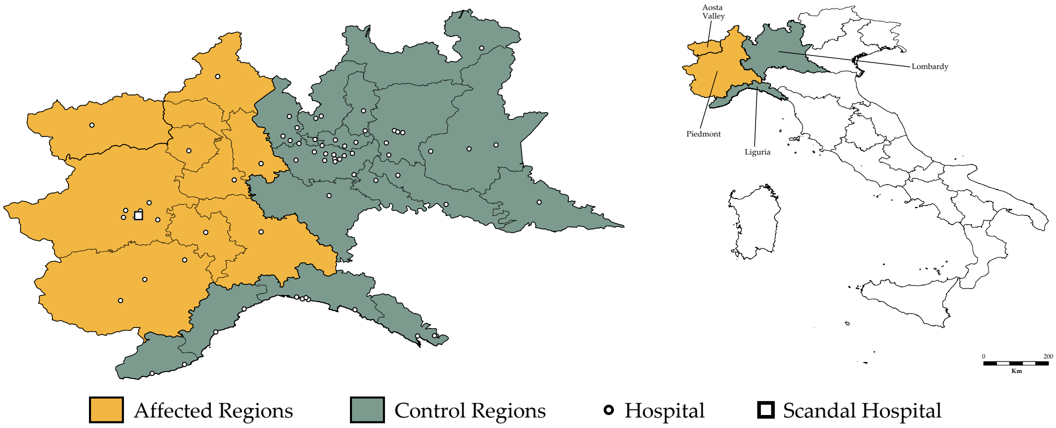
Table 3.A.4: TV News Coverage of Corruption and Reported Donors

	Dependent Variable: Number of Reported Donors					
	OLS Model			Poisson Model		
	(1)	(2)	(3)	(4)	(5)	(6)
Surgeon	-0.015***	-0.017***	-0.017***	-0.029***	-0.033***	-0.033***
	(0.005)	(0.006)	(0.006)	(0.009)	(0.010)	(0.010)
CEO	-0.000	-0.002	-0.002	0.001	-0.003	-0.003
	(0.003)	(0.004)	(0.004)	(0.004)	(0.006)	(0.006)
Observations	1,320	1,320	1,320	1,320	1,320	1,320
Controls	✓	✓	✓	✓	✓	✓
Fixed Effects						
Hospital	-	✓	✓	-	✓	✓
Month	-	✓	✓	-	✓	✓
Year	-	✓	✓	-	✓	✓
Month × Hospital	-	-	✓	-	-	✓

Note: This table shows the effect of the media coverage of the CEO and surgeon scandals on the number of reported donors. The dependent variable is the number of reported donors at the hospital-month-year level. TV coverage is measured as the number of news about the two corruption scandals broadcast by the regional TV news in Piedmont and the Aosta Valley. Specifications in columns (1)–(3) are estimated using OLS, and columns (4)–(6) are estimated as Poisson regression models. In columns (1)–(6) we include controls for population and newspaper circulation. In columns (2) and (5), we additionally include year, month, and hospital fixed effects. In columns (3) and (6), we additionally interact with month and hospital fixed effects. Bootstrapped (1,000 replications) and wild clustered standard errors at the hospital level are in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

3.B Additional Figures

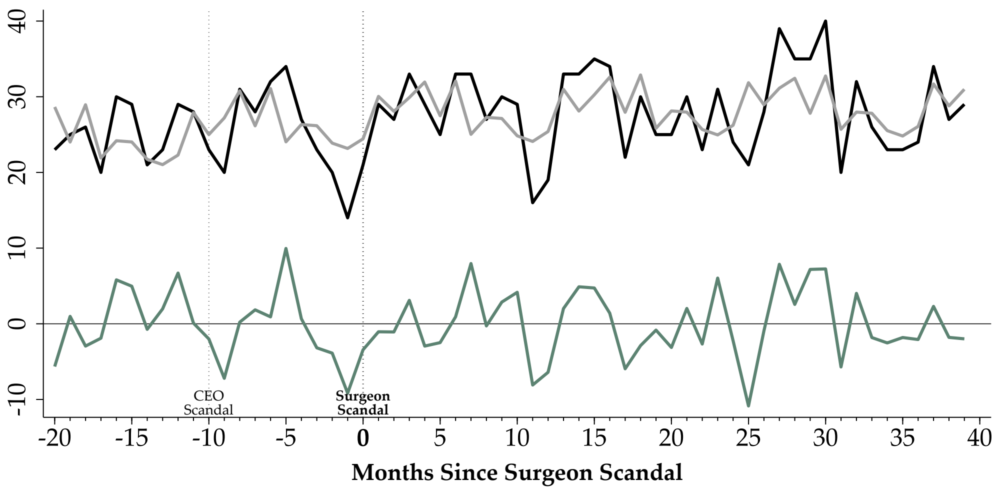
Figure 3.B.1: Hospital and Region Locations



Note: This map shows the location of Piedmont and Aosta Valley (yellow) as well as Lombardy and Liguria (green). The map also illustrates the exact location of organ transplant units in the regions analyzed in this study.

Figure 3.B.2: Counterfactual Effects by Time: Bordering Regions

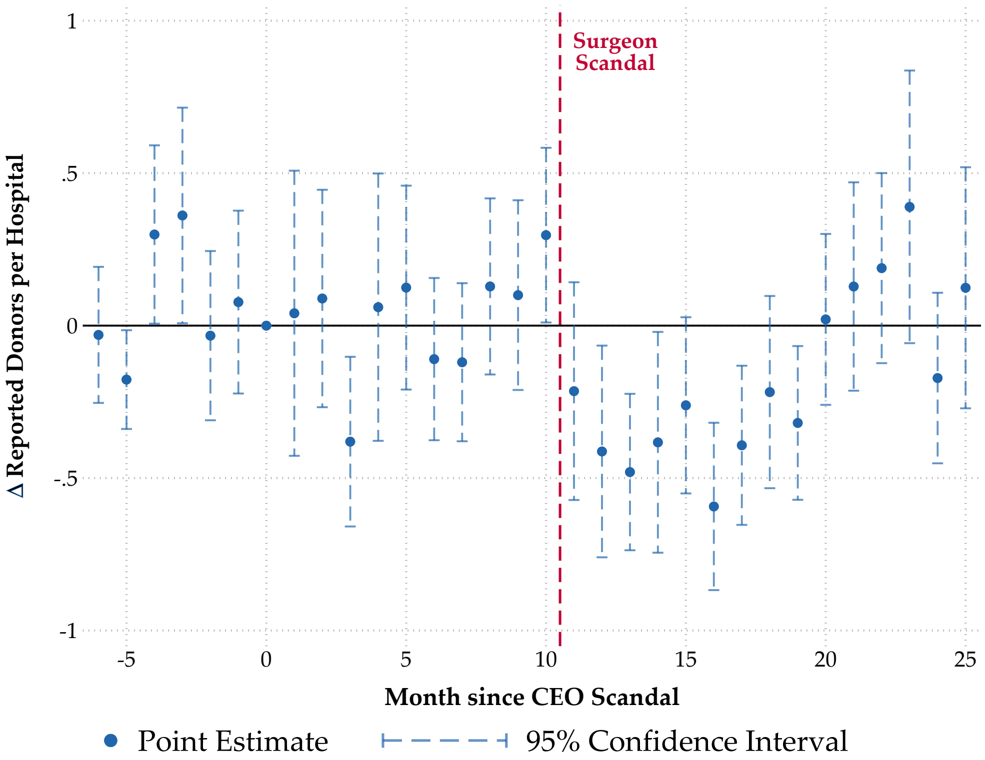
(a) Bordering Regions: CEO Scandal



(b) Bordering Regions: Surgeon Scandal

Note: The figure depicts actual and counterfactual reported organ donations across time for the case of bordering regions. The counterfactual is constructed by predicting reported donors using the estimates from Table 3.3, column (2) and by setting to zero the coefficient for the number of articles on the surgeon case. The time series on reported donors results from averaging the hospital-level data in the bordering regions by year-month and subsequently centering it around the onset of the surgeon scandal.

Figure 3.B.3: Joint Event-study Estimates



Note: The figure displays point estimates and 95 percent confidence intervals on the effect of the Surgeon scandal on the number of reported donors. All estimates are based on the regression model in Equation 3.2. Standard errors are clustered at the hospital level.

3.C Additional Analyses

In addition to the DiD specification described in Section 3.4 we estimate the following specification:

$$D_{ht}^{R} = \beta N_{p(h)t} + \boldsymbol{X}_{p(h)t}\gamma + f(\theta_h, \delta_t) + \varepsilon_{ht}, \tag{3.7}$$

where D_{ht}^{R} represents the monthly number of reported organ donors identified by the medical staff in each hospital h located in province $p(h)$ in period (month-year) t. We proxy the perception of corruption using the number of newspaper articles dealing with the two scandals at Hospital 'Molinette' between 2001 and 2005. N is the weighted number of monthly (t) newspaper articles about the two corruption cases circulated in province p. We consider N as a shifter for optimal effort e^{*}. $\boldsymbol{X}_{p(h)t}$ is a vector of controls for population, newspaper circulation at the provincial level, and hospital-specific time trends. Finally, several combinations of hospital, month, and year fixed effects are included in the model. β is the coefficient of interest for the analysis as it represents the effect of an extra article on the corruption scandals on the number of reported donors.

Cases of corruption involving different actors in the organ donation process imply differential effects on reported donors, e.g., the medical staff may perceive the information content of a case involving a surgeon versus a CEO differently. Therefore, we enrich the previous model by considering the two cases of corruption separately:

$$D_{ht}^{R} = \beta_1 N_{p(h)t}^{C} + \beta_2 N_{p(h)t}^{S} + \boldsymbol{X}_{p(h)t}\gamma + f(\theta_h, \delta_t) + \varepsilon_{ht} \tag{3.8}$$

where N^{C} is the weighted number of newspaper articles referring to the corruption case involving the hospital CEO, and N^{S} is the weighted number of newspaper articles about the case of corruption involving the hospital surgeon. In this case, we are interested in the coefficients β_1 and β_2 that measure the effect on the number of reported donors in response to an additional article about the CEO or surgeon scandal, respectively.

Declaration

This dissertation is the result of my own work, and no other sources or means, except the ones listed, have been employed.

Mannheim, April 17, 2025

Maximilian Mähr

Curriculum Vitæ

2020–2025	University of Mannheim (Germany) *Ph.D. in Economics*
2017–2020	University of Bonn (Germany) *M.Sc. in Economics*
2014–2017	University of Münster (Germany) *B.Sc. in Economics*