On the Mechanics Behind Academic Progress

Zur Erlangung des akademischen Grades eines
Doktors der Wirtschaftswissenschaften

(Dr. rer. pol.)

von der KIT-Fakultät für Wirtschaftswissenschaften
des Karlsruher Instituts für Technologie (KIT)

genehmigte
DISSERTATION

von

B.Sc. Wi.-Ing. Lars Herberholz

Tag der mündlichen Prüfung:	16.12.2020
Referent:	Professor Dr. Berthold U. Wigger
Korreferentin:	Professor Dr. Nora Szech

Karlsruhe, Dezember 2020

Contents

List of Tables

List of Figures

Acknowledgements

Over the last years, I have benefitted from the amazing support of several people. First and foremost, I am deeply indebted to my primary advisor, Prof. Dr. Berthold U. Wigger, for his encouragement, guidance, and trust. I owe my understanding of economic research mainly to him. I am also very grateful to my secondary advisor, Prof. Dr. Nora Szech, for her insightful feedback and ideas. Moreover, I wish to sincerely thank Prof. Dr. Martin Klarmann and Prof. Dr.-Ing. Thomas Lützkendorf for their vital roles on my dissertation committee.

I had the privilege to be surrounded by many colleagues and friends who supported me throughout my dissertation project. In particular, Dr. Benjamin Bittschi has been a great co-author, Ezra Setio and Nicolas Renkamp provided tremendous assistance with the data collection, and Dominik Zimmermann and Timo Reinheimer spared plenty of time to discuss my research plans, both during regular working hours and beyond. I am immensely thankful to all of them.

Most importantly, I wish to thank my parents for their unconditional support and Maria for her endless patience and love along the way.

1 Introduction

For centuries, the merits of knowledge acquisition have been highlighted by economists, philosophers, and politicians alike. Prominent examples include Benjamin Franklin, Alfred Marshall, or Karl Marx who all described knowledge as a valuable asset, partly in view of its growing importance for production purposes. At the present time, this view is almost universally acknowledged. Returns to education are not only confined to higher earnings but may also take non-pecuniary forms, such as increased health and happiness (see e.g., Oreopoulos, 2007). At the aggregate level, education contributes to societal welfare and economic prosperity. The latter aspect can largely be seen as resulting from technological change, i.e., the flow of new ideas. As the shift towards the knowledge society continues, both innovation and highly skilled labour are expected to gain further importance. Yet, as recent evidence suggests, ideas are getting harder to find (Bloom, Jones, van Reenen, & Webb, 2020). In a similar vein, Jones (2010) reports that innovators (need to) undertake more and more training prior to their active careers, which caused an estimated drop in life cycle innovation potential of 30% over the 20th century. Against this background, the present thesis investigates the mechanisms underlying academic progress, upon which research and higher education may be fostered. Conceptually, I will address three related topics from the economics of science and innovation; these are presented below.

The second chapter is titled "Collapsing Stars and the Diffusion of Scientific Knowledge" and builds on joint work with Benjamin Bittschi. The focus of this chapter lies on the role of star scientists in the knowledge production process. Not only do eminent scientists account for a great portion of research contributions, but they also tend to occupy central positions in their collaboration networks. Prior studies have established that these stars constitute the origin of spillover effects, thus lifting the output trajectories of their co-authors. Given that the existing evidence is limited to specific fields of science, we aim to broaden the understanding of these effects by exploring the entire subject spectrum. For this purpose, we compile a rich bibliometric database comprising metadata of 15.6 million publications over the period from 1996 to 2015. By means of these data, we are able to define star scientists through performance criteria and further delineate a group of 162 stars that died both prematurely and unexpectedly. In the aftermath of these lethal shocks, we estimate a publication and citation deficit of treated collaborators that ranges from 4.2 to 7.8%, relative to a matched control group. Yet field-specific analyses reveal considerable heterogeneity masked by the aggregate figures. While star effects are most dominant in life sciences, they are less pronounced in physical sciences, take only nuanced forms in health sciences, and seem absent in social sciences. Furthermore, we discover an interplay of three main effect channels which shed light on the transmission of scientific knowledge. More specifically, in certain fields, spillovers are driven by spatial elements and both eminent collaborators and collaborators with markedly different field expertise than their star are more severely affected by the death event.

The third chapter studies the "Efficiency of European Universities" and is based on joint work with Berthold U. Wigger. In the course of this chapter, we present insights into the input-output transformation in the higher education sector. Methodologically, we rely on frontier techniques rooted in the Operations Research literature that allow for (relative) efficiency assessments. We argue that existing applications of these tools to universities are prone to unsuitable comparisons since inherent differences between institutions are largely left unconsidered. For instance, substantially higher expenses in medical relative to social studies are not necessarily a display of inefficiency but rather a reflection of distinct technologies. Instead of pooling the university landscape, we therefore apply peer-group selection methods centred around subject space proximity to obtain unbiased efficiency estimates of 450 European universities. This analysis partly builds on bibliometric data from the previous chapter but further includes official statistics such as financial figures or student and graduate numbers, which are accessible from 2011 to 2014. Provided with efficiency measures, potential efficiency drivers are moved into focus. Consistent with the study on star scientists, subject orientation is again identified as a source of heterogeneity, i.e., the relation between efficiency and several structural attributes varies depending on the university's subject profile. However, both the ability to seek third-party funding and institutional size are mostly associated with higher efficiency.

The fourth chapter provides new empirical evidence on the implications of "Competitive Funding in Academia". Higher education sectors in most of continental Europe have been targeted by New Public Management inspired reforms over the last decades. Competitive funding schemes are a key element of this policy trend as they are hoped to improve the allocation of resources and fuel innovation. Yet, New Public Management has (frequently) drawn criticism due to unintended consequences. I add to this debate by examining the Quality Pact for Teaching, a large-scale funding program that aims to promote the quality of academic teaching in Germany. The program comprises almost 2 billion euros, which are spent over 10 years from 2011 onwards. The intriguing feature of this funding scheme is that German higher education institutions are for the first time exposed to competition for a notable amount of teaching aids. From exploring the grant allocation, I find evidence for a Matthew effect pattern. Stated differently, past third-funding volume emerges as a significant determinant of funding success at the Quality Pact for Teaching. Accordingly, it might be worried that the program unintentionally discriminated against institutions that were unaccustomed to grant competition which would contradict the stated goal of supporting the academic landscape at a broad range.

In the end, the fifth chapter offers a brief conclusion of the mechanics behind academic progress by reflecting on the previous chapters. Apart from linking the central findings of my thesis, an outlook on future research avenues will be presented.

2 Collapsing Stars and the Diffusion of Scientific Knowledge[†]

2.1 Introduction

Some stars collapse straight into darkness (Adams, Kochanek, Gerke, Stanek, & Dai, 2017). Our understanding of these rare events is limited in an astronomical sense, as it is of the consequences for the scientific community once it loses its brightest minds. The present paper adds to the second line of inquiry. Star scientists are known to play a central role in the production of knowledge (Zucker & Darby, 1996), hereby fostering economic growth and social welfare (Romer, 1990). If their contributions were to end abruptly, what scars would be left behind?

Our attempt to answer this question revolves around the fate of scientists that formerly collaborated with a star. Unlike the romantic ideal, innovation is rarely achieved through the creativity of lone genius. Instead, teamwork has become increasingly prevalent and impactful in today's science and technology (Bercovitz & Feldman, 2011; Singh & Fleming, 2010; Wuchty, Jones, & Uzzi, 2007). Star scientists, in particular, are embedded in large co-author networks (de Solla Price & Beaver, 1966; Zuckerman, 1967). Given the level of freedom scientists are provided with, it can reasonably be assumed that these networks result from active search-and-matching processes (see e.g., Stephan, 2012, Chapter 4); in other words, they are formed endogenously. The end of a collaborative tie, in contrast, might occur exogenously and therefore open up a pathway for causal inference. To be more precise, we use the premature and unexpected death of outstanding scientists as a quasi-experiment and explore empirically how these lethal shocks affect the research productivity and quality of former co-authors. In doing so, we shed light on the nature of interpersonal knowledge spillovers.

The process of human capital formation is central for any modern society, but, as the stock of knowledge grows, also demands more and more effort from scientists on their way to the research frontier. As a consequence, it might be suspected that innovative phases are on the decline (Jones, 2009). Against this background, it appears all the more important to investigate spillover effects as a potential means to spur scientific progress. In approaching this topic, we build on a number of studies, most notably the work of Azoulay, Graff Zivin, and Wang (2010) who laid the conceptual foundations by disclosing how collaborators fare in the aftermath of "superstar extinction". Yet, we aim to extend the existing literature along several dimensions. First, hitherto evidence is drawn from specific scientific areas including physical sciences (Waldinger, 2012, 2016), life sciences

[†] This chapter is based on joint work with Benjamin Bittschi.

(Azoulay et al., 2010), medicine (Mohnen, 2018), economics (Ductor, Fafchamps, Goyal, & van der Leij, 2014) mathematics (Borjas & Doran, 2012, 2015; Waldinger, 2010) or even more narrow disciplines such as evolutionary biology (Agrawal, McHale, & Oettl, 2017) or immunology (Oettl, 2012). In contrast, our dataset allows us to examine spillover effects over the entire subject spectrum and further compare the fields of life, health, physical, and social sciences within a uniform framework. Second, we offer new insights into the origins of spillover effects. In particular, we explore the extent of knowledge flows through interdisciplinary avenues and inspect in how far results are confined to the science system in the United States, which served as the focal point for most previous studies.

Our analysis builds on a (dynamic) conditional difference-in-difference (DiD) design, where the treatment originates from the unexpected passing of 162 star scientists. We identify these stars from a larger group of eminent scientists that either belong to the National Academy of Sciences (NAS) or possess outstanding publication records. In order to define the latter criterion, we compile a rich bibliometric dataset from Scopus, which comprises meta-information of 15.6 million publications over the period from 1996 to 2015. These core data are further complemented with information from Google Maps and GenderAPI, which enables us to follow the scientific footprints of 9.2 million individuals. Moreover, we can assign star status by means of performance indicators such as the H-Index or citation metrics. Delineating star scientists is not only required for the treatment identification, but also essential for the effect estimation. More specifically, we use the set of stars, who did not pass away, and their respective co-authors to assemble a matched control group for the scientists that experience the unexpected loss of a star collaborator.

On aggregate, we discover that the abrupt ending of a star collaboration causes a lasting decline of 4.2% in published articles of treated scientists. Accounting for output quality, we find a pronounced effect in form of a 7.8% decrease in citation-weighted articles. In neither case are recovery patterns observable. However, field-specific estimations reveal that the aggregate view masks substantial variation across the scientific spectrum. While life sciences is characterised by increased treatment effects in both output dimensions, we solely denote a quality-adjusted effect in physical sciences and nuanced, but no overall, effects in health sciences. Lastly, we cannot detect any statistically significant treatment consequences in social sciences. In the subsequent course of analysis, we focus on the mechanisms behind these effects. It hereby becomes evident that the omission of future cooperation is only a partial treatment aspect. Similarly, neither the frequency nor the timing of interaction before the stars' death offers an explanation for the effect origins. Exploring further effect channels also leads us to reject a gatekeeping story based on editorial goodwill. Yet an interplay of three main effect drivers becomes apparent. First, spillovers are in part spatially confined. More concretely, co-location is related to steeper output declines in physical sciences, while, on a broader geographical scale, dyads within the United States largely account for the effects in life and health sciences. Second, we find horizontal spillovers (or peer effects) in life sciences since the treatment especially affects scientists that are likewise stars. Third, in health and physical sciences, we further discover that the break-up of dyads with markedly different field expertise induces more severe effects, which underlines the relevance of interdisciplinary knowledge transmission.

Our paper is linked to several strands of literature. Importantly, we adopt the research design of Azoulay et al. (2010) who discover that spillovers are primarily transmitted in idea space. The work of both Oettl (2012) and Mohnen (2018) is methodologically close, but focuses on different effect channels. The former study reveals that helpful stars play a crucial role for the performance of collaborators, while the latter study yields a similar conclusion for stars with central network positions. In a related setting, Jaravel, Petkova, and Bell (2018) investigate how inventors are affected by the premature death of a (star) co-inventor and document long-lasting declines in patents and earnings. Further research on scientific spillovers has utilised identification strategies other than death events. For instance, Waldinger draws findings from the expulsion of scholars during the Nazi regime (2010, 2012, 2016) and World War II bombing campaigns (2016), while Borjas and Doran (2012, 2015) exploit the collapse of the Soviet Union as a natural experiment. Moreover, our paper relates to the growing "science of team science" literature, which is bound by the question of how to enhance the effectiveness of collaborative research (see Hall et al., 2018, for a recent review). Team composition and especially team diversity are vital aspects of this debate (National Research Council, 2015), to which our results contribute. Thematic overlap also exists with the work of Akcigit, Caicedo, Miguelez, Stantcheva, and Sterzi (2018), König, Lorenz, and Zilibotti (2016), and Lucas and Moll (2014) who examine interaction-based spillover effects through the lens of endogenous growth models. Lastly, our paper belongs to a wider literature that uses (premature) death cases as a source of identification (apart from the aforementioned studies, see e.g., Aizenman and Kletzer, 2011, Jäger and Heining, 2019, Jones and Olken, 2005, or Nguyen and Nielsen, 2010).

The remainder of the paper is structured as follows. Section 2.2 describes our data and the research design. Section 2.3 details the econometric approach and presents our aggregate results. Section 2.4 focuses on effect heterogeneity across scientific disciplines and further explores different channels through which the diffusion of scientific knowledge operates. Section 2.5 offers a discussion of our findings and concludes.

2.2 Data and Research Design

2.2.1 Data Compilation

Our research design is centred on star scientists and their potential spillovers onto collaborators. With this approach in mind, we assembled our core data laying focus on the science systems in North America and Europe, the latter extended by Israel. Especially US-based scientists and inventors have been the subject of previous studies, while their

European counterparts have received markedly less attention in this stream of literature. Constructing a dataset that spans both continents therefore helps to fill a void, but also offers the opportunity to examine whether spillover effects could be rooted in structural differences between the North American and European scientific landscapes (see e.g., Aghion, Dewatripont, Hoxby, Mas-Colell, & Sapir, 2010).

Within the US, we rely on the Carnegie Classification of Institutions of Higher Education and in particular on the category of doctoral universities to delineate our set of research institutions. As of 2015, this category listed 334 institutions, which we manually linked to their Scopus profiles. Elsevier's database covers a wide range of scientific literature and enables us to collect metadata for each publication with affiliative ties to (at least one of) these research institutions. We proceeded in a similar manner with regard to Europe's higher education sector using the European Tertiary Education Register (ETER) as the start point. From this database, we compiled a set of 724 institutions from 26 countries that were consistently classified as universities based on their right to grant doctoral degrees.[1] Lastly, we added universities from both Israel and Canada to the institutional collective given that both countries are home to internationally renowned scientific communities and geographically adjacent. The outlined procedure resulted in an overall list of 1,146 research institutions, on which grounds we collected 15.6 million publication records over the period from 1996 to 2015, each comprising a citation horizon until 2016.

In the following step, we constructed an author-centric dataset for the over 9.2 million academics listed on these publications. The depth of available metadata already allowed depicting publication activities, co-author networks, affiliation histories, or research topics. Yet we further queried Scopus for each author to access data beyond our observation period, e.g., the year of first publication in order to proxy career starts. In addition, we complemented our core data with gender predictions from Gender API, site coordinates from Google Maps, and biographic information from the NAS, which will be explained in more detail over the next subsections. Taken together, our data approach enables us to track a multitude of academic careers over 20 years in time.

2.2.2 The Scientific Elite

Our decision to focus on the brightest scholars is guided by a fundamental property of scientific progress. As already observed by Lotka (1926), the distribution of scientific output is remarkably skewed, illustrating that a prolific minority is responsible for a great amount of contributions. In a similar vein, as Newton claimed, science is found to largely advance

[1] Access to Scopus is limited by quotas. Thus, instead of collecting data for the entire university sample from ETER, we focused on institutions from Austria, Belgium, Bulgaria, Croatia, Cyprus, the Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Iceland, Ireland, Italy, Latvia, Lithuania, Malta, the Netherlands, Norway, Poland, Portugal, Spain, Sweden, Switzerland, and the United Kingdom. We feel confident that this sample provides an adequate level of star power for our empirical purposes.

"on the shoulders of giants" (Bornmann, De Moya Anegón, & Leydesdorff, 2010; Cole & Cole, 1972). Considering their dominant role in the production of knowledge, star scientists are (most) likely to shine on their surroundings and thus provide a natural starting point to investigate spillover effects.

A first look at our raw data underlines the importance of elite scientists. Within the cohort that first published in the year 2000, we find the median scientist to record three career publication, whereas the marginal star, defined as the scientist that marks the start of the top percentile, accumulates 145 publications. Their equivalents with regard to the citation distribution see themselves separated by a comparable margin of 66 versus 3,485 career citations. Defining stardom based on relative performance is indeed common practice.[2] We follow this approach and maintain the top percentile as the threshold for awarding star status. However, we rely on a more refined set of metrics to measure accomplishments. We begin by delineating a set of stars according to their H-Index, which we calculate over both a 5-year window for publications and citations. For instance, our first star cohort is compiled in 2001 and comprises the group of scholars with the highest H-Indices based on their research output from 1996 to 2000. To account for different timings in output, we include citations if they accrued within the first five years after publication. The H-Index is essentially designed to provide a balanced measure of research quantity and quality, so that stars of either domain are not captured by it. While we are not concerned with omitting scientists that only generate large quantities of work, we do intend to include scientists that even occasionally shift the research frontier through seminal papers. Thus, we add forward citations (i.e., citation-weighted publication counts) as a second star criterion to our performance catalogue, again employing 5-year windows for calculation.

Next, we extend our star criteria with co-author adjusted versions of both metrics. We are generally in favour of measures that account for variations in team size. Yet when it comes to classifying star scientists, unadjusted metrics are likely to be thought of as providing complementary value. More specifically, they carry a network component and may help identifying stars that are very well positioned and possibly facilitate knowledge flows by connecting numerous co-authors (Mohnen, 2018).[3]

[2] Rothaermel and Hess (2007) assign star status to scientists that accumulate publications and citations three standard deviations above the mean. Using similar outcome measures, Jaravel et al. (2018) refer to stars upon exceeding the 98th percentile, while Waldinger (2016) sets the cut off at the 95th percentile.

[3] In case of forward citations, we divide the number of citations of each publication by the number of its authors before aggregating these counts at the scientist level. As for the H-Index, we use a modification that counts publications fractionally, again according to the number of listed authors (see Schreiber, 2008, for details). It is worth noting, however, that adjusted and unadjusted metrics do not capture achievement in a completely different sense since almost half of the final star sample satisfy both types of criteria.

Based on these four performance criteria, we identify stars on a yearly basis from 2001 to 2012. Since publishing practices differ considerably across the scientific spectrum, we do so separately by field. We hereby follow Elsevier's classification system and assign scientists to one of 26 scientific fields (see Appendix A.1)[4] based on the distribution of their past publications. More specifically, these distributions denote how often scientists have published in each specific field, i.e., in journals that are classified under a given field. We assign scientists to their mode field and, if necessary, break ties at random. Moreover, we restrict the star delineation to research articles to ensure that a certain editorial standard is met by all papers, but also to avoid potential double counting of articles and former conference papers. Overall, this procedure yields a set of 154,205 eminent scientists. As a fifth and final criterion, we define 3,458 members of the NAS as stars, who are among our author collective.[5] Accounting for the overlap, the final sample consists of 155,720 scientists, which corresponds to 1.7% of the observable scientific community. It should be stated though that the composition of the star sample changes over time. Scientists are referred to as stars as of the year they fulfil a performance criterion or are inducted to the NAS, yet once proven keep their status thereafter.

Only a small circle of the scientific elite is of immediate interest for our research design. To allow for a causal interpretation of spillover effects, we focus on stars whose careers ended abruptly due to unexpected death at a maximum age of 65 years. We identify these cases by inspecting publication histories. Once a star's publication activity falls off rapidly while being at a career age where retirement appears doubtful, we manually search for bibliographic information online. This approach leads to 594 stars that died between 2001 and 2012. After imposing the age constraint, we further exclude scientists whose research efforts already came to a halt before their death and, most importantly, scientists whose passing might have been anticipated from prolonged illnesses. We draw the distinction between unexpected and anticipated deaths primarily based on information provided by obituaries, but also from personally contacting former colleagues in a few unclear cases. Altogether, we end up with 162 deceased stars to constitute the origin of our treatment (see Appendix A.2). From a field perspective, we note that 40 stars belong to life sciences, 44 to health sciences, 54 to physical sciences, and 24 to social sciences. It further becomes apparent that heart attacks and accidents are mentioned most frequently among the treatment cases, while cancer is the dominating cause of death among the (unreported) group of anticipated deaths.

---

[4] We omit the narrow field of multidisciplinary studies, which is not part of either of the main fields, i.e., life sciences, health sciences, physical sciences, and social sciences.

[5] We sort NAS members into our four-field taxonomy based on their affiliated section and the scheme reported in Appendix A.1. NAS sections are thus given priority over our publication-based classification, yet both approaches agree on 3,438 of 3,458 cases.

Variable	P5	P25	P50	Mean	P75	P95	SD
Age at death	37	48	55	53.41	60	64	8.12
Female	0	0	0	0.049	0	0	0.217
U.S. affiliated	0	0	0	0.494	1	1	0.502
No. of distinct co-authors	5	16	42	67.22	92	183	72.99
No. of articles	3	11	21.5	25.32	34	63	18.57
No. of citations	75	214	431.5	801.1	1,024	2,608	972.5
Year of death	2001	2004	2007	2006.6	2009	2012	3.18

Tab. 2.1: SUMMARY STATISTICS ON TREATMENT STARS

Notes: The sample comprises 162 outstanding scientists whose active careers ended abruptly between 2001 and 2012 due to unexpected death at a maximum age of 65 years. All time-varying variables refer to the year preceding the death event. Article, citation, and distinct co-author numbers are aggregated over a prior 5-year span.

Table 2.1 depicts the sample of treatment stars. On average, these stars died at 53.4 years of age. Almost precisely one half was affiliated with a research institution located in the US and the vast majority was male. Female underrepresentation is fairly unsurprising in view of the collective evidence on scientific gender gaps (Ding, Murray, & Stuart, 2006; Shen, 2013).[6] Moreover, star scientists published an average of 25.3 articles, worked with 67.2 different co-authors, and received just over 800 citations over the course of five years prior to their passing.[7]

2.2.3 Matching Approach

Identifying 162 deceased star scientists allows us to circumscribe the treatment group, i.e., their former co-authors. While this task is straightforward, more diligence is needed to find an appropriate control group. On which outcome trajectory would the treatment group be, had they not been exposed to the death of a star collaborator?

One possibility would be to rely on the full population of scientists to derive an answer. However, treated co-authors likely form a positive selection, as star collaborations are not random. Instead, we expect assortative matching by both age and ability, which makes it doubtful to assume that the full population were to provide an accurate projection for the treatment group's outcome path (even conditional on a variety of fixed effects). A second option would be to employ an implicit control group composed of treated co-authors that experience the death at either earlier or later points in time. Yet, this approach could also pose threats to identification if, for instance, the death event leads to a change in outcome

[6] Differences in health status (Williams, 2003) and risk attitude (Hartog, Ferrer-i-Carbonell, & Jonker, 2002) could also play a part. To be clear, a heart attack or stroke might be a sudden and unexpected event, but still dependent on lifestyle factors.

[7] Cumulative figures are calculated over a fixed range to account for staggered death years.

trends. Azoulay et al. (2010) discuss this methodological issue and present a strategy to circumvent it. We follow their example and therefore build our control group based on a one-to-one matching procedure.

Although the matching algorithm iterates over the years from 2001 to 2012, we focus on the year 2005 for illustration purposes. We begin by compiling the pool of potential control scientists, which consists of scientists that essentially meet two conditions. First, as of 2005, they must have collaborated with a star scientist who does not pass away, regardless of cause. Second, at no time do they become co-authors of one of the 162 deceased stars (i.e., they remain spared from treatment). For each treated scientist with an associated death in 2005, we aim to select an appropriate control scientist from the defined pool. In order to be matched, we require that treated and control scientists have similar career ages, are embedded in co-author networks of comparable size, and show congruent outcome trends up to 2004. In addition to individual characteristics, we include further criteria to ensure that both groups are balanced regarding features of their star relationship, i.e., the number of past collaborations and the elapsed time since last collaborating, and their stars' standing as proxied by the amount of citations received until 2004. Before deferring further (technical) details to Appendix A.3, we note that the algorithm is implemented year by year, separately for the four main scientific fields, without replacement, and utilises the idea of coarsened exact matching introduced by Iacus, King, and, Porro (2011, 2012).

Finally, we add two constraints to ensure that we are exploring spillover effects between established scientists. Junior scientists and PhD students, to begin with, might experience a conceptually different treatment effect in the sense that the death of a senior colleague or supervisor could have career-ending consequences. We thus restrict the analysis to (both treated and control) scientists with a career age of at least five years at the time of death. Moreover, we exclude a small number of scientists whose career starts coincide with the beginning of their star collaboration to prevent our results from being intertwined with mentoring effects.[8]

The outlined procedure leads to a set of 9,297 matched collaborator pairs representing a successful matching rate of 93.6%. Summary statistics are reported in Table 2.2. Note that control collaborators inherit the year of star death from their matched counterparts, so that treatment timing is identically distributed in both groups. Time-varying variables are again calculated as of the year preceding the (inherited) year of star death to depict the sample right before treatment onset. Overall, we detect only minor differences between treated and control collaborators. The average treated collaborator published 14.5 articles, received 517 citations, and held co-authorship ties to 65.9 scientists over a past five-year period, while his/her control group pendant recorded 14.1 articles, 503 citations, and 65.2

---

[8] Both Waldinger (2010) and Azoulay, Liu, and Wang (2017) provide insights into this strand of literature.

co-authorship ties. The performance balance is further reflected by the share of stars in both groups – 25.8% of the treated and 25.0% of the control collaborators are considered stars, which indicates that assortative matching influences network formation in science. We also document a close resemblance in career ages, i.e., publication activities in both groups span 18.2 years on average. Achieving a high age balance is clearly important for our research design since scientific output typically follows life cycle patterns (Levin & Stephan, 1991).

Variable	Group	P5	P25	P50	Mean	P75	P95	SD
Career age	Treated	6	11	17	18.24	25	34	8.78
	Control	5	11	17	18.17	25	34	8.88
Female prediction	Treated	0	0	0	0.239	0	1	0.426
	Control	0	0	0	0.256	1	1	0.437
U.S. affiliated	Treated	0	0	0	0.425	1	1	0.494
	Control	0	0	0	0.409	1	1	0.492
Star status	Treated	0	0	0	0.258	1	1	0.438
	Control	0	0	0	0.250	0	1	0.433
No. of distinct co-authors	Treated	5	16	37	65.92	81	227	82.78
	Control	5	17	38	65.20	86	212	76.67
No. of articles	Treated	1	3	8	14.53	18	49	18.92
	Control	1	3	8	14.12	18	48	17.91
No. of citations	Treated	6	58	189	517.0	549	2,053	1,018.7
	Control	4	64	203	503.0	552	1,908	937.6
No. of collaborations	Treated	1	1	1	2.29	2	7	3.87
	Control	1	1	1	2.23	2	7	3.95
Years since last collaboration	Treated	0	1	3	4.04	6	11	3.61
	Control	0	1	3	3.96	6	11	3.53
No. of citations (star)	Treated	202	517	1,133	1,642.3	2,178	5,839	1,546.9
	Control	158	512	1,033	1,495.5	1,901	4,449	1,613.8

Tab. 2.2: SUMMARY STATISTICS ON MATCHED COLLABORATORS

Notes: The sample consists of 9,297 pairs of treated and control collaborators. All time-varying variables refer to the year preceding the (inherited) year of star death. Article, citation, and distinct co-author numbers are aggregated over a prior 5-year span. Gender information are inferred through name and country data and are available for 85.3% of the sample.

Turning to the dyadic variables, we first note that the mean number of collaborations (i.e., jointly published articles) between stars and co-authors amounts to 2.29 in the treated and to 2.23 in the control group. However, and to some degree surprising, the median dyad in both groups denotes only one collaboration. Moreover, an average of 4.04 years passed since stars and treated co-authors last collaborated, while 3.96 years elapsed in the control group. The reported time gaps in Table 2.2 are indeed long enough to assume that neither the average nor the median co-author was engaged in ongoing research projects with their star at the time of death. Another observation is related to the star scientists' standing. In particular, we find both treated and control stars to receive more citations than the initial star sample portrayed in Table 2.1. What might seem striking at first glance is merely due to a (harmless) selection effect. Deceased stars with greater citation numbers usually record both higher article and co-author numbers, which causes them to appear more frequently in the matched sample. Treated stars are slightly more accomplished than control stars, but the magnitude is not concerning.[9]

We further achieve balance on two variables that were not part of the matching process.[10] First, about one quarter of the collaborator sample is predicted to be female. Neither does Scopus provide gender information nor is it feasible to collect these data manually (as we did for deceased stars). For these reasons, we rely on the gender inference by GenderAPI, which has been found to offer the most accurate application for this task (Santamaría & Mihaljević, 2018). In essence, gender data are inferred from first names, optionally in combination with a country information. While our Scopus data cover first names, there is no direct country indication. We thus derive a home country proxy from the affiliations listed on the earliest publication records. In sum, we hereby manage to classify 85% of our sample.[11] Probing the validity of this approach, we find gender predictions to be correct for over 99% of the full sample of deceased stars. Second, we observe a little over 40% of the collaborators to be US-affiliated. Again, we denote slight uncertainty regarding this number, as some collaborators are linked with multiple affiliations as of their most recent publications (note that our data do not include within-year publication dates). In these instances, we infer a collaborator's location based on his/her mode affiliation(s), breaking possible ties at random.

In view of our field-specific estimations in Section 2.4, we lastly note that the described matching approach also creates balance between treated and control groups if the sample is split by fields (see Appendix A.4 for corresponding summary statistics).

---

[9] Besides, we would expect that the benefits of having a star collaborator increase with his/her standing. Building a control group with stars that received fewer citations than their treated counterparts should therefore rather serve as a conservative estimation approach.

[10] The number of matching variables is limited due to the curse of dimensionality. In other words, it would become considerably more difficult to find matches if we extended our variable set any further.

[11] We started by querying first names only and considered gender predictions valid if GenderAPI reported an accuracy of over 98%. In a second step, we used first names combined with the home country proxy to classify the remaining collaborators, hereby setting the accuracy threshold to 95%.

2.3 Identification of Main Effects

2.3.1 Outcome Paths

To set the stage for the DiD framework, we begin with a purely graphical illustration of the treatment impact. To be more precise, we plot the publication output of treated and (matched) control collaborators before and after the star scientists' death. This approach gives a basic yet compelling impression of the stars' influence on the outcome trajectory of their co-authors without the need for parametric assumptions.

Figure 2.1 displays the output trends centred symmetrically around the time of death.[12] We will confine our assessment of publication output to two main measures, i.e., article count and forward citations, both adjusted for co-authorships. As depicted in the upper panel of Figure 2.1, treated and control collaborators show hardly any difference in article counts before the year of star death. On average, both groups vary synchronically between 0.50 and 0.55 annual articles. After the treatment, however, an evident gap emerges in favour of the group of scientists that does not experience the sudden passing of an outstanding co-author. The relative performance deficit of the treatment group is apparent in every year after the death event and thus, albeit slight variations in magnitude, permanent. The graphic further underlines the importance of the matching design. As can be seen, article counts tend to rise over time, even for the treatment group, which could be reflective of life cycle and/or year fixed effects. In absence of the counterfactual output path provided by the matched collaborator sample, it would remain ambiguous how to disentangle these effects from the actual treatment effect.

It is conceivable that collaborators adjust to the treatment shock by raising their effort devoted to each published paper. In this scenario, scientists would (try to) maintain their overall quality of output despite experiencing a decline in productivity in form of lower article counts. We explore this possibility by plotting citation-weighted article counts, i.e., forward citations, in the lower panel of Figure 2.1. This measure provides a common proxy for scientific quality (see e.g., Jaravel et al., 2018, or Kahn and MacGarvie, 2016).[13] At first sight, we note that both groups show decreasing forward citation trends, which can be

---

[12] The number of yearly observations monotonically decreases as the temporal distance to the year of death increases, which can lead to imprecisely estimated effects at both ends of the observation period. Uncertainty in later years can also arise from collaborators becoming inactive, likely due to retirement. In line with Jaravel et al. (2018), we address these concerns by confining observations to a nine-year window around star death and by excluding observations if collaborators exceed a career age of 45 years. Note that we apply the age constraint simultaneously to each matched pair to ensure that treated and control collaborators keep their balance in calendar and experimental time.

[13] Following the cited literature, we employ winsorized forward citations. We apply this adjustment at the 99.9th percentile, separately for each year and scientific field (life, health, physical, and social sciences). Robustness checks in Appendix A.5 show that our results do not rely on winsorizing.

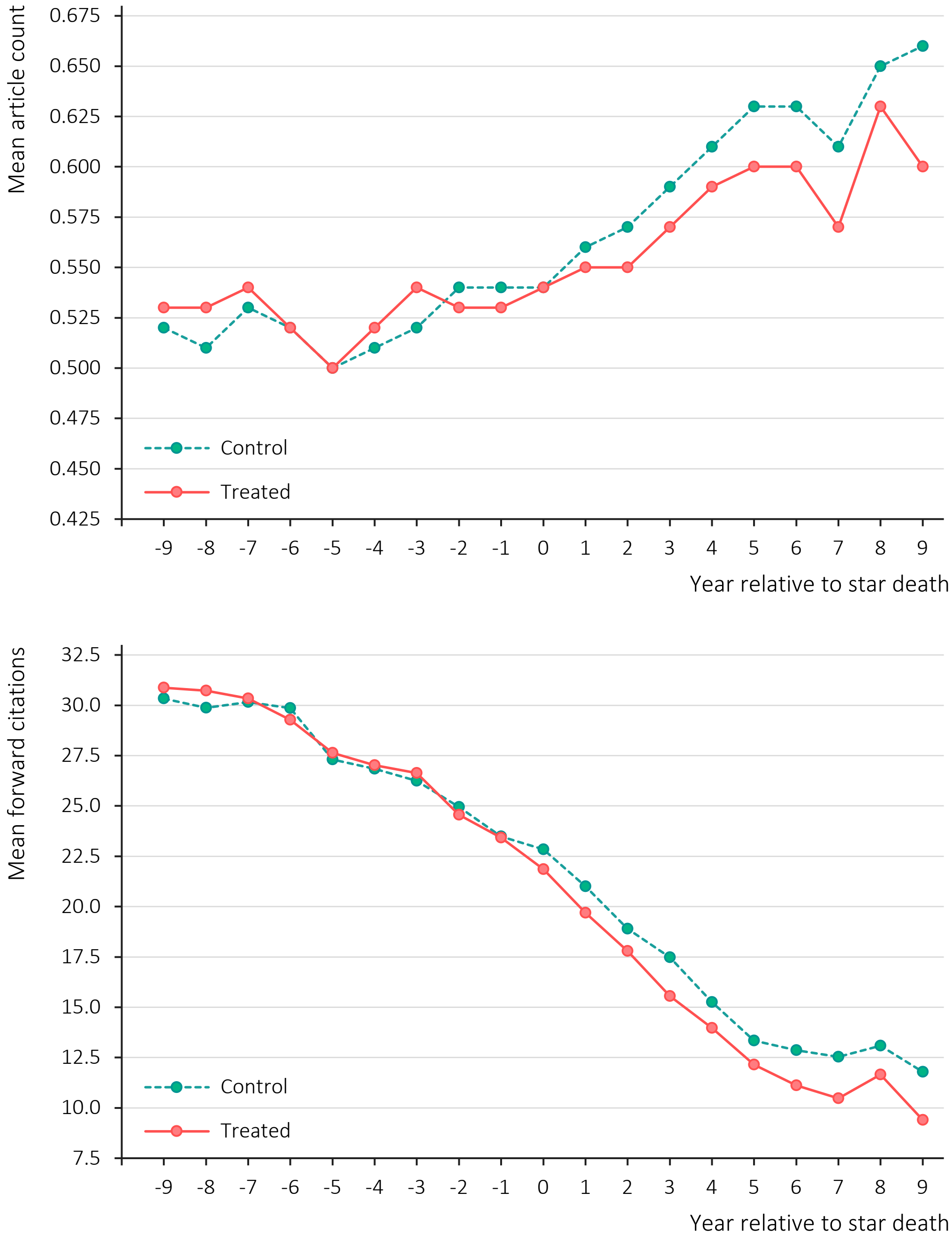


Fig. 2.1: OUTCOME PATHS AROUND STAR DEATH

Notes: The sample consists of 9,297 pairs of treated and control collaborators confined to a nine-year window around the (inherited) year of star death. The panel is equally unbalanced for treated and control collaborators. Both output measures are co-author adjusted. Forward citations are winsorized at the 99.9th percentile.

attributed to the truncated nature of the variable. Stated differently, forward citations represent the total number of citations received as of 2016 by all articles published in a given year, which makes high numbers at the end of the observation period less likely. However, this mechanical effect can be neglected since treatment and control scientists cover the exact same time spans. More importantly, both groups closely resemble each other until the year of death, yet treated collaborators are again outperformed by their matched counterparts in all subsequent years.

The presented evidence suggests that scientists suffer in the realms of both productivity and quality after the abrupt end of a star collaboration. Figure 2.1 is indeed a (raw) preview of our main results that further underlines the effectiveness of our matching approach by visually confirming the parallel course of pre-trends. We evaluate these findings in more econometric detail over the next subsections.

2.3.2 Econometric Model

We apply a straightforward econometric methodology that has been employed in related contexts (Azoulay, Fons-Rosen, & Graff Zivin, 2019; Jäger & Heining, 2019; Jaravel et al., 2018). Assured through the matching procedure, treated scientists are paired with control scientists that possess a multitude of similar characteristics. Moreover, matched scientists are also temporally aligned, implying that each control scientist inherits a counterfactual death year from his/her treated counterpart. These design properties allow us to estimate a dynamic DiD equation, where the causal effect of star death is identified through yearly differences in the research output of both groups (adjusted for a range of fixed effects). Our econometric approach takes the following form:

$$Y_{it} = exp\left[\alpha + \sum_{k=-9}^{9} \beta_k^{All}\, \mathbb{1}\,(L_{it} = k) + \sum_{k=-9}^{9} \beta_k^{Real}\, \mathbb{1}\,(L_{it} = k) \times Treated_i \right. \qquad (2.1)$$
$$\left. + \vartheta_{it} + \delta_t + \gamma_i + \epsilon_{it}\right],$$

where Y_{it} denotes either the article count or forward citations of co-author i in calendar year t. Both dependent variables are bound by a considerable fraction of zero values. We therefore estimate Equation (2.1) by means of Poisson pseudo-maximum likelihood (PPML) techniques. Apart from handling the skewed, non-negative distribution of the dependent variables, the PPML estimator offers compelling robustness properties. Importantly, it can be ensured that coefficient estimates are consistent as long as the conditional mean of the dependent variable is correctly specified (Gourieroux, Monfort, & Trognon, 1984). The data generating process is thus not required to be Poisson. In addition, employing robust standard errors, clustered at the star level in our application, allows for correct inference irrespective of any form of serial correlation (Wooldridge, 1997).[14]

We address the staggered treatment onset by including lead and lag terms, denoted by L_{it}, in Equation (2.1). Each of these terms represents an indicator variable that switches to

---

[14] For the estimation in Stata, we employ the ppmlhdfe command by Correira, Guimarães, and Zylkin (2019), which implements PPML regressions with multiple high-dimensional fixed effects. In contrast to conventional commands, ppmlhdfe proves robust to typical convergence issues in Poisson contexts.

1 if an observation is k years apart from the death event. As shown by Jaravel et al. (2018), the first set of leads and lags, whose effects will be identified by the β_k^{All} coefficients, fulfils a role similar to the post dummy in classic DiD frameworks. Its practical relevance stems from the concern that career age fixed effects (ϑ_{it}), calendar year fixed effects (δ_t), and individual fixed effects (γ_i) may not entirely capture trends in productivity or research quality around the time of star death.[15] One possible cause for such trends could refer to the sample construction, where we condition on star collaboration, which could coincide with unobservable factors that may change regardless of the star's passing (e.g., funding outlooks or work environments). Any of these transitory processes are absorbed by the common lead and lag terms. The second set of leads and lags, which is interacted with the indicator variable for treatment status, $Treated_i$, therefore isolates the causal treatment effect. We split the overall effect into yearly elements, each of which will be identified by their respective β_k^{Real} coefficient.

The key identifying assumption of our model is that star deaths are exogenous conditional on the covariates in Equation (2.1), which implies that treated and control scientists would have developed parallel output paths if the death event had not occurred. Ensuring this assumption motivates our research design, which builds on manually screened obituaries and a thorough matching procedure. While it is not possible to verify the parallel trends assumption post-treatment, its validity can be bolstered by means of pre-treatment data. Specifically, Equation (2.1) enables testing if death events are accompanied by preceding effect patterns, which would render the analysis doubtful. Apart from that, decomposing the effect post-death allows us to explore treatment consequences in dynamic fashion. We present our estimation results in the following subsection and note that any of these estimates can be interpreted as semi-elasticities after coefficients are exponentiated and decreased by one.

2.3.3 Results

Figure 2.2 provides a graphical depiction of the annual β_k^{Real} coefficients by plotting point estimates along with 95% confidence intervals derived from Equation (2.1). The upper panel depicts the treatment dynamics in terms of article counts, while the lower panel refers to forward citations. Technically, the point estimate that corresponds to the year preceding the treatment year is normalised to zero, implying that this lead marks the reference point for the presented effects.

[15] Career age fixed effects account for output shifts over the course of a scientist's career, while calendar year fixed effects capture all time-related influence factors such as the expansion of academic journals. Finally, individual fixed effects control for variation that originates from characteristics that are constant across individual scientists, e.g., innate ability, but also cover time-invariant dyadic features as, e.g., the age gap between stars and collaborators. Given that the three classes of fixed effects induce collinearity, we omit two (out of 45) career age fixed effects, which is standard practice (Jaravel et al., 2018).

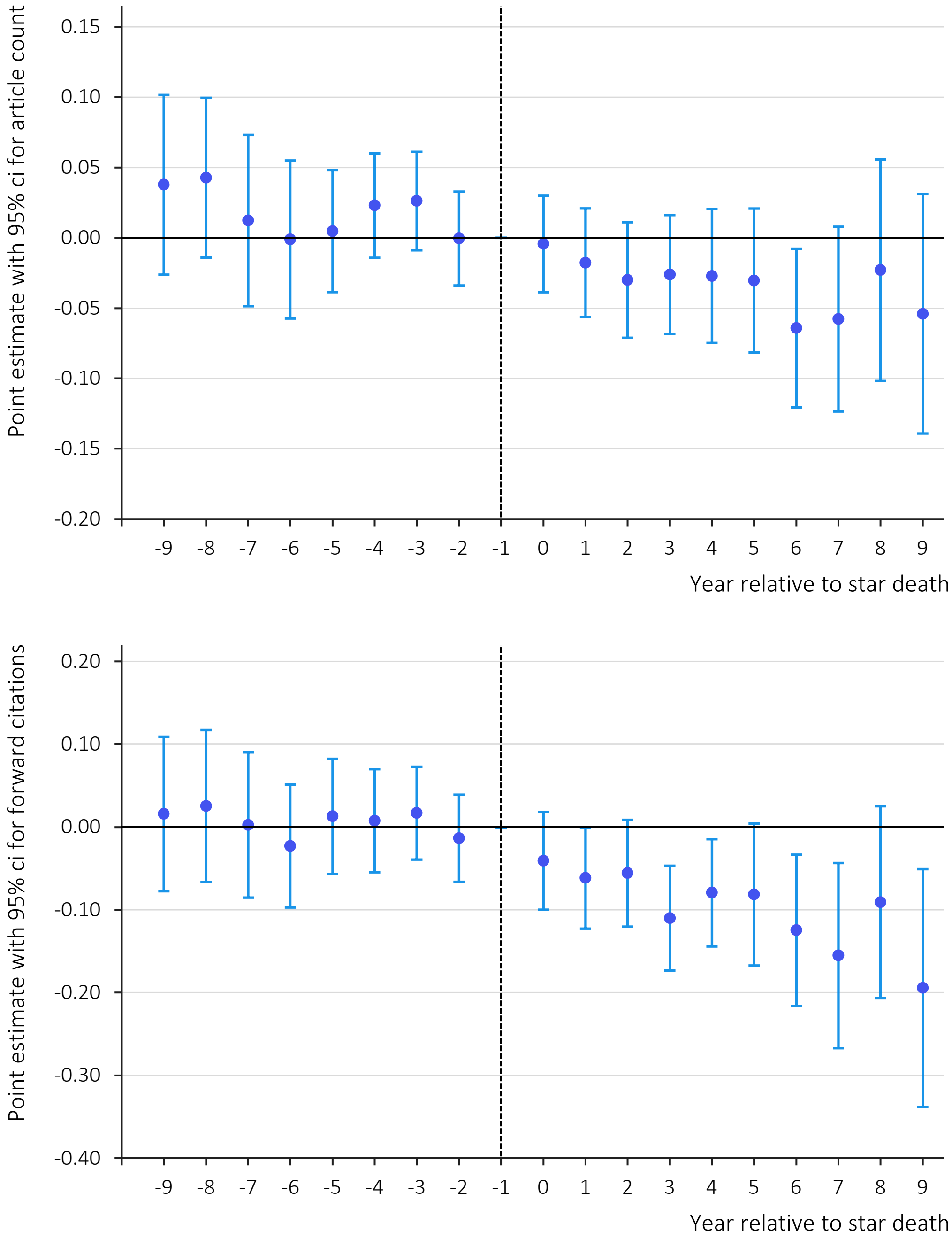


Fig. 2.2: TREATMENT EFFECT DYNAMICS

Notes: The sample consists of 9,297 pairs of treated and control collaborators confined to a nine-year window around the (inherited) year of star death. The panel is equally unbalanced for treated and control collaborators. Both output measures are co-author adjusted. Forward citations are winsorized at the 99.9th percentile. Collaborators without variation in output over the observation period are dropped by the estimation routine. For the econometric approach, refer to Section 2.3.2.

In view of the effect patterns for article counts, we first note the absence of pre-trends. While most point estimates leading up to the treatment year are slightly positive, neither of them is statistically significant, which is in line with the non-parametric résumé. After the death event, we notice a gradual shift in point estimates, which turn consistently negative. The productivity decline induced by the treatment shock appears to increase in the long run, but the picture is not entirely conclusive. The 6th lag is associated with a statistically significant effect that translates into a 6.2% reduction ($exp\,[-0.064]-1$) in article counts, but the remaining lags are smaller in magnitude and not statistically significant. Despite statistical uncertainty on the annual level, the aggregate perspective clearly indicates that the unexpected passing of a star leads to a moderately diminished productivity for co-authors without signs of a rebound effect.

As for forward citations, our proxy for output quality, we discover a broadly comparable picture to the article count analysis but with amplified effect magnitudes. Again, our research design finds support through insignificant point estimates for all leads, which underlines parallel pre-treatment trends. After the treatment, however, point estimates markedly decrease, implying that the stars' death puts collaborators on career paths with less impactful publications. In six out of nine post-treatment years, we estimate a statistically significant decline in forward citations. Reduced output quantity could play into this finding, but the absolute effect sizes are notably higher. In fact, they tend to rise over time, peaking in the 9th year where the treatment effect equates to a 17.7% decrease in forward citations. This again illustrates that the star loss unfolds long-term consequences that transcend the mere disruption of ongoing projects. From comparing both panels of Figure 2.2, it can be inferred that the treatment impact becomes more pronounced when output quality is taken into consideration.

2.4 Variations over the Scientific Spectrum

2.4.1 Main Field Effects

The identification of treatment effects over the complete sample sets the baseline for our next analysis steps, in which we exploit the rich diversity of our data. An integral part of the upcoming investigation is to compare effects across the scientific spectrum. We start with field-specific treatment effects, derived from the following specification:

$$Y_{it} = exp\,\Big[\alpha + \beta^{All} AfterDeath_{it} + \beta^{Real} AfterDeath_{it} \times Treated_i + \vartheta_{it} + \delta_t + \gamma_i + \epsilon_{it}\Big], \tag{2.2}$$

which mirrors Equation (2.1) with the exception of the $AfterDeath_{it}$ variable that takes the place of the former lead and lag terms. $AfterDeath_{it}$ denotes an indicator variable that switches to 1 in the year of star death. Its interaction with the $Treated_i$ variable allows us to determine the treatment effect in a time-averaged form, i.e., pooled over all leads and lags. The β^{Real} coefficient will identify this effect, while the β^{All} coefficient will capture all side effects that relate to the treatment timing but not the actual event. The advantage of Equation (2.2) lies in the ease of discussing total effect magnitudes but also in the improved statistical power. The latter aspect is particularly relevant for estimations on smaller datasets, which applies to the present setting, where the overall sample will be split according to the stars' field classification. Equation (2.2) will therefore be estimated separately for the four fields of life, health, physical, and social sciences, although we will also report the outcome of a pooled estimation, which corresponds to our main (dynamic) results. Apart from that, we adopt the inclusion of fixed effects, the level of standard error clustering, and the use of the PPML estimator from Equation (2.1).

	Overall Sample	Life Sciences	Health Sciences	Physical Sciences	Social Sciences
	Article count as dependent variable				
After death × treated	-0.043 *	-0.066 *	-0.044	-0.026	-0.062
	(0.022)	(0.030)	(0.034)	(0.030)	(0.151)
Log pseudo-likelihood	-189,139	-52,754	-79,330	-53,894	-3,021
No. of observations	275,344	83,541	118,212	69,107	4,484
No. of dyads	18,542	5,585	7,940	4,711	306
	Forward citations as dependent variable				
After death × treated	-0.081 **	-0.114 **	-0.043	-0.104 *	-0.015
	(0.028)	(0.041)	(0.038)	(0.043)	(0.205)
Log pseudo-likelihood	-2,800,261	-809,005	-1,154,988	-784,720	-40,122
No. of observations	275,166	83,526	118,148	69,008	4,484
No. of dyads	18,527	5,584	7,934	4,703	306

Tab. 2.3: IMPACT OF STAR DEATH ON COLLABORATORS' OUTPUT

Notes: The overall sample consists of 9,297 pairs of treated and control collaborators confined to a nine-year window around the (inherited) year of star death. The panel is equally unbalanced for treated and control collaborators. Both output measures are co-author adjusted. Forward citations are winsorized at the 99.9th percentile. Collaborators without variation in output over the observation period are dropped by the estimation routine. Field delineation is based on the stars' publication profile or, if available, derived from the classification of the NAS. Robust standard errors are in parentheses, clustered at the level of the star.

* $p < 0.05$, ** $p < 0.01$, and *** $p < 0.001$.

Table 2.3 depicts the results derived by means of Equation (2.2). From the upper panel, it becomes apparent that, regarding the overall sample, the death of a star affects article counts to a statistically significant extent. The effect equates to a 4.2% decline, which represents the pooled counterpart of the dynamic effects that are displayed in Figure 2.2 (upper panel). However, a closer look at the single fields reveals that this productivity shock can only be confirmed for life sciences, where treated collaborators face an even stronger drop of 6.4%. As for the other fields, negative effects may be measured, but point estimates do not reach statistical significance at conventional levels. Turning to the lower panel of Table 2.3, we find treatment effects to increase once publication quality is factored in. Overall, co-authors experience a statistically significant reduction of 7.8% in forward citations. This effect, again, becomes more pronounced in both magnitude and statistical significance for life sciences dyads, where the loss of an eminent scientist is followed by a 10.8% reduction. In case of forward citations, a similar observation can be made for physical sciences, where a 9.9% deficit is identified. Yet, in accordance with the article count results, no evident effects can be stated for the fields of both health and social sciences.[16]

	Overall Sample	Life Sciences	Health Sciences	Physical Sciences	Social Sciences
	Article count excl. star collaboration as dependent variable				
After death × treated	-0.026	-0.051	-0.028	-0.007	-0.054
	(0.022)	(0.032)	(0.032)	(0.032)	(0.160)
Log pseudo-likelihood	-181,703	-50,560	-76,254	-51,936	-2,808
No. of observations	268,794	81,844	115,561	67,138	4,251
No. of dyads	18,000	5,446	7,717	4,549	288
	Forward citations excl. star collaboration as dependent variable				
After death × treated	-0.062 **	-0.112 **	-0.017	-0.080	0.021
	(0.028)	(0.040)	(0.035)	(0.042)	(0.224)
Log pseudo-likelihood	-2,630,686	-760,591	-1,088,715	-735,906	-34,288
No. of observations	268,314	81,780	115,383	66,947	4,204
No. of dyads	17,961	5,440	7,702	4,534	285

Tab. 2.4: IMPACT OF STAR DEATH ON COLLABORATORS' OUTPUT BEYOND JOINT PRODUCTION

Notes: The overall sample consists of 9,297 pairs of treated and control collaborators confined to a nine-year window around the (inherited) year of star death. The panel is equally unbalanced for treated and control collaborators. Both output measures are co-author adjusted. Forward citations are winsorized at the 99.9th percentile. Collaborators without variation in output over the observation period are dropped by the estimation routine. In comparison to Table 2.3, this also applies to collaborators that solely published together with their star. Field delineation is based on the stars' publication profile or, if available, derived from the classification of the NAS. Robust standard errors are in parentheses, clustered at the level of the star.

* $p < 0.05$, ** $p < 0.01$, and *** $p < 0.001$.

A first hypothesis as to what could drive the effects reported in Table 2.3 concerns the role of collaborative output. Naturally, the death of a star scientist renders future cooperation impossible. If we removed this portion from the control collaborators' publication résumé, how would the assessment of treatment consequences change? We explore this question in Table 2.4. Technically, we use modified dependent variables that solely comprise articles

[16] Based on similar research designs, Azoulay et al. (2010), Oettl (2012), and Mohnen (2018) document declines of 8.2%, 12.4%, and 14.3% in impact factor weighted publication counts, respectively, whereas Jaravel et al. (2018) report a 15.6% drop in forward citations of patents.

that were not co-authored by the star. As can be seen from Table 2.4, treatment effects are less accentuated in this scenario. What might have been expected becomes particularly visible in case of articles counts (upper panel) where point estimates derived from neither life sciences nor the overall sample remain statistically significant. The productivity decrease in Table 2.3 can therefore largely be attributed to the unrealised potential of joint work.[17] However, repeating the analysis with forward citations (lower panel) leads to a different conclusion. With regard to the overall sample, the point estimate slightly decreases, but the effect stays statistically significant. Physical sciences adjusts in a comparable manner, although the effect lies at the margin of significance (p-value of 0.056); and life sciences remains virtually unaffected. In summary, control scientists are thus found to accumulate more forward citations than treated scientists do, even after subtracting co-publications with their star. Importantly, this finding illustrates that the sudden death of a star clearly unfolds consequences that span beyond the omission of joint work.

In Appendix A.5, we present a series of robustness checks that result from modifying Equation (2.2). First, we technically delay the beginning of the after-death period by one year. Including the year of death into the pre-death period could be justified on grounds of publication lags or if death events occur towards the end of the year. Strictly speaking, the death year can be considered a transition year, where the treatment consequences start to emerge. Second, we follow Azoulay et al. (2010) and Oettl (2012) by capturing life cycle patterns with career age cohort dummies, which could mitigate collinearity concerns between year, age, and individual fixed effects. Third, we extend our fixed effects arsenal by including interacted calendar year and career age fixed effects, thereby probing the implicit separability assumption in Equation (2.2). Fourth, we explore if clustering standard errors at the collaborator level instead of the star level affects our results. Fifth, we re-estimate treatment effects on a (substantially) shortened panel of collaborators that are traceable for a full seven years before and after the death year. Using a balanced panel addresses the concern that collaborators with a surplus of either pre- or post-treatment observations might have a confounding influence on the estimation of true effects. Sixth, we employ forward citations without winsorizing. Taken together, we detect only minor changes in our results due to these alterations. In health and physical sciences, we both note one instance with a statistically significant article count effect, but these singular findings may not be overstated. Importantly, it can be confirmed that the main effects reported in Table 2.3 prove robust to a range of different model specifications.

---

[17] To be clear, the results in Table 2.4 do not imply that treatment effects are non-existent. They rather show how effects shift if joint work is taken out of the equation. In this setting, control collaborators are mainly penalised (post-death), although delayed publications are also removed for treated collaborators.

2.4.2 Distinct Effect Channels

The death of an outstanding scientist affects the performance of former co-authors to an appreciable extent. The documented effects are mainly driven by collaborations in the fields of life and physical sciences and in part, but by no means fully, explainable by the deprivation of future cooperation. Within this section, we aim for a deeper understanding of the effect formation. If we were to determine subgroups of the treated scientists that experience the star death to a particularly great extent, we would have strong evidence for the origins of the treatment effect. Stated differently, where does the star's death leave its primary mark? We explore heterogeneity in the treatment effect employing the following estimation equation:

$$\begin{aligned} Y_{it} = exp\ \Big[&\alpha + \beta^{All} AfterDeath_{it} + \beta^{Real} AfterDeath_{it} \times Treated_i \\ &+ \eta^{All}\, \mathrm{Z}_i \times AfterDeath_{it} + \eta^{Real}\, \mathrm{Z}_i \times AfterDeath_{it} \times Treated_i \\ &+ \vartheta_{it} + \delta_t + \gamma_i + \epsilon_{it}\Big], \end{aligned} \tag{2.3}$$

where Z_i constitutes a time-invariant indicator variable, which we expect to be insightful for the magnitude of the treatment effect. To be clear, Z_i will vary over the course of the analysis and delineate different sets of collaborations based on either individual or dyadic characteristics. The overall treatment effect will, according to this distinction, be divided into a common (β^{Real}) and a specific (η^{Real}) component. The coefficient of interest in this setting becomes η^{Real}, which isolates the differential treatment effect that is additionally yet exclusively felt by the delineated group of collaborators. Consistent with our former models, we incorporate Z_i not only as part of an interaction for treated dyads, but also within a second interaction, which is common to all dyads and thus accounts for general outcome shifts that are attributable to Z_i. All further estimation aspects of Equation (2.1) and (2.2) remain unchanged, as does our strategy to distinguish between scientific fields.

We first direct attention to collaborative features, which could play a moderating role. Intuitively, the assumption would be that scientists that maintained an intensive work relation with their star experience more severe treatment consequences than sporadic dyads. Two reasons lend support for this claim. First, co-authorships are not randomly assigned. Instead, they are more likely to result from a thorough matching process. Collaborations that turn out to be fruitful should thus embody higher chances of being continued. Second, even if we overstated the freedom in choosing co-authors and took potential lock-in effects into consideration (Boudreau et al., 2017), one might still expect repeated collaborations to be more valuable through accumulating team-specific capital (Jaravel et al., 2018). However, there are opposing arguments to be raised too. Notably, upholding a star collaboration could have benefits, e.g., in form of acquired knowledge

	Overall Sample	Life Sciences	Health Sciences	Physical Sciences	Social Sciences
	Article count as dependent variable				
After death × treated	-0.047 *	-0.054	-0.053	-0.029	-0.185
	(0.022)	(0.030)	(0.034)	(0.035)	(0.182)
After death × treated × dyad frequency in 3. tertile	0.015	-0.039	0.036	0.015	0.349
	(0.032)	(0.048)	(0.053)	(0.059)	(0.278)
Log pseudo-likelihood	-189,136	-52,753	-79,329	-53,892	-3,019
No. of observations	275,344	83,541	118,212	69,107	4,484
No. of dyads	18,542	5,585	7,940	4,711	306
	Forward citations as dependent variable				
After death × treated	-0.083 **	-0.089 *	-0.066	-0.098 *	-0.172
	(0.027)	(0.044)	(0.038)	(0.044)	(0.253)
After death × treated × dyad frequency in 3. tertile	0.008	-0.080	0.090	-0.019	0.438
	(0.048)	(0.069)	(0.066)	(0.096)	(0.404)
Log pseudo-likelihood	-2,800,245	-808,892	-1,154,831	-784,652	-40,019
No. of observations	275,166	83,526	118,148	69,008	4,484
No. of dyads	18,527	5,584	7,934	4,703	306

Tab. 2.5: EFFECT HETEROGENEITY BY COLLABORATION FREQUENCY

Notes: The overall sample consists of 9,297 pairs of treated and control collaborators confined to a nine-year window around the (inherited) year of star death. The panel is equally unbalanced for treated and control collaborators. Both output measures are co-author adjusted. Forward citations are winsorized at the 99.9th percentile. Collaborators without variation in output over the observation period are dropped by the estimation routine. Field delineation is based on the stars' publication profile or, if available, derived from the classification of the NAS. Robust standard errors are in parentheses, clustered at the level of the star.
* p < 0.05, ** p < 0.01, and *** p < 0.001.

or access to superior networks, that increase outside options and eventually allow for an easier transitioning towards new collaborations. Turning to our empirical assessment in Table 2.5, we can infer that frequent collaborators, defined as those that belong to the upper third of the distribution of co-authorships with their respective star, do not suffer a treatment effect of a markedly different size than the remaining collaborators.[18] The effect for the former group, with regard to the overall sample, corresponds to a drop of 3.1% ($exp\,[-0.047 + 0.015] - 1$) in article counts and 7.2% in forward citations, while the latter group experiences a decline of 4.6% ($exp\,[-0.047] - 1$) in article counts and 8.0% in forward citations. Importantly, these deviations are not statistically significant,

[18] Analogous to the matching approach, we calculate separate distributions for each treatment year and each scientific field (derived from the stars' classification). Descriptively, frequent collaborators published a mean number of 5.8 joint articles with their star. However, due to the large amount of one-time dyads, frequent collaborators are oftentimes synonymous with repeated collaborators.

neither overall nor in any single field. Additionally, looking into recent collaborations yields a similar conclusion, as does repeating the analysis with multi-year collaborations (see Appendix A.6). In sum, we find no evidence that treatment effects depend on any of these basic interaction features.

	Overall Sample	Life Sciences	Health Sciences	Physical Sciences	Social Sciences
	Article count as dependent variable				
After death × treated	-0.045	-0.058	-0.049	-0.038	0.010
	(0.026)	(0.035)	(0.050)	(0.031)	(0.144)
After death × treated × star wrote editorial	-0.005	-0.029	0.004	0.083	-0.254
	(0.049)	(0.069)	(0.069)	(0.086)	(0.420)
Log pseudo-likelihood	-189,128	-52,754	-79,326	-53,890	-3,018
No. of observations	275,344	83,541	118,212	69,107	4,484
No. of dyads	18,542	5,585	7,940	4,711	306
	Forward citations as dependent variable				
After death × treated	-0.072 *	-0.119 **	-0.020	-0.105 *	0.032
	(0.033)	(0.050)	(0.054)	(0.046)	(0.205)
After death × treated × star wrote editorial	-0.026	0.045	-0.059	0.007	-0.217
	(0.062)	(0.085)	(0.075)	(0.112)	(0.588)
Log pseudo-likelihood	2,800,201	-808,972	1,154,930	-784,554	-40,083
No. of observations	275,166	83,526	118,148	69,008	4,484
No. of dyads	18,527	5,584	7,934	4,703	306

Tab. 2.6: EFFECT HETEROGENEITY BY EDITORIAL INFLUENCE

Notes: The overall sample consists of 9,297 pairs of treated and control collaborators confined to a nine-year window around the (inherited) year of star death. The panel is equally unbalanced for treated and control collaborators. Both output measures are co-author adjusted. Forward citations are winsorized at the 99.9th percentile. Collaborators without variation in output over the observation period are dropped by the estimation routine. Field delineation is based on the stars' publication profile or, if available, derived from the classification of the NAS. Robust standard errors are in parentheses, clustered at the level of the star.
* $p < 0.05$, ** $p < 0.01$, and *** $p < 0.001$.

We proceed by ruling out a mechanism that would paint a less meritocratic picture of the scientific community. In particular, we examine if stars exercise a gatekeeping role, thereby elevating the career paths of their collaborators. If this believe turned out to be true, one should have less faith in fair academic assessment and instead devote more emphasis into forming profitable social ties. Our approach to test this assumption relies on data about editorials. From inspecting publication histories, we find that almost a quarter of all stars in both groups published at least one editorial over the course of five years before the year of death. However, as reported in Table 2.6, there is no indication that editorial goodwill offers an explanation for the treatment effect. To be more precise, co-authors of star scientists with editorial linkage are not subject to a differential effect that approaches statistical significance.[19] A comparable conclusion is indeed derived by Azoulay et al. (2010) who reject the gatekeeping hypothesis from a monetary angle, i.e., influence over the funding apparatus of the National Institutes of Health does not cause effect variations in their study of US life scientists.

While control over journal resources is apparently not a driving force, we do discover local resources to be in part meaningful. Leaning on Azoulay et al. (2010), we base our reasoning on geographical proximity. We pursue an analogous path as in Section 2.2.3 and first assign scientists to institutions as of their most recent publications prior to the treatment. In a second step, we query address data for these institutions from Scopus and third extend them with geographical data from Google Maps. This ultimately enables us to encircle collaborations that were co-located at the time of star death. We refer to dyads as co-located if both scientists were located in the same city. Accordingly, we do not require them to be linked to the same affiliation, in part because Scopus, in some instances, masks (parent) institutions by distinguishing between their sub-entities, which would add noise to this classification. Besides, relying on the city-oriented definition does take into account that a localised dimension of the treatment effect could encompass shared infrastructure facilities (e.g., large computing centres, telescopes, or laboratories). Empirically, we find that co-located dyads represent slightly over one fifth of both the treated and control sample. Furthermore, we detect a statistically significant interaction effect in the productivity sphere of physical sciences, which implies a decline of 13.0% in article counts, in addition to the negligible common treatment effect, for co-located collaborators following the death event (see Table 2.7).[20] We interpret this geographically confined component of the treatment effect as a general reflection of the stars' role in governing research environments. To illustrate this point, one might think of preferential access to expensive or highly-specialised equipment that could be at the star's disposal and may be of particular importance in physical sciences (as conjectured by Azoulay et al., 2019).

---

[19] Colussi (2018) underlines the benefits of being connected to editors of leading economics journals. While we are not able to confirm this result in our setting, it might be interesting to note that the differential treatment impact is largest among social scientists, although very imprecisely estimated.

[20] From a technical standpoint, one might recall that no aggregate effect was found for this field, which, however, does not preclude the possibility of nuanced effects, as presented here. Further examples in relation to health sciences follow over the course of this section.

	Overall Sample	Life Sciences	Health Sciences	Physical Sciences	Social Sciences
	Article count as dependent variable				
After death × treated	-0.029	-0.069 *	-0.034	-0.003	0.009
	(0.023)	(0.030)	(0.037)	(0.032)	(0.179)
After death × treated × co-located	-0.069 *	0.015	-0,050	-0.139 *	-0.433
	(0.033)	(0.072)	(0.044)	(0.064)	(0.310)
Log pseudo-likelihood	-188,788	-52,675	-79,144	-53,834	-2,987
No. of observations	274,210	83,314	117,559	68,907	4,430
No. of dyads	18,461	5,569	7,892	4,698	302
	Forward citations as dependent variable				
After death × treated	-0.072 *	-0.139 ***	-0.026	-0.082	0.109
	(0.031)	(0.042)	(0.043)	(0.044)	(0.250)
After death × treated × co-located	-0.047	0.112	-0.080	-0.136	-0.647
	(0.053)	(0.083)	(0.081)	(0.103)	(0.397)
Log pseudo-likelihood	-2,790,989	-806,053	-1,150,133	-783,394	-39,425
No. of observations	274,032	83,229	117,495	68,808	4,430
No. of dyads	18,446	5,568	7,886	4,690	302

Tab. 2.7: EFFECT HETEROGENEITY IN GEOGRAPHICAL SPACE, CO-LOCATION CHANNEL

Notes: The overall sample consists of 9,297 pairs of treated and control collaborators confined to a nine-year window around the (inherited) year of star death. The panel is equally unbalanced for treated and control collaborators. Both output measures are co-author adjusted. Forward citations are winsorized at the 99.9th percentile. Collaborators without variation in output over the observation period are dropped by the estimation routine. Field delineation is based on the stars' publication profile or, if available, derived from the classification of the NAS. Robust standard errors are in parentheses, clustered at the level of the star.

* $p < 0.05$, ** $p < 0.01$, and *** $p < 0.001$.

Co-location sheds some light on the treatment effect origin but does not deliver a full explanation on its own. We thus turn to a distinct mechanisms class that emphasises stars as being sources of unique knowledge and skills. After the treatment, collaborators might prove incapable of filling the void that star scientists left behind, indicating that parts of their expertise might die with them. The permanent nature of this loss could explain the long-term impact revealed in Figure 2.2. In exploring this hypothesis, we draw on the literature that examines technological distance between firms based on patent data (e.g., Ahuja, 2000, or Rosenkopf and Almeida, 2003). We adapt the methodology to our case and employ publications, instead of patents, to position scientists in subject space. For this purpose, we first compile subject portfolios for each scientist, which are derived from the set of non-dyad publications prior to the treatment year.

	Overall Sample	Life Sciences	Health Sciences	Physical Sciences	Social Sciences
	Article count as dependent variable				
After death × treated	-0.029	-0.061 *	-0.015	-0.017	-0.035
	(0.025)	(0.027)	(0.038)	(0.034)	(0.131)
After death × treated × subject distance in 3. tertile	-0.060	-0.016	-0.108 *	-0.034	-0.113
	(0.037)	(0.072)	(0.054)	(0.067)	(0.478)
Log pseudo-likelihood	-187,601	-52,268	-78,768	-53,444	-2,971
No. of observations	265,707	80,526	114,406	66,550	4,225
No. of dyads	17,819	5,363	7,648	4,520	288
	Forward citations as dependent variable				
After death × treated	-0.049	-0.116 **	0.004	-0.058	0.024
	(0.033)	(0.042)	(0.047)	(0.049)	(0.189)
After death × treated × subject distance in 3. tertile	-0.141 ***	0.000	-0.190 **	-0.216 *	-0.116
	(0.050)	(0.081)	(0.073)	(0.097)	(0.604)
Log pseudo-likelihood	-2,761,000	-796,070	-1,139,737	-775,149	-38,745
No. of observations	265,629	80,526	114,370	66,508	4,225
No. of dyads	17,812	5,363	7,645	4,516	288

Tab. 2.8: EFFECT HETEROGENEITY IN SUBJECT SPACE

Notes: The overall sample consists of 9,297 pairs of treated and control collaborators confined to a nine-year window around the (inherited) year of star death. The panel is equally unbalanced for treated and control collaborators. Both output measures are co-author adjusted. Forward citations are winsorized at the 99.9th percentile. Collaborators without variation in output over the observation period are dropped by the estimation routine. Field delineation is based on the stars' publication profile or, if available, derived from the classification of the NAS. Robust standard errors are in parentheses, clustered at the level of the star.
* $p < 0.05$, ** $p < 0.01$, and *** $p < 0.001$.

Relying on Elsevier's most granular journal classification layer, these portfolios are akin to vectors with 334 elements, with each of them listing the share of publications in a specific subject category. We then calculate the Euclidean distance between these vectors, which enables us to quantify the gap that separates stars and their respective collaborators in subject dimension. As shown in Table 2.8, there is strong evidence that this measure proves central for understanding how treatment effects unfold. More specifically, it becomes apparent that collaborators of subject distant dyads, i.e., the upper third of the year- and field-specific distributions, suffer especially steep outcome declines in health and physical sciences. The differential effects on quality are large in magnitude and imply that these scientists see their forward citations decrease by an additional 17.3% and 19.4% in health and physical sciences, respectively.[21] As for the former field, we further determine a statistically significant drop in productivity that amounts to an extra 10.2%. Considered as a whole, research potential is primarily lost in duos that combined distant expertise. Not only does this finding lend support to the substitution theory formulated above, since stars should become harder to replace if collaborators have less inside knowledge about their colleagues' field, but it also shows that omitted knowledge transmission through interdisciplinary avenues constitutes a main treatment effect component.

	Overall Sample	Life Sciences	Health Sciences	Physical Sciences	Social Sciences
	Article count as dependent variable				
After death × treated	-0.016	-0.005	-0.045	0.020	-0.092
	(0.024)	(0.040)	(0.040)	(0.041)	(0.189)
After death × treated × star-star dyad	-0.044	-0.104 *	0.002	-0.076	0.222
	(0.029)	(0.044)	(0.046)	(0.049)	(0.262)
Log pseudo-likelihood	-189,126	-52,746	-79,328	-53,888	-3,017
No. of observations	275,344	83,541	118,212	69,107	4,484
No. of dyads	18,542	5,585	7,940	4,711	306
	Forward citations as dependent variable				
After death × treated	-0.046	-0.011	-0.058	-0.049	-0.079
	(0.029)	(0.048)	(0.042)	(0.051)	(0.241)
After death × treated × star-star dyad	-0.051	-0.160 **	0.029	-0.082	0.192
	(0.039)	(0.059)	(0.059)	(0.075)	(0.399)
Log pseudo-likelihood	-2,796,337	-807,991	-1,153,051	-783,440	-40,008
No. of observations	275,166	83,526	118,148	69,008	4,484
No. of dyads	18,527	5,584	7,934	4,703	306

Tab. 2.9: EFFECT HETEROGENEITY BY COLLABORATOR STATUS

Notes: The overall sample consists of 9,297 pairs of treated and control collaborators confined to a nine-year window around the (inherited) year of star death. The panel is equally unbalanced for treated and control collaborators. Both output measures are co-author adjusted. Forward citations are winsorized at the 99.9th percentile. Collaborators without variation in output over the observation period are dropped by the estimation routine. Field delineation is based on the stars' publication profile or, if available, derived from the classification of the NAS. Robust standard errors are in parentheses, clustered at the level of the star.

* $p < 0.05$, ** $p < 0.01$, and *** $p < 0.001$.

[21] The total decrease in physical sciences is presumably even higher but cannot be stated with certainty since the common part of the treatment effect now turns statistically insignificant. However, estimating Equation (2.2) on the subsample of subject distant dyads yields a precisely estimated total decrease of 24.0%, which is almost identical to the additive effect in Table 2.8.

To this point, it remains puzzling, which mechanisms account for the (pronounced) treatment consequences faced by life sciences dyads. As will become clear, looking into this matter gives rise to a two-fold explanation. We first investigate if scientists of higher and lower calibre are differently affected upon the stars' passing. From a theoretical viewpoint, one could emphasise that collaborators of lower calibre may generally be more reliant on the stars' influence and therefore bear the higher costs of treatment. However, this influence might not prove to be overly substantial since impact analysis shows that the success of collaborative work is rather restrained by lower-ability members than lifted by higher-ability members (Ahmadpoor & Jones, 2019). Moreover, one might be sceptical about the likelihood of future interactions if dyads comprise a (too) severe ability or performance gap. In order to resolve this question empirically, we differentiate between collaborators based on their scientific achievement prior to the death event. Drawing a line between regular and star co-authors, we discover the latter group to take up almost the entire treatment effects in life sciences. As reported in Table 2.9, stars experience additional consequences in form of a dual decrease of 9.9% in article counts and 14.8% in forward citations. These differential effects are statistically significant yet bound to the life sciences spectrum. The stars' deaths thus turn out to be particularly harmful for related star scientists, indicating that horizontal rather than vertical spillovers fuel knowledge production in this field.

The second channel, which allows insights into the effect formation in life sciences, pertains to the (broader) geographical dimension. A priori, it is unclear if variations in the treatment impact could be attributed to the science systems in which dyads are embedded. This possibility has not been explored by previous studies (Azoulay et al., 2010; Mohnen, 2018; Oettl, 2012), yet it seems conceivable that organisational aspects as institutional autonomy, competition, or stratification could alter a star's (external) value. We shed light on this matter by focussing on intra-US dyads, i.e., collaborations where both scientists are affiliated with an US institution at the time of treatment. These dyads represent 35% of the treated sample (versus 33% of the control sample) and evidently experience treatment consequences of a higher degree in health and life sciences. Given the statistically significant interaction terms in Table 2.10, we determine a differential productivity decline of 12.0% in the former field and a differential quality decline of 17.5% in the latter field. Although we remain limited in assessing the exact reasons for these effects, the US science system appears to be more star dependent.

	Overall Sample	Life Sciences	Health Sciences	Physical Sciences	Social Sciences
	Article count as dependent variable				
After death × treated	-0.013	-0.036	-0.003	-0.017	-0.177
	(0.028)	(0.042)	(0042)	(0.034)	(0.228)
After death × treated × US-US dyad	-0.090 *	-0.076	-0.128 *	-0.024	0.265
	(0.036)	(0.054)	(0.055)	(0.062)	(0.304)
Log pseudo-likelihood	-188,782	-52,674	-79,137	-53,836	-2,986
No. of observations	274,210	83,314	117,559	68,907	4,430
No. of dyads	18,461	5,569	7,892	4,698	302
	Forward citations as dependent variable				
After death × treated	-0.029	-0.026	-0.009	-0.088	-0.083
	(0.031)	(0.047)	(0.041)	(0.053)	(0.297)
After death × treated × US-US dyad	-0.141 **	-0.192 *	-0.115	-0.044	0.208
	(0.047)	(0.075)	(0.062)	(0.080	(0.412)
Log pseudo-likelihood	-2,789,367	-805,569	-1,149,362	-783,637	-39,389
No. of observations	274,032	83,299	117,495	68,808	4,430
No. of dyads	18,446	5,568	7,886	4,690	302

Tab. 2.10: EFFECT HETEROGENEITY IN GEOGRAPHICAL SPACE, US CHANNEL

Notes: The overall sample consists of 9,297 pairs of treated and control collaborators confined to a nine-year window around the (inherited) year of star death. The panel is equally unbalanced for treated and control collaborators. Both output measures are co-author adjusted. Forward citations are winsorized at the 99.9th percentile. Collaborators without variation in output over the observation period are dropped by the estimation routine. Field delineation is based on the stars' publication profile or, if available, derived from the classification of the NAS. Robust standard errors are in parentheses, clustered at the level of the star.

* $p < 0.05$, ** $p < 0.01$, and *** $p < 0.001$.

Finally, our results should be put into perspective. Uncovering heterogeneity in the causal effect of star death does not itself permit a causal interpretation. To be more concrete, our estimations do not identify how treatment effects would change if collaborators were (exogenously) moved along certain covariate dimensions. However, our analysis does reveal which types of collaborators, in fact, are exposed to higher treatment impacts, thus helping to develop a better understanding of the processes that shape scientific advancement. Before we turn to a discussion of our main results, we shortly allay some robustness concerns, which are detailed in Appendix A.6. First, horizontal spillovers and intra-US effects operate independently as both interactions remain statistically significant if included in the same estimation. Second, intra-US effects are neither a mere reflection of intra-country effects nor entangled with the co-location channel. Third, distance in subject space is not to be confounded with distance in topic, which is less predictive for the treatment effects. Fourth, our results do not hinge on the specific threshold definition that delineates subject distant dyads.

2.5 Discussion and Conclusion

The unexpected and premature death of 162 prolific scientists provides us with a quasi-experimental setting, in which we investigate how valuable a star collaborator's presence is for individual research performance. We find that scientists suffer average declines of 4.2% in article counts and 7.8% in forward citations following the exogenous passing of a star co-author. Furthermore, there are no signs of recovery patterns. Instead, treatment consequences seem permanent and rather increase over time, thus indicating that star exposure constitutes an irreplaceable asset.

Attempting to uncover the origins of the treatment effect, we first perform field-specific estimations, which we deem necessary given that cultures and practices differ along the scientific spectrum. In the course of this analysis, we generally confirm the findings of Azoulay et al. (2010) and Oettl (2012) as we determine a clear treatment impact in the field of life sciences, spanning both the productivity and quality sphere. In addition, we detect a quantitatively similar quality decrease for physical sciences dyads, which adds to the evidence presented by Borjas and Doran (2015) on high-quality mathematicians. On the contrary, collaborations in health and social sciences are (initially) found to escape any statistically significant treatment consequences. The absence of overall effects in these fields might have several reasons, which we cannot ascertain. In case of health sciences, for instance, one might argue that formal co-authorship could be less informative about true research interaction. Several studies raise the concern that guest or gift authorships lead to inflated co-author number in medical journals (Bhopal et al., 1997; Flanagin et al., 1998; Wislar, Flanagin, Fontanarosa, & DeAngelis, 2011).[22] In a separate vein, the death of a star scientist, as macabre as it may sound, could also emerge as beneficial for future performance. Prestigious research positions, journal space, funding, and accolades are all examples of scarce resources in academia, access to which could become less restrictive after the star is exempt from competition (Borjas & Doran, 2015). Although we find no evidence of positive treatment outcomes, it seems conceivable that the competition channel might (partly) offset negative effects. Moreover, we believe that this argument could carry particular relevance in fields where elite scientists form a small interlinked community, as it tends to the case for social sciences (Goyal, Van Der Leij, & Moraga-González, 2006).

[22] Our data could be reflective of this phenomenon (to some extent) as we observe health scientists to record the highest collaborator numbers (see Table A.3).

In a subsequent step, we exploit the rich heterogeneity in individual and dyadic data to develop an understanding of the mechanisms that give rise to the treatment effects. There are three findings that stand out. First, we provide evidence that knowledge production comprises spatial elements. On a broader scale, we determine US-located dyads to be a primary effect driver in both life and health sciences. The observation that US scientists that lose a star collaborator, who is likewise located in the US, experience steeper output declines points to systemic causes. What could cause them to be especially vulnerable? A probable answer relates to increased inequality levels in the US biomedical sector as documented by Katz and Matter (2019) who highlight rich-get-richer effects in terms of patents, publications, and research grants that reinforce the role of elites and limit the degree of upward mobility. Star contact could thus be more important for career paths in this environment. On a local scale, we further find co-location effects in physical sciences. Although several studies underline the general tendency towards distant collaborations (Jones, Wuchty, & Uzzi, 2008; Laband & Tollison, 2000; Waltman, Tijssen, & Eck, 2011), our analysis suggests that close workspaces can still be a relevant factor for knowledge production. More specifically, we find that some part of the spillovers generated by stars are locally confined. We take the view that the diverse range of specialised equipment and material used in physical sciences could offer an explanation, yet our conclusion is not clearly verifiable. Including data on physical capital, similar to Baruffaldi and Gaessler's (2018) approach, would therefore be a promising extension to our analysis.

Our second main result pertains solely to life sciences collaborations. In stark contrast to other fields, we notice that the sudden death of a star primarily casts a shadow on fellow star scientists. Horizontal rather than vertical spillovers are thus characteristic for frontier research in life sciences. While it lies beyond our scope to determine the exact reasons for this finding, we offer two plausible explanations. Unrealised joint production, to begin with, appears to play a minor role. Spillovers are, however, by no means restricted to activities within conventional research projects, but can likewise originate from informal interactions, e.g., from "frequent exchanges with strong minds and powerful scientific imaginations that have a deep understanding of the problems one is struggling with" (Stigler, 1988, p. 36) or from "testing out new ideas in casual conversations" (Borjas & Doran, 2015, p. 1116). We expect informal channels to be shaped by social proximity, so that knowledge sharing is primarily facilitated between scientists of similar standing and intellectual ability. Strong informal channels could thus explain why stars suffer the main treatment effects. An alternative explanation is borrowed from Azoulay et al. (2019) who shed light on the nature of entry barriers in life sciences. Following the star's death, they discover an influx of outsiders that, at the expanse of incumbent scientists, successfully challenge the leadership in the star's research domain. These dynamics illustrate that stars, while alive, can also serve as a protection that ensures that like-minded scholars keep the knowledge reins in their hands.

Third, our analysis discloses spillovers in subject space. The idea that linking divergent scientific backgrounds can accelerate the innovative process is indeed not new. Models of creativity have long highlighted that new ideas typically emerge from a recombination or synthesis of existing ideas (Campbell, 1960; Hadamard, 1945; Schumpeter, 1934; Usher, 1954; Weitzman, 1998). On a historical note, Robert Oppenheimer stated about the rise of atomic physics that it “was not the doing of any one man”, but instead “involved the collaboration of scores of scientists from many lands” (cited by Becker, 1957, p. 54). More recently, several bibliometric studies have explored the relationship between disciplinary diversity and citation impact. The conclusions drawn are not entirely consensus, but mostly supportive of a positive relation (Larivière, Haustein, & Börner, 2015; Leahey, Beckman, & Stanko, 2017; Uzzi, Mukherjee, Stringer, & Jones, 2013; Wang, Thijs, & Glänzel, 2015). Interdisciplinary research might not only lead to impactful results, it could also become more of a necessity. Scientific collaborations are oftentimes motivated by gaining access to specific competences, equipment, or data (Beaver, 2001; Melin, 2000). These reasons are rather pragmatic and can be considered to reflect specialisation tendencies in several research fields (Katz & Martin, 1997), which likely continue to increase due to the (ever) growing stock of knowledge (Jones, 2009). Our results align with this literature. More concretely, in health and physical sciences, we find that research potential is mainly lost in duos that combine markedly different field expertise, which is indicative of knowledge transmission through interdisciplinary avenues.

Finally, this paper presents the first causal estimation of spillover effects over the entire spectrum of scientific fields. On aggregate, we discover that the presence of a star scientist benefits the research performance of his or her collaboration network. However, exploring the domains of life, health, physical, and social sciences separately reveals that the star effect is neither visible in each of these fields nor traceable to one common origin. To this end, our study may be viewed as a contribution that can help to develop an improved understanding of knowledge production functions and their potentially heterogeneous forms. Future research could continue in a similar (or complementary) spirit, but address some of our limitations. Importantly, our coverage of social sciences dyads is limited, in the first place due to considerably smaller collaboration networks in this field, but also because of a moderate number of treatment cases. A related question would arise from a change of scenery. Do our findings translate to fields outside of the university sector? Oettl (2012) raises this point and illustrates the perception that tech companies typically value exceptional engineers to an extent that resembles star status in academia. After all, knowing how human capital accumulates by means of interaction would clearly have far-reaching implications and ultimately shed light on a key component of economic growth (Akcigit et al., 2018; Lucas & Moll, 2014).

3 Efficiency of European Universities: A Case of Apples and Peers[†]

3.1 Introduction

As a consequence of European integration, universities in Europe are more and more competing for students, research funding, and scientific personnel across borders (Erkkilä & Piironen, 2014; Teixeira, 2016). It is therefore all the more important not only to assess these universities relative to their national peers, but to conduct comparisons on a broader geographical scale. Yet most studies on higher education efficiency have so far confined attention to one country at a time. Against this backdrop, the present paper adds to the scarce literature on cross-country studies by investigating how efficiently universities from 16 European countries use the resources at their disposal. Apart from estimating relative efficiency scores, we further aim to identify relevant efficiency drivers, such as funding or personnel structure, by means of regression techniques.

Adopting a European perspective enables us to make a methodological contribution to the literature on university efficiency. Given that input-output patterns notably depend on subject composition, universities are far from being considered homogeneous. We propose to address this issue with clustering methods and individual peer-group selection that both build on distance in subject space. We hereby avoid any kind of unreasonable comparison of, for instance, technical universities and business schools. Such an approach naturally comes with the requirement of a sufficiently large number of institutions, which we meet by using the European Tertiary Education Register (ETER) along with Elsevier's database Scopus. In total, we hereby manage to compile a unique micro-dataset on 450 universities that contains information from nearly 2 million publications.

Overall, our main results can be summarised as follows. First, it becomes evident that efficiency comparisons should only be made for universities with similar subject focus. Otherwise, efficiency scores would be more reflective of subject differences, e.g., higher costs in medical or technical studies relative to social studies, than of more or less efficient resource use. Second, efficiency drivers show substantial effect heterogeneity between subject clusters, which illustrates that universities are shaped by different technologies. However, we third provide evidence that third-party funding shares and institutional size are to a large extent efficiency enhancing.

[†] This chapter is based on joint work with Berthold U. Wigger. A former version of our study is available as a CESifo working paper; see Herberholz and Wigger (2020).

The remainder of the paper is organised as follows. Section 3.2 discusses related literature strands and further elaborates on our key ideas. Section 3.3 establishes the classification scheme. Section 3.4 introduces the statistical Data Envelopment Analysis (DEA) approach. Section 3.5 explores estimated efficiency scores. Section 3.6 identifies efficiency drivers. Section 3.7 offers robustness checks. Section 3.8 briefly concludes.

3.2 Related Literature

Detecting inefficiencies within educational institutions has attracted much scholarly attention, most notably leading to empirical studies that employ various frontier efficiency techniques. Starting in the 1980s, numerous studies have focused on different types of institutions including primary and secondary schools, universities as well as university departments, or countries as a whole. For comprehensive reviews, see e.g., Worthington (2002) or De Witte and López-Torres (2017).

Focussing on higher education, the foundations have been laid by studies that were conducted on a single-country basis. Historically, Anglo-Saxon countries were at the centre of most early frontier analyses. For instance, within the United States, Ahn, Charnes, and Cooper (1988) and Ahn and Seiford (1993) were both concerned with comparing public and private doctoral-granting institutions, whereas Breu and Raab (1994) confined attention to the nation's top ranked institutions. Australian universities have also been subject to frequent assessment by, e.g., Coelli (1996), Avkiran (2001), and Abbott and Doucouliagos (2003). The same applies to institutions in the United Kingdom that have been analysed in depth. While the first studies on academic efficiency in Britain were conducted at the department level (see Tomkins and Green, 1988, Beasley, 1990, and Johnes and Johnes, 1995, on accounting, chemistry and physics, and economics departments, respectively), several contributions at the university level soon followed, for instance by Sarrico et al. (1997), Athanassopoulos and Shale (1997), and Johnes (2006). Moreover, McMillan and Datta (1998) provided insights into the relative performance of Canadian universities, while Taylor and Harris (2004) addressed the topic in the South African context. Apart from the origins in the Anglo-Saxon area, higher education efficiency has emerged as a (research) topic of global interest as can be inferred from studies covering institutions in Austria (Leitner, Prikoszovits, Schaffhauser-Linzatti, Stowasser, & Wagner, 2007), Germany (Kempkes & Pohl, 2010), Italy (Agasisti & Salerno, 2007), Greece (Katharaki & Katharakis, 2010), Brazil (Zoghbi, Rocha & Mattos, 2013), Mexico (Sagarra, Mar-Molinero, & Agasisti, 2017), or China (Johnes & Yu, 2008).

Frontier techniques are essentially driven by the number of (decision-making) units under assessment since efficiency refers to an endogenous concept, where benchmarks are set by the best performing subgroup. The aforementioned studies are therefore bound to national efficiency frontiers, which is apparently at odds with the widespread view of universities competing on a global scale. The limitations of country-specific studies have indeed motivated a novel stream of literature, i.e., cross-country studies. Among them, Joumady and Ris (2005) were arguably the first to make a contribution by exploiting a postal survey sent to young professionals three years after graduation. In total, they were able to assess 209 institutions from eight European countries regarding their capacity to prepare students for labour market transition. Due to the unique survey setting, this work marks a rather special case. In contrast, most subsequent studies pursued an alternative path by using administrative data derived from national agencies. The need for manual data adjustments might be a reason why several studies started to adopt a two-country perspective. For instance, Agasisti and Johnes (2009) compared universities from England and Italy and noted that the latter ones were largely outperformed in the academic year 2003/04. Following a similar methodology, Agasisti and Pérez-Esparrells (2010) conducted an analysis of Italian and Spanish universities. As of the academic year 2004/05, Italy was this time found to operate at higher efficiency levels. In fact, further two-country studies were centred around Italy based on data from 2000 onwards. Agasisti and Pohl (2012) observed a lower efficiency of Italian universities relative to their German counterparts, while comparisons to Polish (Agasisti & Wolszczak-Derlacz, 2016) and Dutch (Agasisti & Haelermans, 2016) institutions revealed that efficiency differentials are mostly model-dependent.

Overall, two-country studies can be regarded as a first step to account for increasing internationalisation in higher education. However, comparisons on a broader geographical scale are still required to obtain a more complete picture of cross-border competition and production possibilities. Apart from Joumady and Ris (2005), only a handful of studies have addressed this need to date, which, for the most part, can be explained by the lack of comparable micro-data at the institutional level (Wolszczak-Derlacz, 2017). Wolszczak-Derlacz and Parteka (2011) approached this issue by means of a multitude of sources that led to a dataset on 259 universities from seven European countries. Extending the scope of analysis, Wolszczak-Derlacz (2017) compared 152 US to 348 European universities from ten countries, again based on manually collected data. Both studies clearly show that efficiency scores vary not only within but also between countries. Further studies were built on the projects Aquameth and Eumida, which were initial attempts by the European Commission to construct a unified database on higher education institutions. Exploiting these data, Daraio, Bonaccorsi, and Simar (2015a) investigated economies of scale and specialisation, Bolli et al. (2016) emphasised the role of competitive funding, while Daraio, Bonaccorsi, and Simar (2015b) proposed an advanced approach to rank universities with frontier techniques. Albeit these recent contributions, cross-country studies on academic efficiency are evidently still scarce. We therefore aim to extend this strand of literature with the help of a novel dataset. To the best of our knowledge, we are the first to utilise ETER for efficiency purposes, which enables us to expand our scope beyond past research and present a strong case for the validity of our findings.

Apart from providing an extended cross-country perspective on university efficiency, we propose a methodological contribution to a second stream of literature, which has rather been neglected in recent work. Specifically, we take the view that subject mix differentials have to be addressed comprehensively to avoid a well-known pitfall of DEA applications, i.e., comparing non-homogeneous units (Dyson et al., 2001). In fact, several studies have highlighted various systematic differences between academic fields. For instance, Tierny (1980) provides early evidence on costs per student at liberal arts colleges and shows that chemistry departments exceed political science departments by up to 100%. The general notion that social sciences incur lower cost levels than physical sciences is also confirmed by Dundar and Lewis (1995), who, additionally, discover the highest costs in the field of engineering sciences. Further cost studies have come to similar conclusions. Zimmerman and Altonji (2018) examine instructional spending in the Florida State University System and discover substantial heterogeneity. According to their results, engineering graduates entail costs that are almost double the amount found in low-cost majors such as business. Consistently, Filipini and Lepori (2007) explore expenditure levels of Swiss universities and report the highest values for technical sciences along with medicine. In view of the sharp differences between disciplines, they emphasise that cost comparisons of universities are at risk of being distorted if subject composition is left unconsidered. Johnes (1990) adds to this line of reasoning by stating that over two thirds of the variation in unit costs of UK universities is attributable to subject mix alone. There appear to be two main reasons for these patterns. On the one hand, STEM-related fields but also medicine generally require physical resources to a different extent and magnitude, e.g., basic materials, clinical and mechanical equipment, laboratories, and other costly facilities. On the other hand, some of these fields are considered to be more labour-intensive with higher levels of interaction between students and faculty, which is reflected by a different personnel structures (Kempkes & Pohl, 2010).[23]

The relevance of external research funding is closely related to the cost dimension and equally well documented to be subject-dependent. The findings present a clear picture in so far as STEM-related fields and medicine are most active in third-party collaborations. Social sciences and humanities, on the contrary, are clearly underrepresented, likely due to less commercial potential in these fields (Bonaccorsi, Secondi, Setteducati, & Ancaiani, 2014; Gulbrandsen & Smeby, 2005; Hornbostel, 2001). Relying on third-party funding as a proxy for research performance, as often favoured in the absence of bibliometric data, thus becomes a two-fold problem. Not only does this practice raise economic concerns about confounding inputs with outputs (Johnes & Johnes, 1995), but it also introduces unfair judgement. Publication and citation counts provide preferable output measures,

[23] The latter argument is presumably not restricted to STEM-subjects and medicine. For instance, music and art are also characterised by high levels of instruction.

are, however, prone to bias too. Shin and Cummings (2010), for instance, conclude that field differences constitute the main source of variance in faculty publications. More specifically, publication rates in engineering, natural, and medical sciences are found to exceed those in social sciences and humanities. Piro, Asknes, and Rørstad (2013) confirm this pattern while also emphasising the effects of alternative counting methods. Once fractional publication counts are employed to account for higher co-author numbers in natural sciences, the picture clearly changes with humanities and social sciences ranking first and second, respectively. Field differences are even more visible when it comes to the distribution of citations (Waltman, 2016). To illustrate this point, Radicchi, Fortunato, and Castellano (2008) state that publications with 100 citations are about 50 times more common in developmental biology than in aerospace engineering, while Waltman et al. (2011) find citation counts in biochemistry to be roughly one order of magnitude higher than in mathematics.

Moreover, there is evidence that educational processes are also subject to considerable heterogeneity. According to Smith and Naylor (2001a, 2001b), both completion rates and degree results are affected by the field of study. For instance, in comparison with social sciences, it is more likely to receive a good degree in humanities and biological sciences, whereas dropout risk is increased in computer sciences.[24]

In conclusion, disciplinary differences are substantial, take various forms, and have long been studied. Yet we are not aware of any study on higher education efficiency that has addressed this issue thoroughly. Some attempts were based on the distinct features of medical studies, which led to separating institutions with and without medical schools (Agasisti & Salerno, 2007; Ahn et al., 1988; Thanassoulis, Kortelainen, Johnes, & Johnes, 2011) or to adjusting data of medical schools (Hanke & Leopoldseder, 1998). In fact, the most comprehensive approach might have been presented by Athanassopoulos and Shale (1997), who divided UK universities into three groups with different science orientation levels. It thus seems that the ensuring homogeneity has mostly been overlooked.

While efficiency scores may thus be suffering from a considerable bias, subject mix has attracted (newfound) interest when it comes to explaining efficiency scores within two-stage frameworks. From an economics perspective, running a second-stage regression is of particular importance to gain insights into efficiency drivers on which grounds policy implications can be drawn. Medical faculties have frequently been included within these regressions (Agasisti & Pohl, 2012; Agasisti & Wolszczak-Derlacz, 2016; Kempkes & Pohl, 2010; Wolszczak-Derlacz, 2017; Wolszczak-Derlacz & Parteka, 2011) and in several cases found to have a significant impact. However, it may be questioned whether universities

---

[24] Both studies by Smith and Naylor include gender specific estimations and cover a variety of subjects. Their approach leads to numerous findings, which are not covered in detail at this point. The results presented are therefore rather illustrative albeit significant and representative for both genders.

can be commonly assessed without accounting for subject composition. Universities with a strong life sciences profile might simply be incapable of reaching an efficiency frontier composed of universities that are primarily engaged in social sciences. In fact, they might not even consider these universities their peers, which essentially casts doubt upon the managerial side of relative efficiency techniques. Moreover, regression results become prone to misinterpretation once biased efficiency scores are utilised. Rather than being a source of inefficiency, medical faculties might be more likely to illustrate systematic differentials between academic fields that, of course, become more or less pronounced depending on the respective choice of inputs and outputs.

3.3 Clustering Analysis

3.3.1 Methodology

There is an extensive number of clustering methods, from which we select the K-means algorithm. It is widely considered an elegant method for splitting a dataset into distinct clusters.[25] The idea behind K-means can be formalised in an intuitive way: Let $C_1, \dots, C_K$ denote K sets of distinct, non-overlapping clusters. Since clustering aims at grouping observations that tend to be similar, one can assess clusters based on their within-cluster variation, which should be as small as possible. The problem to be solved by the K-means algorithm can thus be stated as

$$\min_{C_1,\dots,C_K} \sum_{k=1}^{K} W(C_k), \tag{3.1}$$

where $W(C_k)$ denotes the within-cluster variation of cluster C_k. A common way to measure the within-cluster variation of C_k refers to the sum of squared distances between each observation $x \in C_k$ and the cluster's mean μ_k. Using squared Euclidean distance, we can redefine the optimisation problem as follows

$$\min_{C_1,\dots,C_K} \sum_{k=1}^{K} \sum_{x \in C_k} \|x - \mu_k\|^2. \tag{3.2}$$

While the logic underlying K-means gives little cause of concern, one regularly faces the practical issue of selecting the parameter K. Since there is no universal approach for this task, we mainly follow Makles (2012) and determine the optimal cluster number based on the total within-cluster variation, i.e., the target value of the optimisation problem defined above. More specifically, we consider any increase in K desirable as long as it is

[25] Albeit the growing popularity of clustering applications, the higher education landscape has only partially been explored by these techniques. Notable examples are Stanley and Reynolds (1994) and Valadkani and Worthington (2006), who study performance differences within the Australian university system. In a similar vein, Shin (2009) groups South Korean universities based on research performance, whereas Bonaccorsi and Daraio (2009) and Lepori, Baschung, and Probst (2010) develop classification schemes for European universities.

accompanied by a sufficient reduction of that value. For this purpose, we emphasise comparing the proportional reduction of error for different values of K. Formally, this coefficient is defined as

$$PRE(K) = 1 - \frac{WSS(K)}{WSS(K-1)}, \tag{3.3}$$

where $WSS(K)$ denotes the total within-cluster variation for a solution of $K > 1$ clusters. Once this coefficient drops considerably, we refrain from partitioning our data any further. Additional explanations regarding our clustering methodology are partly provided over the course of the next subsection.

3.3.2 Data and Results

Our analysis covers the period from 2011 to 2014 and exploits two main data sources. The core data are derived from ETER, which provides comparable micro-data on higher education institutions across Europe. In addition, we use data from Scopus, an abstract and citation database hosted by Elsevier, in order to supplement our institutional data with meaningful measures of research output. After restricting our dataset to public and government-dependent universities and eliminating specialist institutions (e.g., music and arts academies), 450 universities from 16 European countries remain to constitute our sample (see Appendix B.1 for a geographical depiction).[26]

As for the clustering analysis, we rely on publication records collected from Scopus. In principle, one could also argue in favour of employing student enrolment data for this task. Yet, we consider research output the more adequate choice primarily because our subsequent efficiency analysis addresses research activities in greater detail. Scopus does not only cover a broad range of scientific literature, but also classifies its content under four main subject areas,[27] i.e., life sciences, social sciences, physical sciences, and health sciences. Building on this classification system and our pooled dataset, we calculate each university's share of publications in these subject areas, which determine its position in subject space. These vectors then serve as the foundation for the clustering analysis. However, it should be noted that Scopus's subject areas are partly overlapping. Articles that, for instance, appear in *Applied Mathematical Finance* are assigned to both social sciences and physical sciences. Multidisciplinary work therefore entails the potential risk of distorting subject profiles towards research areas with greater overlap. To address this issue, we opt for a fractional counting approach that essentially divides each (ambiguous) publication evenly between its subject areas.

[26] Institutions are classified as government-dependent if a government agency provides either more than half of their core funding or their teaching personnel's salary. Due to the reliance on public funding, these institutions are often regulated in very similar ways to public institutions (OECD, 2019, p. 160).

[27] Note that the auxiliary subfield of multidisciplinary studies is omitted from the clustering analysis.

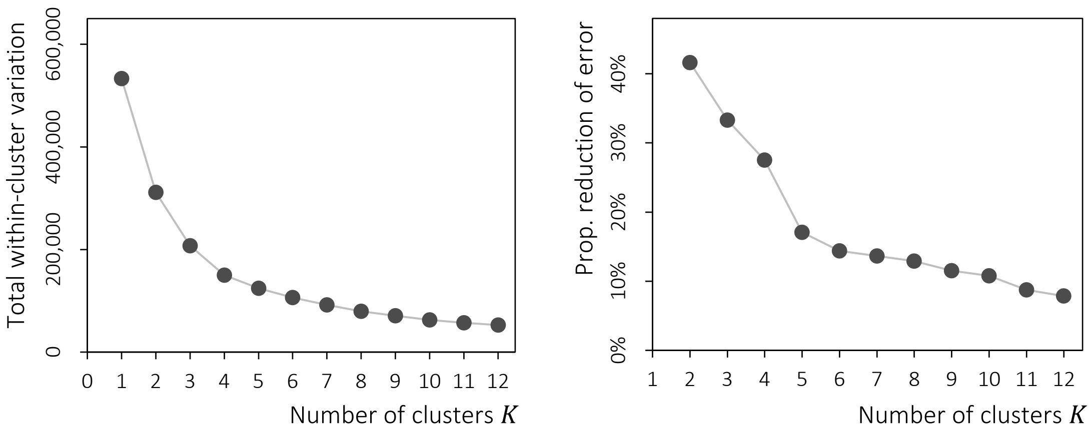


Fig. 3.1: HEURISTICS FOR OPTIMAL CLUSTER NUMBER
Notes: Values are averaged over 1,000 replications of K-means with random starting centres.

The task of selecting an appropriate cluster number is addressed in Figure 3.1. The left-hand panel depicts the total within-cluster variation for different values of K. Raising the number of clusters apparently reduces variation; however, there is a diminishing benefit along with it. This effect becomes even more evident in view of the right-hand panel, which plots the proportional reduction of error. Based on this criterion, adding a second cluster has the biggest impact, reducing total within-cluster variation by 42%. By adding a third and fourth cluster, the variation continues to fall by 33% and 28%, respectively. Afterwards, the graph shows a relatively steep decline. A fifth cluster would decrease variation by (merely) 17%, which would barely differ from adding a sixth, seventh, or eighth cluster, so that any of these solutions appears rather arbitrary. While deciding on the optimal number of K is usually not a clear cut, the heuristics are mostly supportive of a four-cluster solution in the present case. This also pertains to the Caliński–Harabasz (1974) index that we calculated as an additional check (see Appendix B.2).

The final step of our clustering approach is to apply the actual K-means algorithm. One shortcoming of K-means arises from the fact that it tends to converge to local instead of global optima. We thus ran the algorithm multiple times with random starting centres and selected the solution with the lowest total within-cluster variation as suggested by James, Witten, Hastie, and Tibshirani (2013, pp. 388–389). In sum, the final clustering was obtained in 51 of 1,000 replications. Aggregate statistics on subject space location by cluster are presented in Table 3.1.

Cluster	*N*	Life Sciences	Social Sciences	Physical Sciences	Health Sciences	Subject Focus
CLUSTER 1	57	6.25%	56.10%	22.21%	15.11%	*Social Sciences*
CLUSTER 2	140	8.49%	10.74%	74.02%	6.30%	*Physical Sciences*
CLUSTER 3	49	21.41%	12.41%	14.43%	51.24%	*Health Sciences*
CLUSTER 4	204	20.15%	14.95%	40.80%	23.32%	*General*
Sample	450	14.90%	18.57%	45.91%	20.02%	

Tab. 3.1: MEAN COMPOSITION OF RESEARCH OUTPUT BY CLUSTER

According to Table 3.1, our sample consists of universities that, on average, account for scientific output, 15% of which falls under the life sciences category, 19% under social sciences, 46% under physical sciences, and 20% under health sciences.[28] However, the data further show that the European public university landscape can hardly be regarded as homogeneous. In fact, there are significant differences in terms of subject focus. This becomes particularly apparent with regard to specialist clusters such as CLUSTER 1. On average, 56% of the publications by a CLUSTER 1 university belong to social sciences. In contrast, other clusters display mean values that are up to five times smaller ranging from 11 to 15%.[29] A similar degree of specialisation can be observed by CLUSTER 3, which is composed of universities that lay its emphasis on health sciences. These two clusters also resemble each other from the size perspective, as they are considerably smaller than the remaining clusters. Albeit comprising a lot more universities, CLUSTER 2 can still be viewed as a specialist cluster that is directed towards physical sciences. Lastly, CLUSTER 4 contains universities, which most closely align with the sample mean (thereby sharing the general tendency towards physical sciences). We thus consider them as generalist institutions.

[28] Physical sciences seem to be overrepresented, which is partly attributable to database coverage. However, this bias is found to be even larger within the Web of Science, which may have served as an alternative data source. Besides, broad-scale comparison reveals that Scopus exceeds the Web of Science in terms of journal coverage in every disciplinary field (see Mongeon and Paul-Hus, 2016, on both aspects) and thus provides a more reliable basis for efficiency assessment.

[29] It should be noted that mean values can be somewhat misleading. For instance, a few universities outside of CLUSTER 1 are visibly engaged in social sciences along with their primary cluster focus. Yet only two of them marginally exceed the lower bound of CLUSTER 1 set by the Birmingham City University (35.57%). Still, this observation is indicative of cluster boundaries being partly fluid (see Section 3.7).

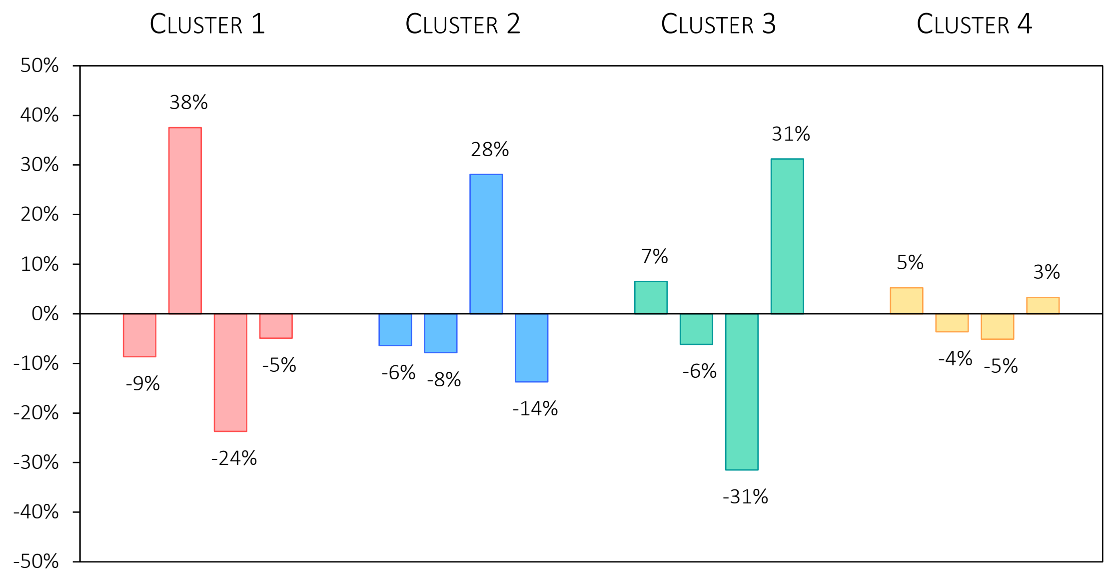


Fig. 3.2: SUBJECT DEVIATION BY CLUSTER IN COMPARISON TO SAMPLE MEAN
Notes: Bar order = life sciences, social sciences, physical sciences, health sciences.

Cluster characteristics are further illustrated by Figure 3.2, which depicts how far clusters deviate from the sample mean, and by Figure 3.3, which emphasises comparisons inside the subject space boundaries that are effectively set by our data. The latter approach is particularly relevant given that the maximum degree of specialisation is found to differ notably between subject areas. For instance, while we discover universities with output fractions of above 90% in social and physical sciences, peak values in life and health sciences lie within the 60% and 70% region, respectively. Employing an identical scale along each subject dimension could therefore conceal insights. Instead, we apply linear transformations to map our data onto the intervals ranging between the 1st and 99th percentile.[30] Following this approach, it first becomes clear that our former results hold true: Three clusters can be described by a distinct subject focus. Yet differences in the degree of specialisation appear in a partly different light. For instance, Figure 3.3 reveals a comparable level of specialisation for CLUSTER 2 and 3, which is primarily due to higher expansion on the health sciences axis. In other words, both clusters become increasingly similar if we acknowledge that specialisation in the field of health sciences relates to lower output fractions than in physical sciences. This effect is indeed most pronounced in life sciences, where values above 50% rarely occur. As a result, universities in CLUSTER 4 appear increasingly balanced hence approaching the perception of generalist institutions more closely.

[30] Standard rescaling refers to utilising minimum and maximum values. As a result, this approach is known to be sensitive to outliers. Percentile ranks thus provide a reasonable alternative.

Fig. 3.3: SUBJECT FOCUS BY CLUSTER IN COMPARISON TO SAMPLE MEAN

Notes: Data are rescaled to lie between the 1st and 99th percentile. Axes range from 0-100%. The sample mean is depicted by the dotted line.

The clustering analysis clearly sheds light on systematic differences between groups of universities in Europe. It is worth noting that we partly confirm the results of Lepori et al. (2010), who identify specialised institutions in the fields of technical-natural sciences and social sciences and humanities. Subject differences alone could indeed be overlooked by efficiency analyses if they were not linked to further institutional disparities. However, descriptive statistics presented in Table 3.2 point to the contrary. Referring to the output dimension, publications per academic staff, measured as full-time equivalents, constitute a typical indicator for scientific productivity. In line with the cited literature, we discover the lowest values within the social sciences cluster. On average, we find these universities to record an annual number of 0.49 publications per academic employee between 2011 and 2014.[31] In comparison, mean values of 0.83 and 0.91 are achieved by universities

[31] Our study is not restricted to research articles but includes every publication format from Scopus. This is particularly relevant for the field of social sciences and humanities where books and book chapters are known to play an important role in scientific communication.

focussing on physical and health sciences, which represents an increase of 69% and 86%, respectively. Overall, the general cluster is associated with the highest productivity of 1.01 publications per academic staff, which may be an indicator for the existence of economies of scope. A similar picture emerges with regard to the number of citations per publication that we include to capture the impact of scientific contributions. Based on an evaluation window that covers the year of publication plus two subsequent years, we determine an average citation rate of 3.13 within the social sciences cluster, which amounts to less than half of what their counterparts with a health or general profile are able to accomplish.

With respect to the input dimension, we direct attention to current expenditures as a summary measure of resource usage. By comparing the annual expenditure levels per student, we clearly observe the health sciences cluster to be a costly exception. On the contrary, universities from the social sciences cluster record relatively low expenditure numbers despite being exposed to a higher teaching load (as indicated by the ratio of students to academic staff). Again, both of these findings are consistent with the reviewed literature. Lastly, we see that universities focussing on social and health sciences are of similar size accommodating an average of 11 to 12 thousand students. In comparison, we find the number of students to exceed 15 thousand in the physical sciences cluster and approach 22 thousand among generalist universities.

It is crucial to note that systematic differences between clusters are a major cause of concern from the standpoint of efficiency analysis. More specifically, they suggest that production processes are subject to heterogeneous technologies, which would remain unconsidered if universities were pooled together across the entire subject spectrum. Instead, we strongly argue in favour of performing efficiency estimation cluster-wise to ensure a comparison of (true) peers.

Cluster	Publications per academic staff	Citations per publication	Students per academic staff	Expenditures per student	Number of students
1 – Social Sciences					
P5	0.06	1.32	11.09	3,569	1,343
Mean	0.49	3.13	20.91	9,911	11,029
P95	1.35	5.91	32.00	20,199	22,945
2 – Physical Sciences					
P5	0.17	1.91	5.41	3,435	3,043
Mean	0.83	4.86	16.89	12,564	15,575
P95	1.77	9.44	29.12	31,530	35,798
3 – Health Sciences					
P5	0.12	1.87	1.47	4,341	1,608
Mean	0.91	6.73	15.69	38,836	11,921
P95	2.68	12.39	30.84	213,706	30,027
4 – General					
P5	0.32	2.77	5.17	5,228	6,928
Mean	1.01	6.81	15.13	15,189	21,882
P95	1.82	11.01	28.33	38,542	48,150
Sample					
P5	0.16	1.94	5.15	3,944	3,059
Mean	0.88	5.73	16.47	8,933	17,461
P95	1.81	10.72	29.17	39,265	38,515

Tab. 3.2: SUMMARY STATISTICS BY CLUSTER

Notes: Publications comprise all document types listed on Scopus. The citation window covers three years including the year of publication. Academic staff is expressed in FTE. Financial data are converted into real PPP EUR (2014 = 100).

3.4 Statistical DEA Approach

We employ a statistical DEA approach in line with Simar and Wilson (1998, 2000). Thus, let $x \in R_+^p$ denote a vector of p inputs and $y \in R_+^q$ a vector of q outputs. The production possibilities set can then be defined as

$$P = \{\, (x, y) \in R_+^p \times R_+^q \mid x\ can\ produce\ y \,\}. \quad (3.4)$$

Production facilities, in our case universities, in the interior of P are termed technically inefficient, whereas universities located on the boundary, or frontier, of P are considered technically efficient. In order to determine the degree of efficiency, we adopt an output-oriented perspective implicitly assuming that universities have greater control over outputs than inputs.[32] A university located at a given point (x, y) can thus be assessed by

[32] This view is shared by a number of studies, including those by Agasisti and Johnes (2009), Kempkes and Pohl (2010), and Wolszczak-Derlacz and Parteka (2011).

$$\theta(x, y \mid P) = sup \{ \theta > 0 \mid (x, \theta y) \in P \}, \tag{3.5}$$

where $\theta(x, y \mid P) \in [1, \infty)$ measures the largest radial expansion of y that is feasible given x. Higher inefficiency is accordingly indicated by larger values of $\theta(x, y \mid P)$. In theory, inefficiency scores could be obtained through mathematical programming if the set of production possibilities were fully disclosed. However, this is not the case. Instead of observing all possible input-output combinations, one generally encounters a subset of technologies from P, denoted by $\hat{P}$. We thus refer to P and $\theta(x, y \mid P)$ as the true but unknown quantities of interest and to $\hat{P}$ and $\theta(x, y \mid \hat{P})$ as their sample estimators.

By construction, $\hat{P}$ constitutes an inner approximation of P, which causes inefficiency estimates to be downward biased, i.e., $\theta(x, y \mid \hat{P}) \leq \theta(x, y \mid P)$. In dealing with this issue, one generally relies on bootstrap-based inference. This leads to a virtual environment, where $\hat{P}$ and $\theta(x, y \mid \hat{P})$ become the quantities of interest to be estimated by P^* and $\theta(x, y \mid P^*)$, which build on subsets drawn from the original data. Further, let $\hat{F}$ refer to the bootstrap data generating process that mimics the true data generating process F. It then follows that

$$\theta(x, y \mid \hat{P}) - \theta(x, y \mid P^*) \mid \hat{F} \sim \theta(x, y \mid P) - \theta(x, y \mid \hat{P}) \mid F, \tag{3.6}$$

so that a bias-corrected estimator of $\theta(x, y \mid P)$ can be stated as

$$\tilde{\theta}(x, y) = 2\, \theta(x, y \mid \hat{P}) - E(\theta(x, y \mid P^*)). \tag{3.7}$$

Technically, we employ the homogeneous bootstrap algorithm proposed by Simar and Wilson (1998) based on 1,000 replications. In addition, we assume free disposability along with convexity and allow for variable returns to scale when constructing estimates of P. This procedure leads to bias-corrected efficiency scores $\{\tilde{\theta}_i : i = 1, \dots, n\}$ for our set of n universities, which we examine descriptively and further by means of kernel densities. For this reason, let f denote the density of $\tilde{\theta}$. Its standard kernel density estimator at any point u is then defined as

$$\hat{f}(u) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{u - \tilde{\theta}_i}{h}\right), \tag{3.8}$$

where $K(\cdot)$ denotes a kernel function and h constitutes a suitable bandwidth. However, due to efficiency scores being constructed with bounded support, this estimator requires alteration to ensure consistency. We therefore apply the modified estimator

$$\hat{f}_R(u) = \begin{cases} \frac{1}{nh_R} \sum_{i=1}^{n} \left[K\left(\frac{u - \tilde{\theta}_i}{h_R}\right) + K\left(\frac{u - (2 - \tilde{\theta}_i)}{h_R}\right) \right], & u \geq 1 \\ 0, & otherwise, \end{cases} \tag{3.9}$$

where h_R denotes an adjusted bandwidth. Moreover, we opt for a Gaussian kernel and follow Silverman's (1986) rule for bandwidth selection.[33]

In a second stage, we investigate potential efficiency drivers by employing the bootstrap regression framework by Simar and Wilson (2007). More precisely, we expect university i's true efficiency θ_i to depend on a vector z_i of covariates. On the assumption that these covariates exert constant percentage effects, our model resolves to

$$ln(\theta_i) = \psi\,(z_i, \beta) + \epsilon_i, \tag{3.10}$$

where β denotes a vector of coefficients, $\psi\,(\cdot)$ describes a functional form later to be determined, and ϵ_i represents the unexplained residual term, which is assumed to be normally distributed with left-truncation at $-\psi\,(z_i, \beta)$. We estimate this model by means of maximum likelihood, again based on 1,000 replications. With true efficiencies remaining unknown, we rely on their bias-corrected estimates for inference about β.

3.5 Efficiency Estimation

3.5.1 Model Specification

Universities are generally known to engage in two major fields of activity, i.e., teaching and research. With regard to teaching, we include the number of yearly graduates on ISCED levels 5 to 7 as our preferred output measure. Some studies instead opted for the number of enrolled students noting that education received by students who drop out before graduation should not be neglected (Cohn, Rhine, & Santos, 1989). However, this approach could be prone to misjudgement caused by inactive students (so-called phantom students), who are of particular relevance when evaluating public institutions (Teixeira, Rocha, Biscaia, & Cardoso, 2012). In fact, universities that handle these students least effectively, hereby revealing inefficient administrations, were the primary beneficiaries of this measure. Besides, study efforts that fail to be rewarded with degrees should not be overstated since job market returns are considerably decreased after dropping out (Walker & Zhu, 2013).

To capture research activities, we include the number of scientific publications in our model. They are indeed central for knowledge dissemination as scientific contributions usually manifest in some form of publication. However, it seems reasonable to argue that publications only serve as a partial indicator for research output. Following Martin and Irvine (1983), we regard them mainly as a measure of scientific production but not of scientific progress. This distinction rests upon the notion that publications tend to vary in scientific value. Most contributions might be incremental in nature, while some add considerably to the advancement of science. Individuals but also institutions as a whole

[33] See Sickles and Zelenyuk (2019) for further details on density estimation in efficiency contexts.

might have different preferences and abilities regarding these two dimensions, which calls for an additional output measure. We therefore incorporate citation counts, hereby broadening the scope of prior efficiency studies that usually omitted this measure due to data limitations.[34] Citations can be viewed as a quality measure, especially in settings that account for field differences. Yet, in a scientometric context, the right terminology would be to refer to citations as a measure of scientific impact (Martin & Irvine, 1983). Furthermore, one might speak of short-term impact given that our citation window is restricted to a maximum of three years. Moed et al. (1985) point out that not every contribution to the current research frontier eventually becomes accepted knowledge, which motivates the distinction between short- and long-term impact. However, empirical evidence indicates that both concepts are closely linked to each other (Adams, 2005). We further clarify this relation through a separate analysis that underlines the significant correlation between initial and overall citations at the institutional level (see Appendix B.3). Based on these results, our efficiency estimates can be expected to be robust to extended citation windows.

Contrary to the output side, the literature reveals less consistency over the input choice. This becomes especially evident by the numerous expenditure types that have been used, including e.g., expenditure on personnel, central administration, or library services. These differences may partly result from the availability of data but, more importantly, express alternative views of higher education efficiency. Our take on this matter is rather strict. In line with Thanassoulis et al. (2011), we define the current expenditure level (converted in purchasing power parities) as our single input. Two main reasons can be pointed out in support of this approach. First, from a public finance standpoint, it seems hardly relevant in what specific way resources are allocated within an institution. Universities are given a great amount of (operational) freedom which can shape production processes in various forms, with labour- and capital-intensive organisation being two classic examples. It is, at best, of secondary concern to policy makers how efficient universities are making use of certain resources. Their focus is expected to lie on the overall budget. Second, there are technical reasons for limiting the number of inputs. DEA is a flexible technique that allows units to attach individual weights to input and output components so that they appear in the most favourable light relative to their peers. Broadening the set of inputs would therefore open up more opportunities for universities to become efficient, which we consider unreasonable. To illustrate this point, adding the number of students as a second input dimension would permit universities to be assessed based on their citation-to-student ratio, which would be at odds with our efficiency perception.

---

[34] Citation counts have frequently been advocated but rarely been included to capture research output. To the best of our knowledge, only one study exploits citation data for efficiency purposes, i.e., Bonaccorsi, Daraio, and Simar (2006), who examine the Italian university system.

Overall, our model includes publications, citations, and graduates as outputs and current expenditures as an aggregate input measure. Moreover, we employ a fractional counting approach meaning that credit for publications and citations is split between collaborating universities according to the number of contributing authors. In view of a potential bias related to varying centrality within the European university network, it is worth noting that we consider co-authorship ties to any other affiliation for this task and not only links to universities from our sample.

3.5.2 Results

This section proceeds with exploring the results of our efficiency estimation, which we conduct separately for the years from 2011 to 2014. While we advocate cluster-specific technology frontiers, we also present results based on a global frontier to contrast both approaches.

Cluster	*N*	P5	P25	P50	Mean	P75	P95	SD
Cluster-specific Frontiers								
Social	228	1.15	1.30	1.64	2.04	2.45	4.23	1.11
Physical	560	1.16	1.33	1.75	2.05	2.47	4.03	1.02
Health	196	1.14	1.32	1.67	2.22	2.41	5.03	1.47
General	816	1.18	1.47	1.81	2.03	2.27	3.75	0.81
Global Frontier								
Social	228	1.16	1.71	2.20	2.60	2.97	5.39	1.52
Physical	560	1.16	1.41	1.96	2.23	2.66	4.31	1.10
Health	196	1.22	1.75	2.32	3.00	3.53	7.35	2.14
General	816	1.18	1.53	1.90	2.16	2.43	4.15	0.91

Tab. 3.3: SUMMARY STATISTICS ON BIAS-CORRECTED EFFICIENCY SCORES BY CLUSTER

Table 3.3 provides summary statistics on efficiency estimates aggregated by cluster and pooled over the 4-year sample period. From the upper half of this table, we can infer that assessing universities cluster-wise reveals efficiency distributions with a high degree of similarity (with a minor exception being the health cluster where the long tail, i.e., the low-performance segment, appears slightly more accentuated). In contrast, results derived from a global frontier indicate that efficiency differs notably between clusters. This finding becomes increasingly visible as percentile ranks increase. There are indeed highly efficient universities within each cluster; however, beyond the 5th percentile, we see an efficiency gap widening between the social and health cluster on the one hand and the physical and general cluster on the other hand. To evaluate this gap in more detail, we employ the non-parametric Kruskal-Wallis test, which clearly rejects the null hypothesis of equal mean ranks across clusters (χ^2 = 46.986 with an associated p-value of 0.0001). In light of this significant result, one might be inclined to draw the conclusion that some clusters simply outperform others. Again, we consider this rather a display of unreasonable comparison.

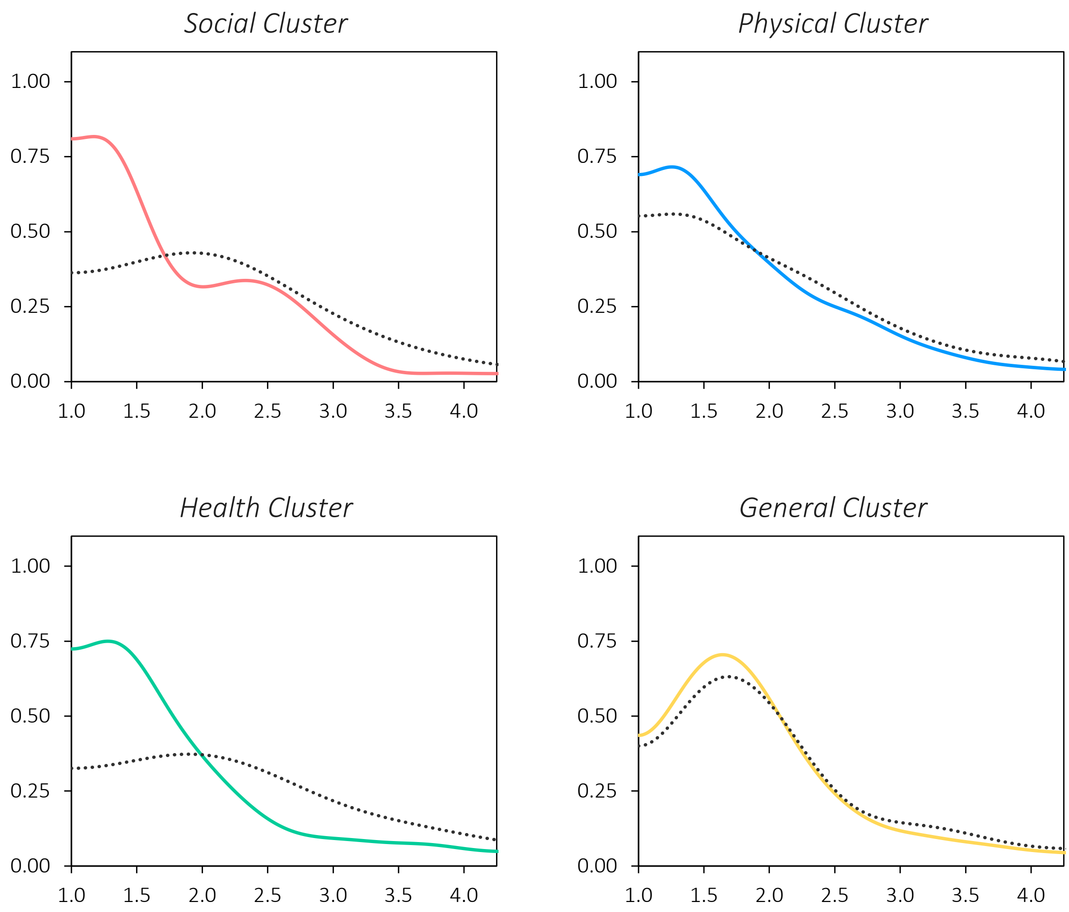


Fig. 3.4: DENSITY ESTIMATES OF BIAS-CORRECTED EFFICIENCY SCORES BY CLUSTER

Notes: Densities refer to estimates derived from an individual cluster frontier (solid line) or from one global frontier (dotted line).

In addition to Table 3.3, density estimates of bias-corrected efficiency scores are visualised in Figure 3.4. Three observations are worth emphasising here. First, and in line with our previous findings, switching from a global to an intra-cluster frontier affects universities focused on social and health sciences most significantly. In the latter scenario, more probability mass becomes assigned towards unity, which is partly due to the relatively high reduction in sample size. Second, all distributions are right-skewed, which marks a frequently expected outcome in efficiency contexts, and leptokurtic. Third, we observe a wide range of efficiency estimates including some extreme values, which indicates not only considerable heterogeneity among universities but also high discriminatory power of our model.

Country	Global Frontier		Social Cluster		Physical Cluster		Health Cluster		General Cluster	
	Mean	P50	Mean	P50	Mean	P50	Mean	P50	Mean	P50
Belgium	1.71	1.57	—	—	—	—	—	—	1.47	1.43
Czech Republic	2.10	2.00	—	—	1.93	1.79	—	—	1.74	1.79
Finland	2.45	2.32	2.09	2.08	1.96	1.76	—	—	2.28	2.23
Germany	3.27	3.12	2.48	2.52	2.72	2.55	3.67	3.74	3.08	3.09
Ireland	1.80	1.80	—	—	—	—	—	—	1.71	1.63
Italy	2.09	1.76	4.22	3.64	1.95	1.69	1.47	1.39	1.71	1.62
Lithuania	5.64	4.55	—	—	3.79	3.64	—	—	—	—
Netherlands	1.77	1.77	—	—	1.34	1.25	—	—	1.61	1.76
Norway	2.65	2.59	—	—	—	—	—	—	2.39	2.31
Poland	1.50	1.39	1.21	1.18	1.43	1.33	1.45	1.41	1.40	1.35
Portugal	1.91	1.81	—	—	1.57	1.20	—	—	1.72	1.79
Sweden	2.96	2.87	—	—	2.38	2.21	2.11	2.05	2.42	2.20
Switzerland	2.82	2.36	—	—	1.98	1.35	—	—	2.48	2.20
United Kingdom	1.92	1.79	1.78	1.49	1.84	1.76	1.44	1.33	1.72	1.68
Sample	2.33	1.99	2.04	1.64	2.05	1.75	2.22	1.67	2.03	1.81

Tab. 3.4: MEAN AND MEDIAN BIAS-CORRECTED EFFICIENCY SCORES BY COUNTRY AND CLUSTER
Notes: Efficiencies scores referring to less than three institutions are not reported. Malta and Cyprus are left out for this reason.

Exploring efficiency levels from a national perspective reveals further insights. According to Table 3.4, mean and median efficiency scores show substantial variation across Europe. The group of top-performing countries mainly comprises Belgium, the Netherlands, and Poland. Relatively high efficiency levels are also reached by universities located in the Czech Republic, Poland, Ireland, and the UK, whereas Scandinavian universities generally offer more room for improvement. Apart from these general patterns, there are cluster distinctions that are worthy of note. Italy, for instance, achieves high efficiency in the health sciences cluster, but clearly lags behind in the social sciences cluster. Interestingly, the reverse picture emerges with regard to Germany despite its overall greater levels of inefficiency. Lastly, Switzerland appears more efficient in the physical sciences cluster than in the general cluster.

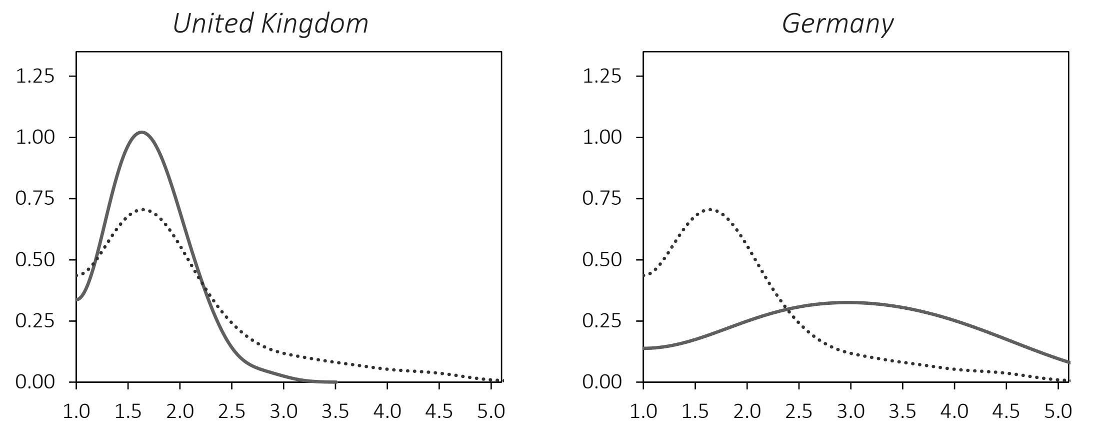


Fig. 3.5: DENSITY ESTIMATES OF BIAS-CORRECTED EFFICIENCY SCORES BY COUNTRY, GENERAL CLUSTER

Notes: Densities refer to a single country (solid line) or to the overall cluster (dotted line).

To offer an outlook beyond measures of central tendency, we take a closer look at the general cluster and, more specifically, at universities from Germany and the UK, which represent its largest subgroups. Based on the density estimates depicted in Figure 3.5, both countries can be considered to differ not only in mean or median efficiency but also in terms of within-country variation. Clearly, the German university landscape reveals a lot more heterogeneity than its British counterpart does. Upon examining the remaining countries, it seems difficult to state a general rule. Yet the illustrated examples seem to indicate that high mean efficiency scores are usually accompanied by greater variation. Additional density plots for single countries are provided in Appendix B.4.

3.6 Identification of Efficiency Drivers

3.6.1 Model Specification

From a policy perspective, detecting inefficiencies in public institutions can only be seen as an intermediate step. The focus of this section will therefore be placed on identifying efficiency drivers, knowledge of which may prove helpful for designing reasonable policy measures to promote higher education efficiency. Overall, we consider the following model specification

$$ln(\tilde{\theta}_{it}) = \beta_0 + \beta_1 \, ln(Size_{it}) + \beta_2 \, Herfindahl_{it} + \beta_3 \, Thirdparty_{it} + \beta_4 \, Fees_{it} + \gamma X'_{it} + \alpha_i + \delta_t + \epsilon_{it}, \tag{3.11}$$

which relates university i's efficiency estimate in year t to various factors expected to be of influence. In particular, we are interested in the potential effects of university size approximated by the number of students ($Size$), subject specialisation calculated as a Herfindahl index ($Herfindahl$), and funding composition characterised by the share of current revenues raised through third-party funds ($Thirdparty$) and student fees ($Fees$).

Moreover, we include a set of year dummies (δ) to control for time fixed effects, a set of country dummies (α) to account for country fixed effects, and further controls (X) related to employee structure, institutional design, and regional productivity. Summary statistics on all covariates by cluster along with more precise descriptions are provided in Table 3.5. Note that these data are derived from ETER except for GDP, which originates from Eurostat and the Swiss Federal Statistical Office.

Among our variables of interest, $Size$ permits investigating potential economies of scale in higher education. From a theoretical standpoint, large universities might benefit from higher utilisation of various assets. These could include shared research infrastructure, e.g., production plants or computing centres that typically require considerable initial investments, but also educational facilities such as libraries. Moreover, advancements in information technology could lead to a reduced demand of interpersonal relations in teaching hence expanding the range of decreasing unit costs presumably in favour of large institutions that tend to offer lectures for greater student numbers. However, administrative tasks potentially are a source of diseconomies of scale since organisational costs are expected to increase disproportionately with size. In view of these opposing arguments, it is understandable that the empirical literature has not yet reached a consensus on this matter (Bonaccorsi, Daraio, & Simar, 2006).

We further aim to shed light on economies of scope by including $Herfindahl$ as a measure of subject specialisation in our model. Based on ETER's distinction of 11 fields of study, this index ranges from 0.1, if students are equally distributed across fields, to 1.0, if students belong to only one field. Even though the Herfindahl index rests upon student numbers, it largely resembles our clustering results. As can be seen from Table 3.5, specialised clusters are characterised by higher index values hereby providing a first indication of the robustness of our approach. Whether efficiency benefits from specialisation or diversification in subject coverage is hard to answer on theoretical grounds. Turning to empirical studies, the overall picture remains mostly unclear. According to Daraio et al. (2015a), specialisation enhances academic efficiency, whereas results from Agasisti and Wolszczak-Derlacz (2016) as well as Wolszczak-Derlacz (2017) point to the contrary, i.e., the presence of economies of scope. Yet another conclusion is derived by Bonaccorsi et al. (2006), who reject any statistically significant relation.

Lastly, our interest lies in evaluating if differences in funding structure are related to university efficiency. Although external funding has become an increasingly central revenue source for European universities, empirical evidence on its performance impact remains relatively scant. Still, we expect universities with larger proportions of third-party funds to be more efficient given that previous studies by Wolszczak-Derlacz and Parteka (2011),

Variable	Sample	Social Cluster	Physical Cluster	Health Cluster	General Cluster
GDP –	*Regional gross domestic product per capita according to NUTS 2 classification*				
P5	16,005	18,802	14,048	17,180	17,180
Mean	32,257	46,018	26,887	38,988	30,480
P95	48,400	159,662	46,954	157,583	47,858
Multisite –	*Binary variable indicating campuses outside a university's main location*				
P5	0.00	0.00	0.00	0.00	0.00
Mean	0.24	0.20	0.25	0.19	0.26
P95	1.00	1.00	1.00	1.00	1.00
Hospital –	*Binary variable indicating the presence of a university hospital*				
P5	0.00	0.00	0.00	0.00	0.00
Mean	0.29	0.04	0.07	0.43	0.47
P95	1.00	0.00	1.00	1.00	1.00
Size –	*Number of enrolled students at ISCED levels 5-7*				
P5	3,059	1,343	3,043	1,608	6,928
Mean	17,461	11,029	15,575	11,921	21,882
P95	38,515	22,945	35,798	30,027	48,150
Herfindahl –	*Herfindahl index based on enrolled students at ISCED levels 5-7 by subject (in %)*				
P5	14.24	15.62	15.34	16.27	13.77
Mean	26.95	37.25	31.17	37.75	18.75
P95	75.27	97.58	76.55	99.80	27.51
Prof –	*Proportion of full professors amongst employees (in %)*				
P5	2.22	1.39	3.03	1.74	2.42
Mean	6.54	6.09	6.91	6.42	6.49
P95	11.32	10.34	11.44	12.12	11.40
Female –	*Proportion of women amongst full professors (in %)*				
P5	7.69	9.60	4.92	6.67	12.29
Mean	22.50	31.54	17.39	24.90	22.80
P95	40.00	50.00	30.38	50.00	36.36
International –	*Proportion of foreigners amongst academic employees (in %)*				
P5	1.74	4.97	1.11	1.04	1.86
Mean	16.17	19.54	14.72	14.15	16.37
P95	41.16	55.70	45.87	39.60	39.72
Thirdparty –	*Proportion of current revenues raised through third-party funds (in %)*				
P5	1.32	0.66	1.16	1.87	2.21
Mean	17.47	11.65	21.16	20.90	16.43
P95	40.34	36.69	42.32	50.06	35.83
Fees –	*Proportion of current revenues raised through student fees (in %)*				
P5	0.13	0.58	0.12	0.05	0.14
Mean	23.02	41.39	14.52	16.90	23.57
P95	69.36	76.05	49.05	68.50	67.71

Tab. 3.5: DESCRIPTION AND SUMMARY STATISTICS ON COVARIATES BY CLUSTER

Notes: Financial data are converted into real PPP EUR (2014 = 100). Breakdown of employee structure is based on headcounts.

Agasisti and Wolszczak-Derlacz (2016), and Wolszczak-Derlacz (2017) discover a negative correlation between the share of core funding and university performance in cross-country contexts. We further extend this strand of research by including the share of student fees, which allows us to disentangle the overall effect of external funding into two separate components. Although employing a parametric model, Bolli et al. (2016) pursue a similar approach and conclude that different mechanisms are potentially in play for these funding sources. More precisely, the share of tuition fees is found to decrease university efficiency, while the opposite is revealed about international public funds.

Apart from investigating a comprehensive set of efficiency drivers, our methodological framework is in particular designed to uncover differences between fields. As indicated by our clustering analysis, universities likely operate under varying technological constraints, which casts doubt on assuming that covariates exert identical effects throughout the subject spectrum. For instance, multidisciplinary work could be of different value across fields. Instead of jointly testing for economies of scope, we thus favour evaluating clusters on an individual basis.

3.6.2 Results

The results of the regression analysis are reported in Table 3.6. In line with the previous section, our focus is twofold. We present our preferred approach that relies on cluster-segmented estimations but also contrast it with the pooling approach, which derives its efficiency estimates from a global technology frontier. It should hereby be kept in mind that the dependent variable constitutes an inefficiency rather than an efficiency measure. Coefficient estimates with a negative sign therefore indicate efficiency-enhancing effects, whereas a positive sign corresponds to efficiency-decreasing effects.

The first result is indeed not linked to a single variable but related to the overall effect heterogeneity, which we find to take various forms. For instance, specialisation supposedly increases university efficiency in the pooled model as indicated by the significant and negative coefficient of $Herfindahl$. However, the segmented approach solely confirms this effect in case of the physical cluster. In a similar vein, a higher share of foreigners amongst academic employees ($International$) is associated with lower efficiency in the pooled model. Not only does this notion appear too general in view of the segmented analysis, it might potentially be misleading. While the effect points to the same direction within the social cluster, universities in the health cluster seem to benefit from increasing levels of internationalisation. Furthermore, the general cluster emerges as an exception regarding the effects of $Female$ in the sense that neither a negative nor a significant relation between the share of female full professors and efficiency can be confirmed.

Variable	Global Frontier	Social Cluster	Physical Cluster	Health Cluster	General Cluster
Natural logarithm of bias-corrected efficiency score as dependent variable					
ln(GDP)	- 0.0412	- 0.1500 **	0.1151	0.0923	- 0.0735 *
	(0.0237)	(0.0466)	(0.0706)	(0.0746)	(0.0362)
Multisite	0.0275	0.0646	0.1108 *	0.1947	0.0095
	(0.0203)	(0.0534)	(0.0484)	(0.1091)	(0.0208)
Hospital	0.0867 ***	- 0.2130	- 0.0190	0.2849 ***	0.1065 ***
	(0.0197)	(0.1302)	(0.0624)	(0.0646)	(0.0233)
ln(Size)	- 0.2843 ***	- 0.2909 ***	- 0.3985 ***	- 0.1034 *	- 0.1737 ***
	(0.0127)	(0.0469)	(0.0268)	(0.0484)	(0.0199)
Herfindahl	- 0.0029 ***	0.0018	- 0.0033 **	- 0.0026	- 0.0012
	(0.0006)	(0.0020)	(0.0010)	(0.0016)	(0.0029)
Prof	- 0.0212 ***	- 0.0255 **	- 0.0217 **	- 0.0669 ***	- 0.0162 **
	(0.0035)	(0.0085)	(0.0084)	(0.0141)	(0.0052)
Female	0.0040 ***	0.0073 ***	0.0066 **	0.0131 ***	- 0.0014
	(0.0010)	(0.0021)	(0.0024)	(0.0031)	(0.0016)
International	0.0045 ***	0.0043 *	0.0016	- 0.0227 **	0.0006
	(0.0011)	(0.0020)	(0.0033)	(0.0069)	(0.0019)
Thirdparty	- 0.0073 ***	- 0.0130 ***	- 0.0074 ***	- 0.0072 *	- 0.0071 ***
	(0.0009)	(0.0033)	(0.0017)	(0.0036)	(0.0014)
Fees	0.0001	0.0072 ***	- 0.0040 *	- 0.0267 ***	- 0.0029 **
	(0.0008)	(0.0019)	(0.0018)	(0.0035)	(0.0011)
No. of observations	1,285	182	305	143	655

Tab. 3.6: REGRESSION RESULTS

Notes: Results are obtained from 1,000 bootstrap repetitions. Constants as well as time and country dummies are included but not reported. Bootstrap standard errors are in parentheses. * p < 0.05, ** p < 0.01, and *** p < 0.001.

So far, we have confined attention to effects that, although significant according to the pooled model, do not withstand cluster-specific examination. In addition to this dimension of effect heterogeneity, there is a second group of variables whose influence on efficiency might be overlooked without further scrutiny. Among them, the share of student fees certainly stands out. While the pooled model rejects any notable impact, we observe significant coefficients of $Fees$ in each cluster. More importantly, universities that rely more heavily on student fees are considered more efficient in the physical, health, and general cluster. On the contrary, the opposing relation is found in the social cluster. A similar pattern, though to a lesser extent, becomes visible with regard to regional gross domestic product per capita (GDP) and the indicator for multisite institutions ($Multisite$); i.e., coefficients turn significant only in a single cluster.

To draw an interim conclusion, several efficiency drivers appear to differ in relevance between subject clusters. Yet some variables show consistent effects. More specifically, our analysis indicates that efficiency is in general positively related to the share of full professors ($Prof$), the share of third-party funding ($Thirdparty$), and university size ($Size$). A closer look at the magnitude of these coefficient estimates further reveals their economic significance. On average, we would expect inefficiency to decrease in a range between 0.7 and 1.3% if the share of third-party funding increased by 1 percentage point. In comparison, raising the share of full professors by an equal margin should lower inefficiency by 1.6 to 6.7%. It is worth noting, however, that the latter adjustment might require greater efforts given that personnel structures are supposedly less flexible than revenue compositions.[35] Turning to the impact of institutional size on inefficiency, we estimate point elasticities between -0.1 and -0.4. Despite some variation in effect sizes, we hereby provide evidence for economies of scale in higher education and additionally infer that avenues for efficiency improvement exist on both the personnel and financial level.

3.7 Robustness Analysis

Within this section, we provide further evidence probing the robustness of our results. To be more precise, we report a series of model checks that involve variations in peer-group construction, output selection, and regression design.

The initial clustering solution marks the starting point for these analyses. As a first step, we assess the quality of this solution by determining silhouette coefficients for each university. Following Rousseeuw (1987), silhouette coefficients indicate how well (data) objects have been classified by a given partitioning. In more concrete terms, they are derived by comparing an object's proximity to its cluster members with the proximity to the members of its neighbouring cluster, i.e., the cluster with the highest proximity among those the object is not part of. In general, silhouette coefficients can range between -1 and 1, with higher values denoting stronger structures. Consistent with the K-means algorithm, we rely on squared Euclidean distance in subject space to measure proximity between universities. Silhouette coefficients are illustrated in descriptive form in Table 3.7 and depicted graphically in Appendix B.5. Two aspects stand out from these displays. First, each cluster is characterised by an average silhouette coefficient higher than 0.5, which is commonly referred to as a threshold for reasonable cluster structures. Second, however, some universities with silhouette coefficients close to 0 appear to be classified rather vaguely.

[35] Within our regression sample, $Prof$ indeed shows less variation than $Thirdparty$, which is reflected by standard deviations of 2.9 and 13.3 percentage points, respectively.

Cluster	N	Min	P25	P50	Mean	P75	Max
Social	57	-0.123	0.341	0.666	0.535	0.758	0.819
Physical	140	0.105	0.557	0.819	0.696	0.859	0.882
Health	49	-0.042	0.289	0.661	0.524	0.744	0.797
General	204	0.024	0.456	0.611	0.575	0.725	0.828
Sample	450	-0.123	0.470	0.665	0.602	0.783	0.882

Tab. 3.7: SUMMARY STATISTICS ON SILHOUETTE COEFFICIENTS BY CLUSTER

The second observation hardly comes as a surprise. We may find the European university landscape to feature four subject clusters, yet it is to be expected that not all universities fit into this classification. Some institutions obviously occupy niches, which suggests that cluster boundaries are partly fluid. As a consequence, it could be suboptimal to compare universities solely to their cluster members (in some cases). Instead, certain universities, especially if near the boundaries, might possess relevant peers outside their own cluster. We thus extend our approach by constructing tailored peer-groups for each university, which are not bound by cluster affiliation but purely based on proximity in subject space. The advantage of this approach, which we term nearest neighbourhood approach, clearly lies in greater homogeneity. It does, however, require 450 bootstrap DEA estimations per year and hence call for markedly more computing resources.

Cluster	Social Peers	Physical Peers	Health Peers	General Peers	Δ Distance
Social	67.95%	1.19%	3.41%	27.44%	-32.41%
Physical	0.24%	83.00%	0.00%	16.76%	-20.87%
Health	3.27%	0.00%	61.05%	35.67%	-34.92%
General	2.68%	11.60%	4.47%	81.24%	-25.52%

Tab. 3.8: AVERAGE PEER-GROUP COMPOSITION BY CLUSTER, NEAREST NEIGHBOURHOOD APPROACH
Notes: Δ Distance refers to the change in average squared Euclidean distance between peers resulting from peer-group construction free of cluster constraints.

Peer-group compositions based on our modified approach are reported in Table 3.8. For comparability reasons, we stick to identical peer-group sizes as in our baseline model, so that universities in the social cluster, for instance, are assessed relative to their 56 closest peers. The average distance between peers is reduced by a substantial margin of 21 to 35% as we switch to the nearest neighbourhood approach mainly due to general cluster universities that frequently enhance peer-groups of universities from specialised clusters. Bias-corrected efficiency scores are then estimated based on these individual peer-groups and regressed on our set of covariates, which leads to the results in Table 3.9. In line with Section 3.6, we again observe considerable effect variation across subject fields. Although efficiency drivers remain (in)significant in the majority of cases, some previous findings need to be refined. To be more concrete, institutional size and the share of third-party funds are no longer strictly linked to higher efficiency given that these effects are not persistent in the health cluster. Similarly, we notice that the share of full professors affects efficiency less clearly within the social cluster (p-value of 0.07).

Variable	Social Cluster	Physical Cluster	Health Cluster	General Cluster
Natural logarithm of bias-corrected efficiency score as dependent variable				
ln(GDP)	- 0.0819	≙ 0.1000	≙ 0.0935	≙ - 0.0687 *
	(0.0587)	(0.0746)	(0.0679)	(0.0350)
Multisite	0.1443 *	≙ 0.1031 *	0.2350 *	≙ 0.0087
	(0.0653)	(0.0506)	(0.1012)	(0.0208)
Hospital	≙ - 0.3071	≙ - 0.0927	≙ 0.3690 ***	≙ 0.0887 ***
	(0.1605)	(0.0625)	(0.0578)	(0.0216)
ln(Size)	≙ - 0.2254 ***	≙ - 0.3967 ***	- 0.0316	≙ - 0.2352 ***
	(0.0597)	(0.0262)	(0.0425)	(0.0187)
Herfindahl	≙ 0.0003	≙ - 0.0033 **	≙ - 0.0005	≙ - 0.0027
	(0.0027)	(0.0011)	(0.0015)	(0.0028)
Prof	- 0.0199	≙ - 0.0295 ***	≙ - 0.0398 **	≙ - 0.0175 ***
	(0.0112)	(0.0086)	(0.0124)	(0.0051)
Female	- 0.0007	≙ 0.0094 ***	≙ 0.0135 ***	≙ - 0.0018
	(0.0029)	(0.0024)	(0.0029)	(0.0016)
International	0.0006	≙ 0.0029	≙ - 0.0239 ***	≙ 0.0008
	(0.0027)	(0.0034)	(0.0063)	(0.0018)
Thirdparty	≙ - 0.0125 **	≙ - 0.0082 ***	- 0.0040	≙ - 0.0067 ***
	(0.0039)	(0.0017)	(0.0030)	(0.0014)
Fees	≙ 0.0111 ***	≙ - 0.0043 *	≙ - 0.0208 ***	≙ - 0.0022 *
	(0.0025)	(0.0019)	(0.0028)	(0.0011)
No. of observations	182	305	143	655

Tab. 3.9: REGRESSION RESULTS, NEAREST NEIGHBOURHOOD APPROACH

Notes: Results are obtained from 1,000 bootstrap repetitions. Constants as well as time and country dummies are included but not reported. ≙ marks coefficient estimates that stay either significant or insignificant relative to Table 3.6. Bootstrap standard errors are in parentheses. * $p < 0.05$, ** $p < 0.01$, and *** $p < 0.001$.

An additional robustness check refers to altering the concept of university efficiency. To account for increasing levels of technology transfer activities, we now opt for granted patents (instead of publications) to complement citations and graduates as a third output component.[36] As pointed out by Geuna and Nesta (2006), patenting efforts tend to be concentrated in the areas of life sciences and technology, which could partially explain why many European universities did not obtain any patents in the past (OECD, 2003). Our data

[36] Patent records were derived from Scopus, which contains data from five major patent offices (Elsevier, 2017).

generally confirm this picture as we find patent engagement to vary considerably across clusters and, besides, note that a number of universities received zero patents in certain years. To ensure a comparison of appropriate peers, we thus confine attention to the subset of our sample that recorded at least one patent in every year between 2011 and 2014. Overall, this leaves us with 281 universities with measurable pursuit of applied forms of research output.

All of these universities serve as potential peers as part of the nearest neighbourhood approach. The subsequent regression, however, requires reasonable sample sizes, which leads us to focus on the physical cluster with 79 and the general cluster with 163 universities.[37] Results are documented in columns 4 and 5 of Table 3.10. With regard to the consistent effects reported before, the positive relation between efficiency and both university size and third-party funding shares can be confirmed. In contrast, the share of full professors turns insignificant. Interestingly, there appear to be different reasons for this finding. While it is caused by model shift in the physical cluster, it is attributable to sample composition in the general cluster (see columns 2 and 3). In other words, a higher share of full professors neither improves efficiency in the publication nor in the patent model if we review patent-active general cluster universities. One might assume that these universities rely more on well-run administrations as they tend to be larger and supposedly more complex. In comparison, patent-active physical cluster universities seem to benefit from a higher full professor share as long as we refer to an efficiency concept that builds upon publications instead of patents.

From the standpoint of generalisability, these additional checks allow us to conclude that institutional size and the ability to seek external funding are the main factors to impact university efficiency. With the exception of the health domain, both variables are consistently identified as efficiency enhancing. To allay concerns about the direction of causality, and to account for possibly delayed effects, further regression analyses with time-lagged covariates are presented in Appendix B.6. Irrespective of model choice, we find the stated interpretation encouraged by these estimations.

[37] Cluster sizes still constitute the reference points, so that groups of 79 and 163 are used to assess universities from the physical and general cluster, respectively.

Variable	**Physical Cluster** Publication Model	**General Cluster** Publication Model	**Physical Cluster** Patent Model	**General Cluster** Patent Model
Natural logarithm of bias-corrected efficiency score as dependent variable				
ln(GDP)	≙ - 0.0376	≙ - 0.1691 ***	0.1564	- 0.2059 ***
	(0.0810)	(0.0404)	(0.1080)	(0.0458)
Multisite	≙ 0.1681 **	≙ - 0.0454	0.1251	- 0.0960 ***
	(0.0608)	(0.0257)	(0.0795)	(0.0286)
Hospital	≙ 0.0077	≙ 0.1582 ***	0.1342	0.1870 ***
	(0.0553)	(0.0261)	(0.0758)	(0.0286)
ln(Size)	≙ - 0.4847 ***	≙ - 0.2863 ***	- 0.5433 ***	- 0.2946 ***
	(0.0417)	(0.0255)	(0.0555)	(0.0280)
Herfindahl	≙ - 0.0048 **	≙ - 0.0018	0.0015	0.0010
	(0.0017)	(0.0035)	(0.0019)	(0.0038)
Prof	≙ - 0.0380 **	- 0.0005	- 0.0230	- 0.0075
	(0.0129)	(0.0061)	(0.0175)	(0.0066)
Female	≙ 0.0158 ***	≙ 0.0002	0.0127 *	0.0003
	(0.0043)	(0.0021)	(0.0058)	(0.0022)
International	≙ 0.0020	0.0052 *	0.0077	0.0052 *
	(0.0036)	(0.0021)	(0.0047)	(0.0023)
Thirdparty	≙ - 0.0132 ***	≙ - 0.0072 ***	- 0.0167 ***	- 0.0083 ***
	(0.0026)	(0.0016)	(0.0034)	(0.0018)
Fees	- 0.0058	- 0.0011	- 0.0070	- 0.0012
	(0.0037)	(0.0013)	(0.0048)	(0.0015)
No. of observations	176	534	176	534

Tab. 3.10: REGRESSION RESULTS OF MODEL VARIANTS, NEAREST NEIGHBOURHOOD APPROACH
Notes: Results are obtained from 1,000 bootstrap repetitions. Constants as well as time and country dummies are included but not reported. ≙ marks coefficient estimates within the publication model that stay either significant or insignificant relative to Table 3.6. Both models are estimated on identical samples, i.e., the group of patent-active universities in each cluster. Bootstrap standard errors are in parentheses. * $p < 0.05$, ** $p < 0.01$, and *** $p < 0.001$.

3.8 Conclusion

The present paper studies the relative efficiency of 450 European universities between 2011 and 2014. Our approach is built on the notion that the higher education landscape in Europe is too diverse to be considered one homogeneous peer-group. In particular, differences in subject focus prove indicative for numerous institutional characteristics. We uncover these systematic patterns by means of clustering techniques and identify four groups of universities that either possess a balanced subject profile or lay clear emphasis on social sciences, physical sciences, or health sciences. Given that efficiency estimation naturally relies on relative assessment, it is crucial to differentiate between these distinct groups. Otherwise, one would run the risk of comparing apples and pears. To illustrate this point, we discover that health cluster universities incur expenditure per student levels that are, on average, almost four times higher than of social cluster universities.

We address homogeneity concerns firstly by employing intra-cluster efficiency frontiers. In an extension to this approach, we secondly construct individual peer-groups for each university based on subject space proximity. With bias-corrected efficiency scores at hand, we direct attention to potential efficiency drivers, which are investigated within a subsequent regression analysis. It becomes evident that different, even partly opposing, mechanisms are in play depending on the cluster under review. Yet institutional size and the ability to seek external funding are largely found to be efficiency enhancing. Apart from the health cluster, inefficiency is expected to fall by 6.7 to 16.7% if the share of third-party funds increased by 10 percentage points and by 1.7 to 5.4% if universities were to expand their capacities by 10%.

Overall, this paper underlines the high degree of diversity in Europe's higher education sector and provides a framework for further in-depth studies. However, our analyses are not without limitations. Incorporating teaching quality would certainly complement our efficiency perception, yet it is hard to think of reliable measures for this domain. Despite the time-lag regression design, it would also be beneficial to adopt additional methods dedicated towards causal inference. Lastly, future research may emphasise the distinction between private and public sources of external funding to broaden the understanding of university efficiency beyond the findings presented in this study.

4 Competitive Funding in Academia: Back to the Bags of Gold

4.1 Introduction

In reference to "The Parable of the Bags of Gold" (Matthew 25: 14-30 NIV), Merton (1968) coined the term "Matthew effect" for self-reinforcing effects in the scientific reward and communication systems. More concretely, he discovered that eminent scientists tend to receive disproportionally great recognition for their work. In a more general sense, early accomplishments are likely to lead to even more success in the future. While studies on this topic are traditionally concerned with intangible capital (such as peer recognition), the present study sheds light on the accumulation of financial resources in academia. The concrete setting of my study is provided by a large German funding program which aims at improving the quality of teaching in higher education. By exploring the distribution of these funds, I specifically investigate if the amount of funding acquired in years prior to the program is associated with present funding success.

The program in question is the "Quality Pact for Teaching" (*Qualitätspakt Lehre, QPT*). The QPT was jointly initiated by the federal and state government(s) in 2010 to promote the quality of teaching in higher education. Until 2020, a total of almost 2 billion euros was scheduled for this purpose and first allocated in 2011/12 with a possible extension period starting in 2017. Competition for QPT grants was open not only for universities but also for universities of applied sciences and colleges of arts and music. According to the funding decision, institutions of all three types are represented among the successful applicants and are thus given the opportunity to improve their studying conditions through various measures (e.g., additional professorships, targeted training for academic personnel, or the design of new teaching concepts) (Joint Science Conference, 2011a, 2011b). Despite the major financial commitment, the QPT has only sporadically been covered by the scientific literature. Moreover, most existing studies either confine attention to single institutions (Deicke, Gess, & Rueß, 2014; Koch & Vogt, 2015) or limit their scope to specific subgroups (Bloch, Mitterle, Rennert, & Würmann, 2018; Dehne, Lucke, & Schiefner-Rohs, 2017).[38] In contrast, I pursue a higher-level approach that addresses the higher education sector in its entirety and therefore aims to fill a current research gap. In addition, I contribute to the literature by analysing the QPT with econometric tools, whereas hitherto results are mainly derived through qualitative methods.

[38] One exception is the official evaluation of the QPT, which was commissioned to the Center for Quality Assurance and Development of the University of Mainz in collaboration with the Prognos AG. Yet, their assessment focusses on conceptual and procedural aspects (ZQ Uni Mainz & Prognos AG, 2016, 2018).

The QPT is an important research topic for multiple reasons. At first, it can be expected that insights from the QPT have wide-ranging implications given that the program stands representative for a paradigm shift in academic governance. Starting in the 1980s, most of continental Europe started introducing more market forces into their higher education systems to increase efficiency and yield higher societal returns (Geuna, 1999; Partha & David, 1994). Germany may be viewed as a latecomer to New Public Management, yet has also implemented several measures to spur competition among and within universities (Schimank & Lange, 2009). A growing dependence on third-party funding plays a central role in this respect (Osterloh & Frey, 2008). The QPT is indeed more than a byproduct of this trend. Instead, it can be seen as an evolutionary step since the program marks the first time that a considerable amount of grants is competitively allocated for teaching purposes.[39] In addition to the novel research setting, the QPT is also worth a closer analysis due to its ambiguous reception. The official evaluation presents the program in a positive light and repeatedly stresses the diversity of the 178 funded projects (ZQ Uni Mainz & Prognos AG, 2016). The Federal Court of Auditors (2019), however, expresses substantial concerns. According to its verdict, the QPT largely fails to meet its goal of impacting the academic landscape at a broad range.

The QPT grant allocation visibly draws a line between winners and losers of the funding competition. This binary classification enables employing (non-)linear regression models that relate QPT outcome to a wide range of institutional characteristics. One of the main benefits of these models lies in their intuitive effect interpretation, i.e., each explanatory variable can be considered to affect the probability of QPT success. From my estimations, I can infer that past third-party funding volume (per professor) is a particularly strong predictor of success, thus underlining the Matthew logic. Exploring heterogeneity of the effect offers another key insight, namely that universities of applied sciences and colleges of arts and music experience effects of a notably higher magnitude. A raise of third-party funds by half a standard deviation is associated with a probability plus of 6.6 percentage points for the former and 17.0 percentage points for the latter type of institution, while the university effect adds up to 2.4 percentage points (assuming median profiles). Pronounced effects at the lower end of the funding spectrum illustrate my perception of the Matthew effect in that it takes the form of a learning curve. In this sense, it might be worried that the QPT's application process (unintentionally) penalised institutions that were generally unaccustomed to grant competition, which would certainly contradict the image of a well-balanced funding scheme. Apart from the Matthew result, I find a positive link between QPT chances and both student numbers and STEM focus, while, on the contrary, neither variation in student fees nor excellence status appears to affect the funding probability.

[39] The "Competition for Teaching Excellence" (*Wettbewerb Exzellente Lehre)* preceded the QPT with a similar concept, but on an evidently smaller scale (10 million versus 2 billion euros) (Stifterverband, 2009).

Overall, I conclude that procedural knowledge can be viewed as an important facet of the Matthew effect. Supporting evidence for this notion even emerges from the official QPT evaluation, in which grantees report on their large progress in meeting the specific demands of third-party funded projects (ZQ Uni Mainz & Prognos AG, 2018, pp. 65–66). Such statements warrant attention since they point to a growing divide in administrative expertise that stands in diametric opposition to the QPT's intentions. My general policy recommendation is straightforward: Future programs that rely on broad-scale competition should devote particular effort to ensuring that selection processes are merit-based and not obscured by experience and skilful grant writing.

Apart from extending the aforementioned QPT studies, the present paper contributes to three main strands of literature. First, it belongs to a class of studies that has followed Merton's (1968) seminal work. Evidence for Matthew effect channels (or, more generally, cumulative advantage) has been detected in virtually all branches of science. In fact, at its core, the Matthew effect represents more of an interrelated concept that applies to scientific outcomes such as publications and citations (Allison, Long, & Krauze, 1982; Larivière & Gingras, 2010), just as much as it leaves its mark on awards (Azoulay, Stuart, & Wang, 2014) and resources (Laudel, 2006), and thereby shapes the course of scientific careers (Petersen, Jung, Yang, & Stanley, 2011).[40] Second, my study adds to the literature that reviews higher education policies. With respect to Germany, this strand has largely been concerned with the impacts of the Excellence Initiative (e.g., Bruckmeier, Fischer, & Wigger, 2017; Menter, Lehmann, & Klarl, 2018) and student fees (e.g., Bruckmeier & Wigger, 2014; Fischer & Wigger, 2016; Hübner, 2012). Third, a monetary Matthew effect can also be understood as part of the literature on financial interdependences in higher education, which typically examines the relation between public and private sources of funding (e.g., Grimpe, 2012; Muscio, Quaglione, & Vallanti, 2013).

The remainder of this paper proceeds as follows. Section 4.2 sets the stage for my study by depicting the institutional context and the included data. Section 4.3 elaborates on the methodological approach and presents my results. Lastly, Section 4.4 offers a discussion of the main findings and concludes with a series of policy proposals and avenues for future research.

4.2 Institutional Context and Data

My dataset is primarily compiled from the German Federal Statistical Office (Destatis) and builds on two complementary publication series that are annually distributed. These are the non-monetary *(Fachserie 11, Reihe 4.3.1*) and monetary (*Fachserie 11, Reihe 4.3.2*) statistics on higher education. Given that both series cover data on an aggregate level, I requested special evaluations from Destatis to obtain selected data on the institutional

[40] These studies are only an excerpt of the literature. For a systematic review, see DiPrete and Eirich (2006).

level. I further complemented these data with information on the first two rounds of the Excellence Initiative that are publicly available. Finally, I drew on the QPT project database, which is hosted by the DLR Project Management Agency, to delimit the group of institutions that were granted with funding.[41]

Taken together, I utilise these data sources to identify structural characteristics that are predictive of QPT funding success. In light of this research strategy, I first address some conceptual aspects. The QPT consists of two separate funding lines that either offer support for single institutions or institutional consortia. The former funding line can mainly be regarded as the QPT's core since it encompasses 159 projects, while the latter funding line covers 19 joint projects. In the course of the analysis, I will solely focus on single project funding not only because of its dominance from a volume standpoint, but also because I intend to avoid ambiguity in assessing structural variables. Stated differently, it would certainly require an additional model to evaluate joint projects, which ought to include data on, e.g., geographical proximity or institutional alignment. Furthermore, the QPT program announcement was issued by the German Federal Ministry of Education and Research (*Bundesministerium für Bildung und Forschung, BMBF*) on November 10, 2010. It declared that two application rounds would be conducted with respective submission deadlines in March and September 2011. Applicants were required to provide a project description, a financial plan, and a data-driven assessment of the institution's strengths and weaknesses in teaching. Given this timeline, I refer to 2009 as the base year when gauging the effects of structural variables. Yet, data from 2010 will also be incorporated into the robustness checks.

According to the BMBF (2010) announcement, QPT funds were open for higher education institutions under public ownership (with the exception of the two universities operated by the German Federal Armed Forces as I ascertained upon personal inquiry with the BMBF). Within the realm of eligibility, I managed to compile data for 250 institutions. Missing data prevent me to include another 14 institutions, two of which are among the QPT recipients. These are the Jade University of Applied Sciences and the University of Applied Sciences Emden/Leer that both emerged from an organisational split in 2009. In total, my dataset thus covers 157 of 159 institutions that eventually received support as part of the QPT single project funding line.

Table 4.1 illustrates the outcome of the data compilation process. In general, the German higher education landscape distinguishes between three main types of institutions. These are (regular) universities, universities of applied sciences, and colleges of arts and music (Hüther & Krücken, 2018), which represent 80, 118, and 46 observations of the sample,

[41] See https://www.qualitaetspakt-lehre.de/de/projekte-im-qualitatspakt-lehre-suchen-und-finden.php. If not stated differently, my QPT information are derived from this database.

Variable		P5	P50	Mean	P95	SD
QPT	*QPT single project funding*	0	1	0.628	1	0.484
West	*West Germany, including Berlin*	0	1	0.800	1	0.401
Abolition	*Tuition abolition upcoming*	0	0	0.372	1	0.484
University	*University*	0	0	0.320	1	0.467
Applied	*University of Applied Sciences*	0	0	0.472	1	0.500
Education	*University of Education*	0	0	0.024	0	0.153
Music	*College of Arts and Music*	0	0	0.184	1	0.388
Excellence	*Excellence Initiative success*	0	0	0.144	1	0.352
STEM	*Percentage of STEM students*	0	27.79	32.15	77.59	27.37
Size	*Number of students*	289	4,245	7,656	27,161	8,998
Teachratio	*Students per prof.*	13.14	43.45	49.61	94.88	32.01
Fees	*Fees per student*	0.00	409.86	398.22	843.48	336.39
Basic	*Basic funds per student*	4,284	7,485	10,517	25,055	9,001
Thirdparty	*Public third-party funds per prof.*	0	12,296	54,970	234,465	86,782

Tab. 4.1: SUMMARY STATISTICS ON HIGHER EDUCATION INSTITUTIONS

Notes: The sample consists of 250 institutions that were legally eligible to apply for QPT funding. This applies to all public institutions except for those operated by the German Federal Armed Forces. A further 14 institutions were dropped in the compilation process due to missing data. Professorships are expressed in FTE. Continuous variables refer to the year 2009.

respectively. I further broaden this classification scheme by delineating six institutions that, albeit being legally on par with universities, are characterised by a unique focus on educational sciences. They are accordingly referred to as universities of education (see Blömeke, 2004, for further information). From a territorial viewpoint, 80% of the sample are located in a western German state (including Berlin). Among them, I additionally mark institutions that belong to either Baden-Württemberg, Hamburg, or North Rhine-Westphalia. As Fischer and Wigger (2016) note, these states shared a similar timeline regarding their tuition fees policy. More concretely, tuition fees were in place when the QPT program was announced yet their future abolishment became foreseeable shortly afterwards.[42] I keep note of this circumstance by means of an indicator variable (termed abolition). Tuition fees sparked a widespread public and scholarly debate in Germany; however, at the time, the university sector was subject to further reforms. By initiating the Excellence Initiative, policy-makers visibly broke with the egalitarian tradition of higher education governance in Germany (Hartmann, 2006). During the first two rounds held in 2006 and 2007, the federal and state government(s) jointly provided 1.9 billion euros to fund successful projects over periods of five years. Universities were awarded in three

[42] In all three states, the abolition was to be expected due to a change of government. In practice, tuition fees were first removed in North Rhine-Westphalia (winter term 2011), followed by Baden-Württemberg (summer term 2012) and lastly Hamburg (winter term 2012) (Fischer & Wigger, 2016, Table 2).

different domains, i.e., Graduate Schools, Excellence Clusters, and Institutional Strategies that aim(ed) at promoting young scientists, top-level research in particular fields, and universities as a whole, respectively (German Research Foundation, 2013). In total, 36 universities received support across these areas, which is denoted by the Excellence indicator.

Finally, I characterise the higher education landscape based on selected cardinal data. Building on Destatis's subject classification system, I note that institutions, as of 2009, accommodate an average of 32% of their students in the fields of mathematics, natural sciences, and engineering. I subsume these students under the STEM label and report them separately mainly in view of their augmented dropout rates (Heublein, 2014).[43] In addition, I find that institutions comprise a mean of 7,656 students with a student-to-professor ratio of 50. From the monetary angle, basic funds cover the bulk of expenses. They amount to slightly over 10.5 thousand euros, while student contributions add up to nearly 400 euros, both on average and per student. Lastly, I emphasise the extent of third-party funds, which are raised from public sources – primarily the German Research Foundation, federal ministries, and the European Union. On average, these funds amount to 55 thousand euros per professor, which, for comparison, corresponds to 905 euros per student. Most of these statistics can indeed be expected to vary (markedly) between institutional types. This aspect will not be left unconsidered, but, due to methodological reasons, will be deferred to Section 4.3.3.

4.3 Determinants of Funding Success

4.3.1 Econometric Foundations

In the course of this section, I first lay the econometric foundations for the regression analysis mostly following Long (1997) before I turn to the model specification and the estimation results. To begin with, binary outcomes can typically be interpreted as the manifestation of a latent variable. In the present context, it is accurate to view the QPT's funding decision as the result of a committee evaluation given that almost all eligible institutions participated in the application procedure (Joint Science Conference, 2011b). Thus, let $Eval_i$ denote the evaluation outcome of institution i, which in theory can be thought to range from $-\infty$ to ∞. Despite being unobserved, $Eval_i$ may still be assumed to be linearly related to a vector X_i of observed covariates through

[43] For first-year bachelor students in 2008/09, Heublein (2014) reports an overall dropout rate of 28%. Yet STEM fields are more severely affected. At universities, 39% is recorded for mathematics and natural sciences and 36% for engineering. At universities of applied sciences, dropout is generally less frequent, but still at its peak in these fields with respective rates of 34% and 31%.

$$Eval_i = \beta\, X_i' + \epsilon_i, \tag{4.1}$$

where β denotes a vector of coefficients and ϵ_i represents a residual error term that is symmetrically distributed about zero, independent of X_i.[44] Furthermore, let QPT_i denote the actual funding outcome, which can take values of either one for success or zero for failure. The link between both random variables can then be formulated in the following way: If the committee comes to a positive assessment, i.e., $Eval_i > 0$, funding is granted so that $QPT_i = 1$. In the remaining cases where $Eval_i \leq 0$, however, funding is denied, implying that $QPT_i = 0$. Thus, the conditional probability of a positive funding outcome can be modelled as

$$Pr(QPT_i = 1 \mid X_i) = Pr(Eval_i > 0 \mid X_i). \tag{4.2}$$

After substituting Equation (4.1) and rearranging the inner inequality, the right-hand side of Equation (4.2) can be stated as $Pr(\epsilon_i > -\beta\, X_i \mid X_i)$ or equivalently as $Pr(\epsilon_i \leq \beta\, X_i \mid X_i)$ due to the symmetry of ϵ_i. Lastly, let the cumulative density function of ϵ_i be denoted by G. It follows that

$$Pr(QPT_i = 1 \mid X_i) = G[\beta X_i']. \tag{4.3}$$

This representation illustrates that the choice of G, which fulfils the role of a link function, can be derived from the distributional features of an underlying latent variable model. In presenting my main results, I opt for a standard logistic distribution of ϵ_i and therefore estimate a logit model.[45] However, Equation (4.3) can also be viewed from another angle. Irrespective of a latent variable, one can define a model where the probability of a positive QPT outcome is directly dependent on X_i. In this scenario known as the linear probability model (LPM), G becomes redundant or, more formally, resolves to the identity function. Due to its reduced complexity, the LPM is usually considered a useful counterpart to non-linear variants such as the logit model. Yet, the (effective) absence of a link function also has its downsides. Importantly, it cannot be assured that the predicted probabilities are constrained to the unit interval which can lead to biased coefficient estimates (Horrace & Oaxaca, 2006). I thus refer to the LPM rather as a complementary approach to the main logit model that, on a positive note, requires less functional assumptions.

Both models allow the estimation of marginal effects. In the LPM case, these are (simply) given by the ordinary least squares estimate of β. As for the logit model, the link function calls for consideration. To be more concrete, let x_{ik} and β_k denote the k-th element of X_i and β, respectively. If x_{ik} is continuous, its marginal effect is stated by $g[\beta X_i']\,\beta_k$ with

---

[44] The first element of X_i can universally be set to one to include a constant, which is mostly the case for binary response models. Moreover, each element of X_i can be a function, such as a logarithm, of the respective explanatory variable (Wooldridge, 2010, Chapter 15).

[45] Alternative model choices are discussed as part of the robustness checks at the end of Section 4.3.3.

g indicating G's derivative. If, in contrast, x_{ik} is binary, the marginal effect is akin to the difference of G being computed at two points separated by β_k. Either way, it becomes evident that marginal effects depend on the specific realisation of X_i and the estimate of β, which is obtained through maximum likelihood estimation in the logit case. I thus consider two types of marginal effects as part of the results section. First, I report the average over all observations, which yields the average marginal effect (AME). Second, I present marginal effects for (synthetic) institutions that take on median covariate values. In doing so, I illustrate how the funding chances of, e.g., the median college of arts and music or the median university are altered when a certain explanatory variable changes.

4.3.2 Model Specification

Based on the outlined methodology, I aim to identify institutional characteristics that played a decisive role for the QPT funding allocation. I specify the following model for this purpose:

$$\begin{aligned}\Pr(QPT_i = 1) = G\Big[&\beta_0 + \beta_1\, ln(Thirdparty_i) + \beta_2\, STEM_i + \beta_3\, ln(Fees_i) \\ &+ \beta_4\, Abolition_i + \beta_5\, ln(Fees_i) \times Abolition_i + \gamma X_i'\Big]\end{aligned} \tag{4.4}$$

which relates the conditional probability of funding success to a comprehensive set of explanatory variables.[46] Technically, I refrain from adding a time subscript to the model since the dependent variable lacks a longitudinal dimension, i.e., the initial QPT decision was a one-time event. Yet, it is important to note that Equation (4.4) implies a time lag as the QPT outcome was determined in the year 2011, while the (time-varying) explanatory variables refer to the year 2009.

Among the right-hand side variables, I draw a distinction between variables of primary interest and additional control variables, the latter of which are subsumed under X_i. As for the former category, I direct particular attention to the amount of public third-party funding as a possible reflection of the Matthew logic. If past funding volumes prove to be positively linked to QPT success, we would have empirical evidence for a self-enforcing monetary effect or, stated differently, "For whoever has will be given more, and they will have an abundance" (Matthew 25: 29 NIV). A strong argument in favour of this claim can be seen in the specific expertise regarding competitive funding schemes that well-funded institutions (potentially) acquire over time. Yet, on the contrary, it is worth noting that apart from minor exceptions third-party funds are traditionally targeted towards research activities (Marquardt, 2011). Seeking these grants could require different knowledge than needed in the QPT setting. Moreover, it even seems conceivable that institutions that are

[46] The notation is formally introduced in Section 4.3.1. Furthermore, note that the conditioning variables are omitted from the left-hand side of Equation (4.4) by way of notational convenience.

less research oriented and presumably more dedicated towards teaching might be in an advantageous position when it comes to designing innovative education concepts. Taken together, the relationship remains a priori inconclusive.

The second variable highlighted in Equation (4.4) is the share of students enrolled in STEM fields. STEM studies began drawing public attention at about the same when demand for a QPT-like funding program was first stated by the German Rectors' Conference (2007). More specifically, both low enrolment and high dropout rates in the STEM area were considered key contributing factors to a (looming) labour shortage. While this view was primarily supported by business-oriented research organisations (see e.g., Hetze, 2011, or Koppel and Plünnecke, 2009), it was countered by reports of the Federal Employment Agency (2011) or the DIW Berlin (Brenke, 2010). Still, policy initiatives were introduced to promote STEM qualifications, e.g., by the BMBF in 2008 (under the title "Komm, mach MINT"). Overall, I assume that a strong STEM profile increased QPT funding chances not only due to a potential political desire, but also because STEM-focussed institutions might have felt a particular need to improve their studying conditions.

The third research question relates to the abolition of student fees in three German states. The prospect of an upcoming financial shortfall could provide strong incentives to seek alternative resources to ensure consistent teaching standards. To illustrate this reasoning, Holger Fischer, vice president of the University in Hamburg, stated (hyperbolically) that a potential student fees deficit would cause a relapse to the Stone Age (cited by Volkmann-Schluck, 2011). In acknowledgement of such concerns, it should be pointed out that the legislatives changes in all states eventually included compensatory funds.[47] However, it is still conceivable that institutions invested greater efforts towards the QPT contest in response to rather uncertain financial outlooks. I examine this possibility through $Abolition_i$ and its interaction with $ln(Fees_i)$. The latter variable is further included to capture the main effect of student fees.[48]

Lastly, I control for a series of institutional factors that are described in Section 4.2. These comprise variables that indicate location in Western Germany, success at the Excellence Initiative, and institutional types. Concerning the latter aspect, I select universities of applied sciences to represent the base category since they account for almost half of the sample. Moreover, I include controls that refer to institutional size, teaching ratio and the amount of basic funding. Except for $STEM_i$, which is expressed as a percentage, all

[47] See *Gesetz zur Abschaffung und Kompensation der Studiengebühren und zur Änderung anderer Gesetze* (Baden-Württemberg, 21th December 2011), *Gesetz zur Abschaffung der Studiengebühren* (Hamburg, 20th December 2011), and *Gesetz zur Verbesserung von Chancengleichheit beim Hochschulzugang* (North Rhine-Westphalia, 1st March 2011).

[48] The interaction term allows investigating if the student fees level exerts a differential effect in abolition states. I compute the AME of the interaction term according to Ai and Norton (2003).

continuous variables are in logarithmic form on the assumption that they impact the QPT funding chances at a diminishing rate.

4.3.3 Results

The main regression results are reported in Table 4.2. As can be seen, both the LPM and the logit estimation present an overall consistent picture. With respect to the variables of interest, I find strong evidence for a Matthew effect pattern. According to the LPM model, a 10% increase in third-party funds raises the QPT funding probability by 1.31 percentage points, while, under the same scenario, an AME of 1.15 percentage points is obtained from the logit model. The results further lend support to the STEM hypothesis. To be more precise, QPT chances are expected to rise in the range of 2.54 (LPM) to 2.88 (AME) percentage points given a higher STEM students share of 10 percentage points. Both models not only reveal marginal effects of a similar magnitude, but also agree on their statistical significance. Yet, neither the level of student fees nor its partial abolition are, to a statistically significant extent, related to QPT success. Lastly, two notable effects emerge from the included controls. On the one hand, it becomes apparent that a greater student number is associated with higher funding chances, whereas, on the other hand, a probability premium is detected for universities of education.[49]

Despite their statistical significance, it appears that the documented effects, particularly in case of public third-party funds, are of rather moderate economic significance. However, this may turn out to be a hasty conclusion. To shed more light on this matter, I proceed by exploring the Matthew effect channel in more depth by means of the favoured logit model. At first, it may prove helpful to get a visual impression of the relation between QPT success chances and the prior amount of third-party funds. Since the logit model requires assumptions about all remaining covariate values, I picture the relation for a synthetic institution that is supposed to provide a balanced proxy of the higher education sector. More concretely, it is composed of mean values for binary covariates and median values for continuous covariates. For simplicity, it can thus be thought of as a weighted mixture of all institutional types with median size and median financial figures. As shown in Figure 4.1, opting to log-transform third-party funds implies a markedly intuitive interpretation, i.e., acquiring competitive grants resembles the shape of a typical learning curve.

[49] The estimated beta coefficients of the logit model are listed in the LOGIT column of Table 4.2 but will be omitted from the discussion due to their rather complicated interpretation. Technically, each coefficient denotes the marginal change in the log-odds of QPT success for every unit increase in the respective covariate. Albeit the different meaning, significance levels are mostly consistent with the AME estimates.

Variable	LPM	LOGIT	AME
QPT single project funding as binary dependent variable			
ln(Thirdparty)	0.1309 ***	0.7117 ***	0.1147 ***
	(0.0335)	(0.1771)	(0.0264)
STEM	0.2542*	1.7877*	0.2881*
	(0.1273)	(0.8978)	(0.1397)
ln(Fees)	-0.0828	-0.5814	0.0295
	(0.1876)	(1.0616)	(0.1351)
Abolition	-0.1644	-1.3694	-0.1214
	(0.1096)	(0.8016)	(0.0739)
ln(Fees) × Abolition	0.2344	2.0396	0.3478
	(0.2627)	(1.7112)	(0.3317)
West	-0.0453	-0.2388	-0.0380
	(0.0841)	(0.5104)	(0.0798)
University	-0.2113	-0.8420	-0.1312
	(0.1253)	(0.6811)	(0.1003)
Education	0.5382 **	2.9691 *	0.3065 ***
	(0.1817)	(1.2288)	(0.0682)
Music	0.1950	1.2043	0.1628
	(0.1221)	(0.7828)	(0.0848)
Excellence	-0.0698	-0.3751	-0.0613
	(0.0994)	(0.7610)	(0.1259)
ln(Size)	0.1093 **	0.6637 **	0.1070 **
	(0.0359)	(0.2219)	(0.0338)
ln(Teachratio)	-0.0663	-0.6739	-0.1086
	(0.0774)	(0.5805)	(0.0933)
ln(Basic)	0.0753	0.4928	0.0794
	(0.0709)	(0.4772)	(0.0761)
Adj. R^2 / pseudo R^2	0.2300	0.2576	n/a
F test / Wald test	0.0000	0.0000	n/a
No. of observations	250	250	250

Tab. 4.2: REGRESSION RESULTS

Notes: Constants are included but not reported. Pseudo R^2 accords with McFadden (1973). Continuous covariates refer to the year 2009. Robust standard errors are in parentheses.
* $p < 0.05$, ** $p < 0.01$, and *** $p < 0.001$.

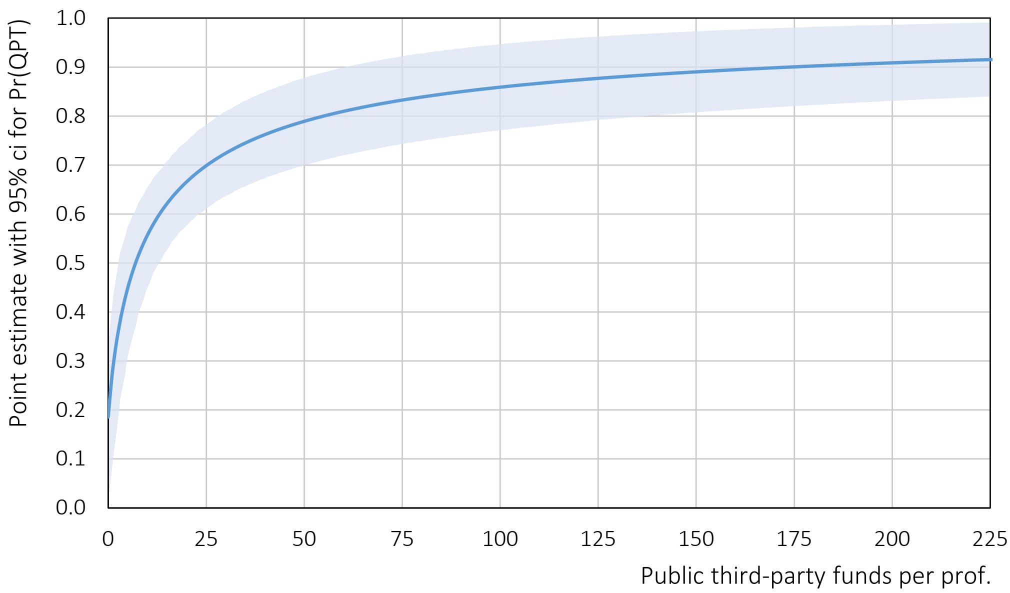


Fig. 4.1: IMPACT OF THIRD-PARTY FUNDS ON QPT FUNDING PROBABILITY

Notes: The estimation is based on a synthetic institution that is composed of mean values for all binary covariates and median values for all continuous covariates. Point estimates are depicted by the solid blue line and 95% confidence intervals are pictured as light blue areas. Public third-party funds per professor are expressed in TEUR.

The learning curve analogy makes it clear that the benefit of additional third-party funds crucially depends on the existing level of funding. If third-party endowments are low, small increases can be expected to be of high impact, whereas an already large financial basis leaves relatively little room for probability improvements. Furthermore, the variability of third-party funds should also be taken into consideration to discuss funding changes of reasonable magnitude. As may be conjectured, both of these aspects differ substantially between institutional types. Universities, for instance, generally rank at the top of the third-party funding list, while colleges of arts and music find themselves mostly at the bottom. To address institutional differences of this kind plus their associated effect heterogeneity, I depict the Matthew effect in segmented form. More specifically, I assess the effect from the viewpoint of median institutions that are described in Table 4.3. Third-party funds aside, it becomes apparent that the median university accommodates the largest student body, whereas the median university of applied sciences is most STEM-oriented. In contrast, the median college of arts and music records the lowest student-to-professor ratio and the highest amount of basic funds per student. Lastly, I find the median university of education to display rather moderate characteristics.

Variable	University	Applied	Education	Music
West Germany, including Berlin	1	1	1	1
Tuition abolition upcoming	0	0	1	0
Excellence Initiative success	0	0	0	0
Percentage of STEM students	26.38	47.42	21.12	0.00
Number of students	14,799	3,980	3,612	592
Students per prof.	71.96	40.98	60.65	16.14
Fees per student	573.63	237.00	534.61	347.35
Basic funds per student	8,667	5,872	5,383	16,616
Public third-party funds per prof.	129,722	8,408	14,309	2,213
No. of observations	80	118	6	46

Tab. 4.3: MEDIAN VALUES BY TYPE OF INSTITUTION

Notes: Columns distinguish between universities, universities of applied sciences, universities of education, and colleges of arts and music. Continuous variables refer to the year 2009.

Figure 4.2 pictures the estimated relation between QPT success and third-party endowment conditional on the (synthetic) median institutions. Each of the four integrated graphics is identically constructed as Figure 4.1 except for two added vertical lines that delimit the interval between the median third-party funding level and an increase of 0.5 standard deviations. The university diagram clearly underlines the diminishing effect nature, as an increase from 129.7 to 178.3 TEUR per prof. would merely result in higher QPT chances of 2.4 percentage points. Yet, as universities of education illustrate, an extensive funding base may not be the only factor that can dampen the effect. In their case, QPT chances are already at an elevated level so that additional third-party funds would be of limited extra value. Expressed in figures, an uptick from 14.3 to 17.6 TEUR per prof. would be associated with a rise in QPT chances of 0.9 percentage points.[50] In comparison, median estimations of both universities of applied sciences and colleges of arts and music reveal effects of a higher magnitude. As for the former, a probability plus of 6.6 percentage points would be predicted following an increase from 8.4 to 13.6 TEUR per prof., while 17.0 percentage points would be noted for the latter given a rise from 2.2 to 7.7 TEUR per prof. Taken together, these two institutional types obviously reflect the Matthew effect in the strongest sense.

[50] As for universities of education, the comparably wide confidence intervals in Figure 4.2 are mainly due to the limited number of observations.

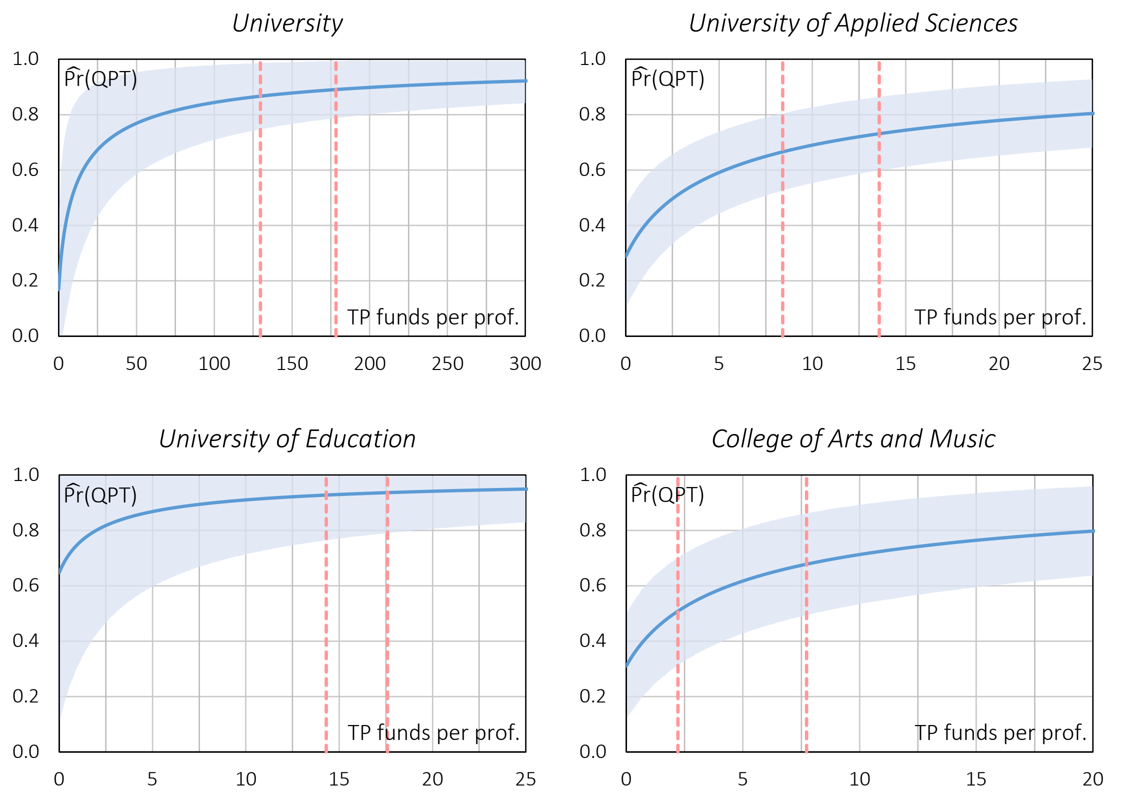


Fig. 4.2: IMPACT OF THIRD-PARTY FUNDS ON QPT FUNDING PROBABILITY BY TYPE OF INSTITUTION

Notes: The estimations are based on synthetic institutions that are composed of median values as reported in Table 4.3. Point estimates are depicted by the solid blue lines and 95% confidence intervals are pictured as light blue areas. Public third-party funds per professor are expressed in TEUR. The median value of this measure is marked by the left vertical line, whereas the value that corresponds to half a standard deviation above the median is marked by the right vertical line. X-axes are capped at about the 95th percentile.

Finally, I present the results of a series of model diagnostics and model alterations to probe the robustness of my findings. An immediate question concerns the proportion of problematic predictions of the LPM. In fact, 6.8% of the predicted probabilities are found to be either negative or greater than one. For this reason, it appears justified to refer to the LPM as more of a supplementary approach. As for the logit model, I conduct the linktest by Pregibon (1980), which aims to detect specification errors. The basic idea behind this test is that if, within a second model, QPT success is regressed on its prediction derived from the original model and the squared prediction, the latter term should not emerge as a relevant explanatory factor. Indeed, the logit model passes the test since the squared prediction is not statistically significant (p-value of 0.19). Moreover, I report the results of three model variants in Appendix C.1. First, I change the link function and employ a probit model. Second, I estimate a heterogeneous logit model that allows the residual variance to differ according to the amount of third-party funds per prof.[51] Third, I utilise data from 2010 within the initial logit model. As becomes apparent, the Matthew effect channel is confirmed by all three models. Similarly, the STEM student share also remains statistically significant although usually at the 10% level.

4.4 Discussion and Conclusion

The perception of academic teaching has seen a shift in recent years. Traditionally, it has mostly been considered a metier for individual engagement among faculty members. Yet, over the last decades, teaching has become more of a public matter that is included into institutional assessments and addressed by national policies (Clegg, 2007; Land & Gordon, 2015). The QPT is a clear manifestation of this trend. Almost 2 billion euros are dedicated towards improving the quality of teaching in German higher education. However, these funds are not evenly spread across the academic landscape but competitively allocated, thus prompting the question about the winners and losers of this funding scheme.

My analysis reveals that the level of third-party funds prior to the QPT is highly indicative of funding success. Simply put, the more grants that were acquired in the past, the higher the QPT chances. Yet my model depicts a diminishing effect of third-party engagement that resembles a typical learning curve. More concretely, it is not to be expected that a gap in third-party volume is related to drastically different QPT chances if institutions are located at the upper end of the funding spectrum. At the lower end, however, the effect becomes notably more pronounced as the estimations for universities of applied sciences and colleges of arts and music show. Starting off median characteristics, an increase of half a standard deviation is associated with a probability plus of 6.6 percentage points for the former and 17.0 percentage points for the latter type of institution. Furthermore, I find higher chances for institutions with a strong STEM-profile or large student numbers. On the contrary, the QPT outcome appears unrelated to student fees, which aligns with the neutrality result of Fischer and Wigger (2016) regarding potential crowding effects.

Before I discuss the implications of these findings, one should be aware of the study's limitations. Despite the diverse set of controls, I cannot rule out that my results are rather a display of multivariate correlations than of causal links. To be specific, reverse causality is not a direct threat to the model given the included time lag. However, it is conceivable that unobserved factors are a source of endogeneity. For instance, some institutions may be (inherently) better at acquiring funds, which could lead to both a larger funding base

[51] Standard non-linear models assume that the error of the latent variable is independent of the covariate vector. In comparison, heterogeneous non-linear models estimate error variances separately (Williams, 2009). These models can lead to more reliable results but are considered inept in situations where the sources of heteroscedasticity are vague (Keele & Park, 2006). Once I allow the error variance to depend on $\ln(Thirdparty)$, the Matthew effect slightly decreases but stays statistically significant.

and increased QPT chances. Interestingly, even if that were the case, my findings would still be consistent with the Matthew logic, which is not based on the premise of causality. A look into the concrete Bible passage illustrates this point. The parable pictures a master who entrusts his property to his servants before setting off on a journey. Yet his property is not arbitrarily split but distributed according to the servants' abilities. Upon his return, the master then discovers that only the better-endowed servants managed to increase his wealth – likely due to their higher financial literacy.

The central takeaway from the Matthew narrative is the path dependency. Should we be worried about this pattern from a welfare perspective? Merton (1968) indeed considers the original Matthew effect a form of misallocation. As for the QPT, this perception does not have to be true if the committee managed to (perfectly) select the projects with the highest social returns. While it lies beyond my scope to resolve this matter, one argument should be given particular thought. As Viner, Powell, and Green (2004) note, the merits of a system that concentrates its resources on its most productive agents are usually based on cumulative advantage. In other words, successful research tends to attract additional funding, which in turn raises the chances of further achievements, thus making the initial recipient a legitimate choice for even more grants. However, the QPT setting casts doubt on the applicability of this rationale for at least two reasons. First, it is unclear on which grounds an organisation with large amounts of third-party funds should be considered a preferred destination for teaching grants. Second, the QPT's program description explicitly states the intention to target the higher education sector at a broad range (BMBF, 2010). In sum, the Matthew effect does not appear to be a purposeful outcome, but rather a reflection of different levels of professionalism in funding acquisition.[52]

I propose a series of policy measures in light of the presented findings. Since the QPT is almost expired, these measures can mainly be seen as recommendations for programs like "Innovation in Higher Education Teaching" (*Innovation in der Hochschullehre*) which is set to succeed the QPT from 2021 onwards (Joint Science Conference, 2019). At first, it is vital to extend the coverage of the official program evaluation. In the QPT case, it is solely restricted to the successful projects. It certainly comes as no surprise that the victorious institutions paint a favourable picture of the competition. However, stating that the QPT's application process received a generally positive assessment from all parties involved (ZQ Uni Mainz & Prognos AG, 2016, p. 24) is not only inaccurate but also misleading. The views of the losing side are equally relevant for a reliable evaluation that is expected to investigate whether the (eventually) rejected institutions may have felt unaccustomed

[52] European higher education has taken a strong managerial turn over the last decades (Krücken, 2011; Teichler, 2003). As for Germany, the ongoing change is clearly visible at the administrative level, where highly qualified personnel is increasingly needed to fill management capacities (Krücken, Blümel, & Kloke, 2013), but also mirrored by the institutionalisation of grant writing practices (Serrano Velarde, 2018).

to the application process.[53] If procedural knowledge were indeed identified as a crucial factor (which would jibe with my Matthew effect interpretation), funding agencies would be advised to offer greater application assistance and contemplate reducing the leeway within their proposal guidelines.

Moreover, data privacy regulations should be revisited. For scientific purposes, it would be extremely valuable to gain (conditional) access to all submitted project proposals and their rating results. Provided with these information, future studies with a scope similar to mine could analyse funding success factors on a more refined level. In addition, it would open up the possibility of exploring if proposal ratings are related to eventual project outcomes. A potential concern could be that proposals that raise disproportionately large expectations obtain systematically higher ratings but also fail to meet their (unrealistic) goals. Briefly put, adverse selection should be considered. As for the QPT, one might counter that the second evaluation round held in 2015, which decided about the project extensions, served as a protection against such risks. Yet, from the initial program depiction (BMBF, 2010), it could already be inferred that cutbacks during the second QPT phase would be highly unlikely.[54] In fact, 87% of all institutions that applied for an extension ended up being successful (BMBF, 2015). Although it is not a trivial task to assess the goal achievement of individual projects, it plays an essential role in the overall program judgement. After all, large-scale funding schemes like the QPT should not just put faith in the applicants' promises given that "someone who is good at 'selling' ideas is not necessarily good at executing them" (Gillett, 1989, p. 28).

---

[53] In view of a potential assessment bias, this part of the evaluation should, of course, be conducted prior to the funding announcement.

[54] This reasoning is based on the BMBF's (2010, p. 5) statement to provide annual budgets of 200 million euros during the extension phase, which even exceeded the annual investment plans for the first phase.

5 Conclusion

One distinctive characteristic of human capital is that it cannot be transferred as easily as financial or physical capital. Knowledge may be codified in textbooks or encyclopaedias which are accessible to a broad public. Yet it still requires years of educational efforts from individuals to partially grasp the current stock of knowledge. The first part of this thesis is devoted to this topic. More specifically, it sheds light on the dissemination of knowledge between scientists. From exploring millions of research collaborations, it can be inferred that star scientists elevate the performance of their co-author networks. Since research networks have grown substantially over time, one could have suspected that individuals translated into diminishing roles. On the contrary, we discover that prolific scientists add to the research frontier not only through their own published work but also by means of spillover effects. However, these effects do not occur over the entire subject spectrum, nor are they traceable to one common origin. Stars enhance their colleagues' publication output most visibly in life sciences and to a lesser extent in physical and health sciences. It also emerges that knowledge is primarily transferred in horizontal direction in life sciences (i.e., between scientists of similar standing), whereas interdisciplinary knowledge flows are characteristic for physical and health sciences. Overall, it can therefore be concluded that the mechanics behind academic progress take different forms in different fields.

This central finding is not restricted to the individual level but can also be derived from an institutional perspective. In the second part of this thesis, emphasis is placed on university production. On the surface, one may receive the impression that public universities form a largely uniform sector given their common dedication towards research and teaching. As far as the efficiency of their activities is concerned, there are, however, great disparities. The important insight is not the mere observation that universities show a varying degree of efficiency (which could have been anticipated), but that efficiency is differently related to a range of structural factors depending on the institution's subject focus. For instance, solely universities with a physical sciences orientation seem to benefit from specialisation (i.e., a narrow subject profile), while a negative link between student fees and efficiency can only be confirmed for universities with a social sciences core. At the same time, it is worth acknowledging that universities are not subject to completely distinct technologies since economies of scale and efficiency gains through third-party funding both represent a general outcome.

Still, it can be conjectured that the academic domain comprises diverse dynamics that call for carefully designed policy measures. It is understandable that governments seek to maximise the returns of their higher education investments. Yet, consensus over the right means for this purpose appears to be missing. Against this backdrop, the third part of this thesis investigates a large German funding program, which for the first time allocated a substantial amount of teaching grants in competitive fashion. In line with the New Public Management paradigm, the Quality Pact for Teaching aims to spur innovation and raise accountability. However, it may have missed its official intention of promoting the quality of academic teaching over a wide range of the academic landscape. More concretely, my analyses reveal that strong commitment to (research-related) third-party funding in years prior to the Quality Pact for Teaching is associated with notably higher success chances. It thus stands to reason that unsuccessful applicants were potentially too inexperienced in competitive funding instruments. In this sense, it can even be worried that the program (unintentionally) contributed to existing inequalities.

In the end, this thesis probably raises as many questions as it answers. Particular interest surrounds the partly diverging effects along the scientific spectrum. What could be at the heart of these heterogeneities? Quantitative research alone may be limited in its ability to explore this question which aims at the inner processes of knowledge creation. Instead, it would be promising to include elements of field research such as Knorr Cetina's (2009) work on epistemic cultures. Future studies could also opt for complimentary indicators to assess academic progress. In light of the knowledge society, altmetrics may offer an opportunity to evaluate the extent to which science contributes to the public discourse. Lastly, it merits further attention if insights from the academic sector can be confirmed in knowledge-intensive branches of the private sector, such as technological companies, law firms, or consultancies.

References

Abbott, M., & Doucouliagos, C. (2003). The efficiency of Australian universities: A data envelopment analysis. *Economics of Education Review*, *22*(1), 89–97.

Adams, J. (2005). Early citation counts correlate with accumulated impact. *Scientometrics*, *63*(3), 567–581.

Adams, S. M., Kochanek, C. S., Gerke, J. R., Stanek, K. Z., & Dai, X. (2017). The search for failed supernovae with the large binocular telescope: Confirmation of a disappearing star. *Monthly Notices of the Royal Astronomical Society*, *468*(4), 4968–4981.

Agasisti, T., & Haelermans, C. (2016). Comparing efficiency of public universities among European countries: Different incentives lead to different performances. *Higher Education Quarterly*, *70*(1), 81–104.

Agasisti, T., & Johnes, G. (2009). Beyond frontiers: Comparing the efficiency of higher education decision-making units across more than one country. *Education Economics*, *17*(1), 59–79.

Agasisti, T., & Pérez-Esparrells, C. (2010). Comparing efficiency in a cross-country perspective: The case of Italian and Spanish state universities. *Higher Education*, *59*(1), 85–103.

Agasisti, T., & Pohl, C. (2012). Comparing German and Italian public universities: Convergence or divergence in the higher education landscape? *Managerial and Decision Economics*, *33*(2), 71–85.

Agasisti, T., & Salerno, C. (2007). Assessing the cost efficiency of Italian universities. *Education Economics*, *15*(4), 455–471.

Agasisti, T., & Wolszczak-Derlacz, J. (2016). Exploring efficiency differentials between Italian and Polish universities, 2001–11. *Science and Public Policy*, *43*(1), 128–142.

Aghion, P., Dewatripont, M., Hoxby, C., Mas-Colell, A., & Sapir, A. (2010). The governance and performance of universities: evidence from Europe and the US. *Economic Policy*, *25*(61), 7–59.

Agrawal, A., McHale, J., & Oettl, A. (2017). How stars matter: Recruiting and peer effects in evolutionary biology. *Research Policy*, *46*(4), 853–867.

Ahmadpoor, M., & Jones, B. F. (2019). Decoding team and individual impact in science and invention. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(28), 13885–13890.

Ahn, T., Charnes, A., & Cooper, W. W. (1988). Some statistical and DEA evaluations of relative efficiencies of public and private institutions of higher learning. *Socio-Economic Planning Sciences*, *22*(6), 259–269.

Ahn, T., & Seiford, L. M. (1993). Sensitivity of DEA to the models and variable sets in a hyphothesis test setting: The efficiency of university operations. *Creative and Innovative Approaches to the Sciences of Management*, *6*, 191–210.

Ahuja, G. (2000). Collaboration networks, structural holes, and innovation: A longitudinal study. *Administrative Science Quarterly*, *45*(3), 425–455.

Ai, C., & Norton, E. C. (2003). Interaction terms in logit and probit models. *Economics Letters*, *80*(1), 123–129.

Aizenman, J., & Kletzer, K. (2011). The life cycle of scholars and papers in economics - the "citation death tax." *Applied Economics*, *43*(27), 4135–4148.

Akcigit, U., Caicedo, S., Miguelez, E., Stantcheva, S., & Sterzi, V. (2018). Dancing with the stars: Innovation through interactions. *National Bureau of Economic Research*.

Allison, P. D., Long, J. S., & Krauze, T. K. (1982). Cumulative advantage and inequality in science. *American Sociological Review*, 615–625.

Athanassopoulos, A. D., & Shale, E. (1997). Assessing the comparative efficiency of higher education institutions in the UK by the means of data envelopment analysis. *Education Economics*, *5*(2), 117–134.

Azoulay, P., Fons-Rosen, C., & Graff Zivin, J. S. (2019). Does science advance one funeral at a time? *American Economic Review*, *109*(8), 2889–2920.

Azoulay, P., Graff Zivin, J. S., & Wang, J. (2010). Superstar extinction. *Quarterly Journal of Economics*, *125*(2), 549–589.

Azoulay, P., Liu, C. C., & Stuart, T. E. (2017). Social influence given (partially) deliberate matching: Career imprints in the creation of academic entrepreneurs. *American Journal of Sociology*, *122*(4), 1223–1271.

Azoulay, P., Stuart, T., & Wang, Y. (2014). Matthew: Effect or fable? *Management Science*, *60*(1), 92–109.

Baruffaldi, S., & Gaessler, F. (2018). Knowledge lost in capital. *Academy of Management Proceedings*, *2018*(1), 18580.

Beasley, J. E. (1990). Comparing university departments. *Omega*, *18*(2), 171–183.

Beaver, D. D. B. (2001). Reflections on scientific collaboration (and its study): Past, present, and future. *Scientometrics*, *52*(3), 365–377.

Becker, B. (1957, October 20). Pioneer of the atom. *New York Times Sunday Magazine*.

Bercovitz, J., & Feldman, M. (2011). The mechanisms of collaboration in inventive teams: Composition, social networks, and geography. *Research Policy*, *40*(1), 81–93.

Bhopal, R., Rankin, J., McColl, E., Thomas, L., Kaner, E., Stacy, R., … Rodgers, H. (1997). The vexed question of authorship: Views of researchers in a British medical faculty. *Bmj*, *314*(7086), 1009.

Bloch, R., Mitterle, A., Rennert, C., & Würmann, C. (2018). Qualitätskapazitäten in der Lehre. Zum Einsatz staatlicher Fördermittel zur Verbesserung der Hochschullehre. *Hochschulmanagement*, *13*(2), 49–55.

Blömeke, S. (2004). Erste Phase an Universitäten und Pädagogischen Hochschulen. In S. Blömeke, P. Reinhold, G. Tulodziecki, & J. Wildt (Eds.), *Handbuch Lehrerbildung* (pp. 262–274).

Bloom, N., Jones, C. I., van Reenen, J., & Webb, M. (2020). Are ideas getting harder to find? *American Economic Review*, *110*(4), 1104–1144.

Bolli, T., Olivares, M., Bonaccorsi, A., Daraio, C., Aracil, A. G., & Lepori, B. (2016). The differential effects of competitive funding on the production frontier and the efficiency of universities. *Economics of Education Review*, *52*, 91–104.

Bonaccorsi, A., & Daraio, C. (2009). Characterizing the European university system: A preliminary classification using census microdata. *Science and Public Policy*, *36*(10), 763–775.

Bonaccorsi, A., Daraio, C., & Simar, L. (2006). Advanced indicators of productivity of universities. An application of robust nonparametric methods to Italian data. *Scientometrics*, *66*(2), 389–410.

Bonaccorsi, A., Secondi, L., Setteducati, E., & Ancaiani, A. (2014). Participation and commitment in third-party research funding: Evidence from Italian universities. *Journal of Technology Transfer*, *39*(2), 169–198.

Borjas, G. J., & Doran, K. B. (2012). The collapse of the Soviet Union and the productivity of American mathematicians. *Quarterly Journal of Economics*, *127*(3), 1143–1203.

Borjas, G. J., & Doran, K. B. (2015). Which peers matter? The relative impacts of collaborators, colleagues, and competitors. *The Review of Economics and Statistics*, *97*(5), 1104–1117.

Bornmann, L., De Moya Anegón, F., & Leydesdorff, L. (2010). Do scientific advancements lean on the shoulders of giants? A bibliometric investigation of the Ortega hypothesis. *PLoS ONE*, *5*(10), e13327.

Boudreau, K. J., Brady, T., Ganguli, I., Gaule, P., Guinan, E., Hollenberg, A., & Lakhani, K. R. (2017). A field experiment on search costs and the formation of scientific collaborations. *The Review of Economics and Statistics*, *99*(4), 565–576.

Brenke, K. (2010). *Fachkräftemangel kurzfristig noch nicht in Sicht*. DIW Berlin.

Breu, T. M., & Raab, R. L. (1994). Efficiency and perceived quality of the nation's "top 25" national universities and national liberal arts colleges: An application of data envelopment analysis to higher education. *Socio-Economic Planning Sciences*, *28*(1), 33–45.

Bruckmeier, K., Fischer, G.-B., & Wigger, B. U. (2017). Status effects of the German excellence initiative. *International Journal of Economics and Finance*, *9*(3), 177–183.

Bruckmeier, K., & Wigger, B. U. (2014). The effects of tuition fees on transition from high school to university in Germany. *Economics of Education Review*, *41*, 14–23.

Caliński, T., & Harabasz, J. (1974). A dendrite method for cluster analysis. *Communications in Statistics-Theory and Methods*, *3*(1), 1–27.

Campbell, D. T. (1960). Blind variation and selective retentions in creative thought as in other knowledge processes. *Psychological Review*, *67*(6), 380–400.

Clegg, S. (2007). The demoic turn: excellence by fiat. In A. Skelton (Ed.), *International perspectives on teaching excellence in higher education: Improving knowledge and practice* (pp. 91–102).

Coelli, T. J. (1996). *Assessing the performance of Australian universities using data envelopment analysis*. University of New England, Department of Econometrics.

Cohn, E., Rhine, S. L. W., & Santos, M. C. (1989). Institutions of higher education as multi-product firms: Economies of scale and scope. *The Review of Economics and Statistics*, *71*(2), 284–290.

Cole, J. R., & Cole, S. (1972). The Ortega hypothesis. *Science*, *178*(4059), 368–375.

Colussi, T. (2018). Social ties in academia: A friend is a treasure. *The Review of Economics and Statistics*, *100*(1), 45–50.

Correia, S., Guimarães, P., & Zylkin, T. (2019). PPMLHDFE: Fast Poisson estimation with high-dimensional fixed effects. *ArXiv Preprint ArXiv:1903.01690*.

Daraio, C., Bonaccorsi, A., & Simar, L. (2015a). Efficiency and economies of scale and specialization in European universities: A directional distance approach. *Journal of Informetrics*, *9*(3), 430–448.

Daraio, C., Bonaccorsi, A., & Simar, L. (2015b). Rankings and university performance: A conditional multidimensional approach. *European Journal of Operational Research*, *244*(3), 918–930.

de Solla Price, D. J., & Beaver, D. (1966). Collaboration in an invisible college. *American Psychologist*, *21*(11), 1011.

De Witte, K., & López-Torres, L. (2017). Efficiency in education: a review of literature and a way forward. *Journal of the Operational Research Society*, *68*(4), 339–363.

Dehne, J., Lucke, U., & Schiefner-Rohs, M. (2017). Digitale Medien und forschungsorientiertes Lehren und Lernen - empirische Einblicke in Projekte und Lehrkonzepte. *Bildungsräume. Proceedings Der 25. Jahrestagung Der Gesellschaft Für Medien in Der Wissenschaft*, 71–83.

Deicke, W., Gess, C., & Rueß, J. (2014). Increasing students' research interests through research-based learning at Humboldt University. *Council on Undergraduate Research Quarterly*, *35*(1), 27–34.

Ding, W. W., Murray, F., & Stuart, T. E. (2006). Gender differences in patenting in the academic life sciences. *Science*, *313*(5787), 665–667.

DiPrete, T. A., & Eirich, G. M. (2006). Cumulative advantage as a mechanism for inequality: A review of theoretical and empirical developments. *Annu. Rev. Sociol.*, *32*, 271–297.

Ductor, L., Fafchamps, M., Goyal, S., & van der Leij, M. J. (2014). Social networks and research output. *The Review of Economics and Statistics*, *96*(5), 936–948.

Dundar, H., & Lewis, D. R. (1995). Departmental productivity in American universities: Economies of scale and scope. *Economics of Education Review*, *14*(2), 119–144.

Dyson, R. G., Allen, R., Camanho, A. S., Podinovski, V. V., Sarrico, C. S., & Shale, E. A. (2001). Pitfalls and protocols in DEA. *European Journal of Operational Research*, *132*(2), 245–259.

Elsevier. (2017). Scopus content coverage guide. *Amesterdam: Elsevier BV*.

Erkkilä, T., & Piironen, O. (2014). Shifting fundaments of European higher education governance: Competition, ranking, autonomy and accountability. *Comparative Education*, *50*(2), 177–191.

Federal Court of Auditors. (2019). *Bericht an den Haushaltsausschuss des Deutschen Bundestages nach § 88 Abs. 2 BHO - Prüfung der Wirksamkeit und zweckentsprechenden Verwendung von Hochschulpaktmitteln des Bundes*. Bonn.

Federal Employment Agency. (2011). *Arbeitsmarktberichterstattung. Kurzinformation Frauen und MINT-Berufe*. Nürnberg.

Federal Ministry of Education and Research. (2010). *Richtlinien zur Umsetzung des gemeinsamen Programms des Bundes und der Länder für bessere Studienbedingungen und mehr Qualität in der Lehre*. Berlin.

Federal Ministry of Education and Research. (2015). *Für bessere Studienbedingungen: Zweite Förderperiode im Qualitätspakt Lehre*. Retrieved September 15, 2016, from https://www.bmbf.de/de/fuer-bessere-studienbedingungen-zweite-foerderperiode-im-qualitaetspakt-lehre-1956.html

Filippini, M., & Lepori, B. (2007). Cost structure, economies of capacity utilization and scope in Swiss higher education institutions. In *Universities and strategic knowledge creation: Specialization and performance in Europe* (pp. 272–304). Edward Elgar Publishing.

Fischer, G.-B., & Wigger, B. U. (2016). Fiscal competition and higher education spending in Germany. *German Economic Review*, *17*(2), 234–252.

Flanagin, A., Carey, L. A., Fontanarosa, P. B., Phillips, S. G., Pace, B. P., Lundberg, G. D., & Rennie, D. (1998). Prevalence of articles with honorary authors and ghost authors in peer-reviewed medical journals. *Jama*, *280*(3), 222–224.

German Rectors' Conference. (2007). *Qualitätsoffensive in der Lehre - Ziele und Maßnahmen: Empfehlung des 105. Senates am 16.10.2007*.

German Research Foundation. (2013). *Excellence initiative at a glance*.

Geuna, A. (1999). *The economics of knowledge production: funding and the structure of university research*. Edward Elgar Publishing.

Geuna, A., & Nesta, L. J. J. (2006). University patenting and its effects on academic research: The emerging European evidence. *Research Policy*, *35*(6), 790–807.

Gillett, R. (1989). Research performance indicators based on peer review: a critical analysis. *Higher Education Quarterly*, *43*(1), 20–38.

Gourieroux, C., Monfort, A., & Trognon, A. (1984). Pseudo maximum likelihood methods: Theory. *Econometrica: Journal of the Econometric Society*, *52*(3), 681–700.

Goyal, S., Van Der Leij, M. J., & Moraga-González, J. L. (2006). Economics: An emerging small world. *Journal of Political Economy*, *114*(2), 403–412.

Grimpe, C. (2012). Extramural research grants and scientists' funding strategies: Beggars cannot be choosers? *Research Policy*, *41*(8), 1448–1460.

Gulbrandsen, M., & Smeby, J. C. (2005). Industry funding and university professors' research performance. *Research Policy*, *34*(6), 932–950.

Hadamard, J. (1945). The psychology of invention in the mathematical field. *The Mathematical Gazette*.

Hall, K. L., Vogel, A. L., Huang, G. C., Serrano, K. J., Rice, E. L., Tsakraklides, S. P., & Fiore, S. M. (2018). The science of team science: A review of the empirical evidence and research gaps on collaboration in science. *American Psychologist*, *73*(4), 532–548.

Hanke, M., & Leopoldseder, T. (1998). Comparing the efficiency of Austrian universities. A data envelopment analysis application. *Tertiary Education and Management*, *4*(3), 191–197.

Hartmann, M. (2006). Die Exzellenzinitiative - ein Paradigmenwechsel in der deutschen Hochschulpolitik. *Leviathan*, *34*(4), 447–465.

Hartog, J., Ferrer-i-Carbonell, A., & Jonker, N. (2002). Linking measured risk aversion to individual characteristics. *Kyklos*, *55*(1), 3–26.

Herberholz, L., & Wigger, B. U. (2020). Efficiency of european universities: A comparison of peers. *CESifo Working Paper No. 8044*.

Hetze, P. (2011). *Nachhaltige Hochschulstrategien für mehr MINT-Absolventen*. Stifterverband für die Wissenschaft.

Heublein, U. (2014). Student drop-out from German higher education institutions. *European Journal of Education*, *49*(4), 497–513.

Hornbostel, S. (2001). Third party funding of German universities. An indicator of research activity? *Scientometrics*, *50*(3), 523–537.

Horrace, W. C., & Oaxaca, R. L. (2006). Results on the bias and inconsistency of ordinary least squares for the linear probability model. *Economics Letters*, *90*(3), 321–327.

Hübner, M. (2012). Do tuition fees affect enrollment behavior? Evidence from a 'natural experiment'in Germany. *Economics of Education Review*, *31*(6), 949–960.

Hüther, O., & Krücken, G. (2018). *Higher education in Germany - Recent developments in an international perspective*. Springer.

Iacus, S. M., King, G., & Porro, G. (2011). Multivariate matching methods that are monotonic imbalance bounding. *Journal of the American Statistical Association*, *106*(493), 345–361.

Iacus, S. M., King, G., & Porro, G. (2012). Causal inference without balance checking: Coarsened exact matching. *Political Analysis*, *20*, 1–24.

Jäger, S., & Heining, J. (2019). How substitutable are workers? Evidence from worker deaths. *Unpublished Manuscript*.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning* (Vol. 112). Springer.

Jaravel, X., Petkova, N., & Bell, A. (2018). Team-specific capital and innovation. *American Economic Review*, *108*(4–5), 1034–1073.

Johnes, J. (1990). Unit costs: Some explanations of the differences between UK universities. *Applied Economics*, *22*(7), 853–862.

Johnes, J. (2006). Data envelopment analysis and its application to the measurement of efficiency in higher education. *Economics of Education Review*, *25*(3), 273–288.

Johnes, J., & Johnes, G. (1995). Research funding and performance in U.K. university departments of economics: A frontier analysis. *Economics of Education Review*, *14*(3), 301–314.

Johnes, J., & Yu, L. (2008). Measuring the research performance of Chinese higher education institutions using data envelopment analysis. *China Economic Review*, *19*(4), 679–696.

Joint Science Conference. (2011a). *Rückenwind für die Lehre: 111 Hochschulen werden im Qualitätspakt Lehre gefördert*. Retrieved September 15, 2016, from http://www.gwk-bonn.de/fileadmin/Pressemitteilungen/pm2011-13.pdf

Joint Science Conference. (2011b). *Zweite Runde Qualitätspakt Lehre: 102 Hochschulen erhalten insgesamt rund 400 Mio. Euro*. Retrieved September 15, 2016, from http://www.gwk-bonn.de/fileadmin/Pressemitteilungen/pm2011-24.pdf

Joint Science Conference. (2019). *"Innovation in der Hochschullehre" - Toepfer Stiftung gGmbH als Trägerinstitution ausgewählt*. Retrieved June 20, 2020, from https://www.gwk-bonn.de/fileadmin/Redaktion/Dokumente/Pressemitteilungen/pm2019-17.pdf

Jones, B. F. (2009). The burden of knowledge and the "death of the renaissance man": Is innovation getting harder? *Review of Economic Studies*, *76*(1), 283–317.

Jones, B. F. (2010). Age and great invention. *The Review of Economics and Statistics*, *92*(1), 1–14.

Jones, B. F., & Olken, B. A. (2005). Do leaders matter? National leadership and growth since World War II. *Quarterly Journal of Economics*, *120*(3), 835–864.

Jones, B. F., Wuchty, S., & Uzzi, B. (2008). Multi-university research teams: Shifting impact, geography, and stratification in science. *Science*, *322*(5905), 1259–1262.

Joumady, O., & Ris, C. (2005). Performance in European higher education: A non-parametric production frontier approach. *Education Economics*, *13*(2), 189–205.

Kahn, S., & MacGarvie, M. J. (2016). How important is U.S. location for research in science? *The Review of Economics and Statistics*, *98*(2), 397–414.

Katharaki, M., & Katharakis, G. (2010). A comparative assessment of Greek universities' efficiency using quantitative analysis. *International Journal of Educational Research*, *49*(4–5), 115–128.

Katz, J. S., & Martin, B. R. (1997). What is research collaboration? *Research Policy*, *26*(1), 1–18.

Katz, Y., & Matter, U. (2019). Metrics of inequality: The concentration of resources in the U.S. biomedical elite. *Science as Culture*, 1–28.

Keele, L., & Park, D. K. (2006). Difficult choices: an evaluation of heterogeneous choice models. *Paper for the 2004 Meeting of the American Political Science Association*.

Kempkes, G., & Pohl, C. (2010). The efficiency of German universities - some evidence from nonparametric and parametric methods. *Applied Economics*, *42*(16), 2063–2079.

Knorr Cetina, K. (2009). *Epistemic cultures: How the sciences make knowledge*. Harvard University Press.

Koch, F. D., & Vogt, J. (2015). Psychology in an interdisciplinary setting: A large-scale project to improve university teaching. *Psychology Learning & Teaching*, *14*(2), 158–168.

König, M. D., Lorenz, J., & Zilibotti, F. (2016). Innovation vs. imitation and the evolution of productivity distributions. *Theoretical Economics*, *11*(3), 1053–1102.

Koppel, O., & Plünnecke, A. (2009). *Fachkräftemangel in Deutschland: Bildungsökonomische Analyse, politische Handlungsempfehlungen, Wachstums- und Fiskaleffekte*. IW Köln.

Krücken, G. (2011). A European perspective on new modes of university governance and actorhood. *Research & Occasional Paper Series: CSHE.17.11*. University of California, Berkeley.

Krücken, G., Blümel, A., & Kloke, K. (2013). The managerial turn in higher education? On the interplay of organizational and occupational change in German academia. *Minerva*, *51*(4), 417–442.

Laband, D. N., & Tollison, R. D. (2000). Intellectual collaboration. *Journal of Political Economy*, *108*(3), 632–662.

Land, R., & Gordon, G. (2015). *Teaching excellence initiatives: modalities and operational factors*. York: The Higher Education Academy.

Larivière, V., & Gingras, Y. (2010). The impact factor's Matthew Effect: A natural experiment in bibliometrics. *Journal of the American Society for Information Science and Technology*, *61*(2), 424–427.

Larivière, V., Haustein, S., & Börner, K. (2015). Long-distance interdisciplinarity leads to higher scientific impact. *PLoS ONE*, *10*(3), e0122565.

Laudel, G. (2006). The 'quality myth': Promoting and hindering conditions for acquiring research funds. *Higher Education*, *52*(3), 375–403.

Leahey, E., Beckman, C. M., & Stanko, T. L. (2017). Prominent but less productive: The impact of interdisciplinarity on scientists' research. *Administrative Science Quarterly*, *62*(1), 105–139.

Leitner, K. H., Prikoszovits, J., Schaffhauser-Linzatti, M., Stowasser, R., & Wagner, K. (2007). The impact of size and specialisation on universities' department performance: A DEA analysis applied to Austrian universities. *Higher Education*, *53*(4), 517–538.

Lepori, B., Baschung, L., & Probst, C. (2010). Patterns of subject mix in higher education institutions: A first empirical analysis using the aquameth database. *Minerva*, *48*(1), 73–99.

Levin, S. G., & Stephan, P. E. (1991). Research productivity over the life cycle: Evidence for academic scientists. *American Economic Review*, *81*(1), 114–132.

Long, J. S. (1997). Regression models for categorical and limited dependent variables. In *Advanced Quantitative Techniques in the Social Sciences* (Vol. 7, p. 219). Sage Publications.

Lotka, A. J. (1926). The frequency distribution of scientific productivity. *Journal of the Washington Academy of Sciences*, *16*(12), 317–323.

Lucas, R. E., & Moll, B. (2014). Knowledge growth and the allocation of time. *Journal of Political Economy*, *122*(1), 1–51.

Makles, A. (2012). Stata tip 110: How to get the optimal k-means cluster solution. *Stata Journal*, *12*(2), 347–351.

Marquardt, W. (2011). *Neuere Entwicklungen in der Hochschulfinanzierung*. Berlin.

Martin, B. R., & Irvine, J. (1983). Assessing basic research. Some partial indicators of scientific progress in radio astronomy. *Research Policy*, *12*(2), 61–90.

McFadden, D. L. (1973). Conditional logit analysis of qualitative choice behavior. In *Frontiers in econometrics* (pp. 105–142).

McMillan, M. L., & Datta, D. (1998). The relative efficiencies of Canadian universities: a DEA perspective. *Canadian Public Policy/Analyse de Politiques*, *24*(2), 485–511.

Melin, G. (2000). Pragmatism and self-organization: Research collaboration on the individual level. *Research Policy*, *29*(1), 31–40.

Menter, M., Lehmann, E. E., & Klarl, T. (2018). In search of excellence: a case study of the first excellence initiative of Germany. *Journal of Business Economics*, *88*(9), 1105–1132.

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

Milligan, G. W., & Cooper, M. C. (1985). An examination of procedures for determining the number of clusters in a data set. *Psychometrika*, *50*(2), 159–179.

Moed, H. F., Burger, W. J. M., Frankfort, J. G., & Van Raan, A. F. J. (1985). The use of bibliometric data for the measurement of university research performance. *Research Policy*, *14*(3), 131–149.

Mohnen, M. (2018). Stars and brokers: Knowledge spillovers among medical scientists. *Unpublished Manuscript*.

Mongeon, P., & Paul-Hus, A. (2016). The journal coverage of Web of Science and Scopus: a comparative analysis. *Scientometrics*, *106*(1), 213–228.

Muscio, A., Quaglione, D., & Vallanti, G. (2013). Does government funding complement or substitute private research funding to universities? *Research Policy*, *42*(1), 63–75.

National Research Council. (2015). *Enhancing the effectiveness of team science* (N. J. Cooke & M. L. Hilton, Eds.). Washington, D.C.: National Academies Press.

Necmi K. Avkiran. (2001). Investigating technical and scale efficiencies of Australian universities through data envelopment analysis. *Socio-Economic Planning Sciences*, *35*(1), 57–80.

Nederhof, A. J. (2006). Bibliometric monitoring of research performance in the social sciences and the humanities: A review. *Scientometrics*, *66*(1), 81–100.

Nguyen, B. D., & Nielsen, K. M. (2010). The value of independent directors: Evidence from sudden deaths. *Journal of Financial Economics*, *98*(3), 550–567.

OECD. (2003). *Turning science into business: Patenting and licensing at public research organisations*. OECD Publishing.

OECD. (2019). *Benchmarking higher education system performance*. Paris: OECD Publishing.

Oettl, A. (2012). Reconceptualizing stars: Scientist helpfulness and peer performance. *Management Science*, *58*(6), 1122–1140.

Oreopoulos, P. (2007). Do dropouts drop out too soon? Wealth, health and happiness from compulsory schooling. *Journal of Public Economics*, *91*(11–12), 2213–2229.

Osterloh, M., & Frey, B. S. (2008). *Anreize im Wissenschaftssystem*. IOU.

Partha, D., & David, P. A. (1994). Toward a new economics of science. *Research Policy*, *23*(5), 487–521.

Petersen, A. M., Jung, W.-S., Yang, J.-S., & Stanley, H. E. (2011). Quantitative and empirical demonstration of the Matthew effect in a study of career longevity. *Proceedings of the National Academy of Sciences*, *108*(1), 18–23.

Piro, F. N., Aksnes, D. W., & Rørstad, K. (2013). A macro analysis of productivity differences across fields: Challenges in the measurement of scientific publishing. *Journal of the American Society for Information Science and Technology*, *64*(2), 307–320.

Pregibon, D. (1980). Goodness of link tests for generalized linear models. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *29*(1), 15–24.

Radicchi, F., Fortunato, S., & Castellano, C. (2008). Universality of citation distributions: Toward an objective measure of scientific impact. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(45), 17268–17272.

Romer, P. M. (1990). Endogenous technological change. *Journal of Political Economy*, *98*(5), S71–S102.

Rosenkopf, L., & Almeida, P. (2003). Overcoming local search through alliances and mobility. *Management Science*, *49*(6), 751–766.

Rothaermel, F. T., & Hess, A. M. (2007). Building dynamic capabilities: Innovation driven by individual-, firm-, and network-level effects. *Organization Science*, *18*(6), 898–921.

Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, *20*(C), 53–65.

Sagarra, M., Mar-Molinero, C., & Agasisti, T. (2017). Exploring the efficiency of Mexican universities: Integrating data envelopment analysis and multidimensional scaling. *Omega (United Kingdom)*, *67*, 123–133.

Santamaría, L., & Mihaljević, H. (2018). Comparison and benchmark of name-to-gender inference services. *PeerJ Computer Science*, *2018*(7), 1–29.

Sarrico, C. S., Hogan, S. M., Dyson, R. G., & Athanassopoulos, A. D. (1997). Data envelopment analysis and university selection. *Journal of the Operational Research Society*, *48*(12), 1163–1177.

Schimank, U., & Lange, S. (2009). Germany: A latecomer to new public management. In *University governance* (pp. 51–75). Springer.

Schreiber, M. (2008). A modification of the h-index: The hm-index accounts for multi-authored manuscripts. *Journal of Informetrics*, *2*(3), 211–216.

Schumpeter, J. A. (1934). The theory of economic development: An inquiry into profits, credit, interest, and the business cycle. *Social Science Electronic Publishing*.

Serrano Velarde, K. (2018). The way we ask for money… The emergence and institutionalization of grant writing practices in academia. *Minerva*, *56*(1), 85–107.

Shen, H. (2013). Mind the gender gap. *Nature*, *495*(7439), 22.

Shin, J. C. (2009). Classifying higher education institutions in Korea: A performance-based approach. *Higher Education*, *57*(2), 247–266.

Shin, J. C., & Cummings, W. K. (2010). Multilevel analysis of academic publishing across disciplines: Research preference, collaboration, and time on research. *Scientometrics*, *85*(2), 581–594.

Sickles, R. C., & Zelenyuk, V. (2019). *Measurement of productivity and efficiency*. Cambridge University Press.

Silverman, B. W. (1986). Density estimation for statistics and data analysis. In *Monographs on Statistics and Applied Probability, London: Chapman and Hall, 1986*.

Simar, L., & Wilson, P. W. (1998). Sensitivity analysis of efficiency scores: How to bootstrap in nonparametric frontier models. *Management Science*, *44*(1), 49–61.

Simar, L., & Wilson, P. W. (2000). A general methodology for bootstrapping in non-parametric frontier models. *Journal of Applied Statistics*, *27*(6), 779–802.

Simar, L., & Wilson, P. W. (2007). Estimation and inference in two-stage, semi-parametric models of production processes. *Journal of Econometrics*, *136*(1), 31–64.

Singh, J., & Fleming, L. (2010). Lone inventors as sources of breakthroughs: Myth or reality? *Management Science*, *56*(1), 41–56.

Smith, J. P., & Naylor, R. A. (2001a). Determinants of degree performance in UK universities: A statistical analyses. *Oxford Bulletin of Economics and Statistics*, *63*(1), 29–60.

Smith, J. P., & Naylor, R. A. (2001b). Dropping out of university: A statistical analysis of the probability of withdrawal for UK university students. *Journal of the Royal Statistical Society. Series A: Statistics in Society*, *164*(2), 389–405.

Stanley, G., & Reynolds, P. (1994). Similarity grouping of Australian universities. *Higher Education*, *27*(3), 359–366.

Stephan, P. E. (2012). *How economics shapes science* (Vol. 1). Harvard University Press Cambridge, MA.

Stifterverband. (2009). *Die Preisträger im Wettbewerb Exzellente Lehre*. Retrieved September 15, 2016, from https://idw-online.de/de/news?print=1&id=339493

Stigler, G. J. (1988). *Memoirs of an unregulated economist*. Basic Books.

Taylor, B., & Harris, G. (2004). Relative efficiency among South African universities: A data envelopment analysis. *Higher Education*, *47*(1), 73–89.

Teichler, U. (2003). The future of higher education and the future of higher education research. *Tertiary Education & Management*, *9*(3), 171–185.

Teixeira, P. N. (2016). Two continents divided by the same trends? Reflections about marketization, competition, and inequality in European higher education. *Research in the Sociology of Organizations*, *46*, 489–508.

Teixeira, P. N., Rocha, V., Biscaia, R., & Cardoso, M. F. (2012). Competition and diversity in higher education: An empirical approach to specialization patterns of Portuguese institutions. *Higher Education*, *63*(3), 337–352.

Thanassoulis, E., Kortelainen, M., Johnes, G., & Johnes, J. (2011). Costs and efficiency of higher education institutions in England: A DEA analysis. *Journal of the Operational Research Society*, *62*(7), 1282–1297.

Tierney, M. L. (1980). An estimate of departmental cost functions. *Higher Education*, *9*(4), 453–468.

Tomkins, C., & Green, R. (1988). An experiment in the use of data envelopment analysis for evaluating the efficiency of UK university departments of accounting. *Financial Accountability & Management*, *4*(2), 147–164.

Usher, A. P. (1954). *A history of mechanical inventions*. Courier Corporation.

Uzzi, B., Mukherjee, S., Stringer, M., & Jones, B. F. (2013). Atypical combinations and scientific impact. *Science*, *342*(6157), 468–472.

Valadkhani, A., & Worthington, A. (2006). Ranking and clustering Australian university research performance, 1998–2002. *Journal of Higher Education Policy and Management*, *28*(2), 189–210.

Viner, N., Powell, P., & Green, R. (2004). Institutionalized biases in the award of research grants: a preliminary analysis revisiting the principle of accumulative advantage. *Research Policy*, *33*(3), 443–454.

Volkmann-Schluck, P. (2011, January 11). SPD will Wegfall der Studiengebühren kompensieren. *Hamburger Abendblatt*.

Waldinger, F. (2010). Quality matters: The expulsion of professors and the consequences for PhD student outcomes in Nazi Germany. *Journal of Political Economy*, *118*(4), 787–831.

Waldinger, F. (2012). Peer effects in science: Evidence from the dismissal of scientists in Nazi Germany. *Review of Economic Studies*, *79*(2), 838–861.

Waldinger, F. (2016). Bombs, brains, and science: The role of human and physical capital for the creation of scientific knowledge. *The Review of Economics and Statistics*, *98*(5), 811–831.

Walker, I., & Zhu, Y. (2013). The impact of university degrees on the lifecycle of earnings: Some further analysis. *BIS Research Paper Number 112*.

Waltman, L. (2016). A review of the literature on citation impact indicators. *Journal of Informetrics*, *10*(2), 365–391.

Waltman, L., Tijssen, R. J. W., & Eck, N. J. van. (2011). Globalisation of science in kilometres. *Journal of Informetrics*, *5*(4), 574–582.

Waltman, L., van Eck, N. J., van Leeuwen, T. N., Visser, M. S., & van Raan, A. F. J. (2011). Towards a new crown indicator: An empirical analysis. *Scientometrics*, *87*(3), 467–481.

Wang, J. (2013). Citation time window choice for research impact evaluation. *Scientometrics*, *94*(3), 851–872.

Wang, J., Thijs, B., & Glänzel, W. (2015). Interdisciplinarity and impact: Distinct effects of variety, balance, and disparity. *PLoS ONE*, *10*(5), e0127298.

Weitzman, M. L. (1998). Recombinant growth. *Quarterly Journal of Economics*, *113*(2), 331–360.

Williams, D. R. (2003). The health of men: Structured inequalities and opportunities. *American Journal of Public Health*, *93*(5), 724–731.

Williams, R. (2009). Using heterogeneous choice models to compare logit and probit coefficients across groups. *Sociological Methods and Research*.

Wislar, J. S., Flanagin, A., Fontanarosa, P. B., & DeAngelis, C. D. (2011). Honorary and ghost authorship in high impact biomedical journals: A cross sectional survey. *Bmj*, *343*, d6128.

Wolszczak-Derlacz, J. (2017). An evaluation and explanation of (in)efficiency in higher education institutions in Europe and the U.S. with the application of two-stage semi-parametric DEA. *Research Policy*, *46*(9), 1595–1605.

Wolszczak-Derlacz, J., & Parteka, A. (2011). Efficiency of European public higher education institutions: A two-stage multicountry approach. *Scientometrics*, *89*(3), 887–917.

Wooldridge, J. M. (1997). Quasi-likelihood methods for count data. In *Handbook of applied econometrics* (Vol. 2, pp. 352–406). Blackwell.

Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data*. MIT press.

Worthington, A. C. (2002). An empirical survey of frontier efficiency measurement techniques in education. *Education Economics*, *9*(3), 245–268.

Wuchty, S., Jones, B. F., & Uzzi, B. (2007). The increasing dominance of teams in production of knowledge. *Science*, *316*(5827), 1036–1039.

Zimmerman, S. D., & Altonji, J. G. (2018). The costs of and net returns to college major. *National Bureau of Economic Research*.

Zoghbi, A. C., Rocha, F., & Mattos, E. (2013). Education production efficiency: Evidence from Brazilian universities. *Economic Modelling*, *31*(1), 94–103.

ZQ Uni Mainz, & Prognos AG. (2016). *Evaluation des Bund-Länder-Programms für bessere Studienbedingungen und mehr Qualität in der Lehre (Qualitätspakt Lehre): Abschlussbericht über die 1. Förderphase 2011-2016*. Berlin/Mainz.

ZQ Uni Mainz, & Prognos AG. (2018). *Evaluation des Bund-Länder-Programms für bessere Studienbedingungen und mehr Qualität in der Lehre (Qualitätspakt Lehre): Ergebnisbericht über den Evaluationszeitraum 2013-2018*. Berlin/Mainz.

Zucker, L. G., & Darby, M. R. (1996). Star scientists and institutional transformation: Patterns of invention and innovation in the formation of the biotechnology industry. *Proceedings of the National Academy of Sciences of the United States of America*, *93*(23), 12709–12716.

Zuckerman, H. (1967). Nobel laureates in science: Patterns of productivity, collaboration, and authorship. *American Sociological Review*, *32*(3), 391–403.

Appendix A

A.1 Classification of Scientific Fields

National Academy of Sciences	Scopus	*Field*
Animal, Nutritional, and Applied Microbial Sciences	Agricultural and Biological Sciences	
Biochemistry	Biochemistry, Genetics, and Molecular Biology	
Biophysics and Computational Biology	Immunology and Microbiology	
Cellular and Developmental Biology	Neuroscience	
Cellular and Molecular Neuroscience	Pharmacology, Toxicology, and Pharmaceutics	
Evolutionary Biology		
Genetics		
Physiology and Pharmacology		
Plant Biology		
Plant, Soil, and Microbial Sciences		
Systems Neuroscience		
		Life Sciences
Immunology and Inflammation	Dentistry	
Medical Genetics, Hematology, and Oncology	Health Professions	
Medical Physiology and Metabolism	Medicine	
Microbial Biology	Nursing	
	Veterinary	
		Health Sciences
Applied Mathematical Sciences	Chemical Engineering	
Applied Physical Sciences	Chemistry	
Astronomy	Computer Science	
Chemistry	Earth and Planetary Sciences	
Computer and Information Sciences	Energy	
Engineering Sciences	Engineering	
Environmental Sciences and Ecology	Environmental Science	
Geology	Materials Science	
Geophysics	Mathematics	
Mathematics	Physics and Astronomy	
Physics		
		Physical Sciences
Anthropology	Arts and Humanities	
Economic Sciences	Business, Management, and Accounting	
Human Environmental Sciences	Decision Sciences	
Psychological and Cognitive Sciences	Economics, Econometrics, and Finance	
Social and Political Sciences	Psychology	
	Social Sciences	
		Social Sciences

Tab. A.1: CLASSIFICATION OF SCIENTIFIC FIELDS

Notes: We divide the scientific spectrum into four main fields based on the *All Science Journal Classification* by Scopus (omitting the field of multidisciplinary studies). Sections in use by the National Academy of Sciences are mapped into this taxonomy according to the reported scheme so that each of its members can be assigned to either life, health, physical, or social sciences. Sections and subfields are listed in alphabetical order.

A.2 List of Treatment Stars

Personal Data	Excerpt from Obituary	Field
KURT JUNGERMANN, 1938 – 2002 University of Göttingen	Died unexpectedly⁺	Life
ROBERT M. MACNAB, 1940 – 2003* Yale University	Fell at home	Life
ROBERT J. KADNER, 1942 – 2005 University of Virginia	Died unexpectedly	Life
DAVID S. SEGAL, 1942 – 2005 University of California, San Diego	Very short and aggressive course of pancreatic cancer	Life
JERRY O. WOLFF, 1942 – 2008* St. Cloud State University	Suicide	Life
DON C. WILEY, 1944 – 2001 Harvard University	Accident	Life
DAVID L. GARBERS, 1944 – 2006 University of Texas Southwestern	Heart attack	Life
UWE CLAUSSEN, 1945 – 2008 University of Jena	Heart attack	Life
REINHART HEINRICH, 1946 – 2006 Humboldt University of Berlin	Died unexpectedly	Life
STEVEN C. HEBERT, 1946 – 2008 Yale University	Sudden death after cardiovascular disease	Life
FRED F. KADLUBAR, 1946 – 2010 University of Arkansas for Medical Sciences	Died unexpectedly	Life
DOMINIQUE DORMONT, 1948 – 2003 CEA Fontenay-aux-Roses	Severe influenza	Life
MARJORIE A. ASMUSSEN, 1949 – 2004 University of Georgia	Bicycle accident	Life
JOHN C. LAWRENCE, 1949 – 2006 University of Virginia	Heart attack	Life
ROBERT W. GOLDBACH, 1949 – 2009* Wageningen University & Research	Trampled to death by an elephant while bird watching	Life
BARBARA K. BURGESS, 1950 – 2001 University of California, Irvine	Suicide	Life
EBBE S. NIELSEN, 1950 – 2001 Australian National Insect Collection	Heart attack	Life
DALE J. BENOS, 1950 – 2010 University of Alabama at Birmingham	Died suddenly while on a walk with his wife	Life
FRANÇOIS TILLEQUIN, 1950 – 2011 Paris Descartes University	Died unexpectedly	Life
THOMAS V. DUNWIDDIE, 1951 – 2001 University of Colorado Medical Campus	Accident while rock climbing	Life
ROBERT B. DICKSON, 1952 – 2006 Georgetown University	Ruptured aorta	Life
VINCENT R. FRANCESCHI, 1953 – 2005 Washington State University	Died unexpectedly	Life
DONALD W. THOMAS, 1953 – 2009 University of Sherbrooke	Stroke	Life
JEFFERY W. WALKER, 1954 – 2010* University of Arizona	Died suddenly and unexpectedly	Life
BAHMAN EGHBALL, 1956 – 2004* University of Nebraska-Lincoln	Swimming accident	Life

(continued)

Personal Data	Excerpt from Obituary	Field
BRIAN M. J. FOXWELL, 1956 – 2008 Imperial College London	Died unexpectedly	Life
RAWIE I. HOLLINGSWORTH, 1956 – 2012 Michigan State University	Collapsed in a hallway due to pulmonary emboli	Life
ANDREAS J. HELBIG, 1957 – 2005 University of Greifswald	Late diagnosed cancer, short illness	Life
ANGEL A. ALONSO, 1957 – 2005 McGill University	Infection with viral encephalitis	Life
KENJI TAKABAYASHI, 1957 – 2006 University of California, San Diego	Died unexpectedly⁺	Life
JASON D. MORROW, 1957 – 2008* Vanderbilt University	Died suddenly	Life
LLOYD R. KELLAND, 1958 – 2008* The Institute of Cancer Research, London	Died suddenly and unexpectedly	Life
ALAN P. WOLFFE, 1959 – 2001 National Institutes of Health, NICHD	Road accident	Life
STEFAN ROSEWICZ, 1960 – 2004 Charité – Berlin University of Medicine	Died suddenly and unexpectedly	Life
MICHAEL BRÜSS, 1961 – 2006 University of Bonn	Died suddenly and unexpectedly	Life
ALAA E. EL-HUSSEINI, 1962 – 2007 University of British Columbia	Drowned while on vacation	Life
MARCO F. RAMONI, 1963 – 2010 Boston Children's Hospital	Heart failure	Life
ANDREA TONTINI, 1966 – 2012 University of Urbino	Suicide	Life
CHARLES A. LOCKWOOD, 1970 – 2008 University College London	Motorcycle accident	Life
EKARAT JANTRATID, 1975 – 2010 Goethe University Frankfurt	Died unexpectedly⁺	Life
LAWRENCE D. JACOBS, 1938 – 2001 University at Buffalo	Brief battle with cancer	Health
SIGRID POSER, 1941 – 2004 University of Göttingen	Died unexpectedly	Health
RICHARD H. WARD, 1943 – 2003 University of Oxford	Died suddenly of cardiac causes	Health
OLOF JOHNELL, 1944 – 2006* Malmö University	Died suddenly and unexpectedly	Health
LARS JANZON, 1944 – 2007* Lund University	Short illness	Health
HAIM RING, 1944 – 2008 Tel Aviv University	Short and serious illness	Health
MICHAEL J. REED, 1944 – 2009 St Mary's Hospital London	Died suddenly	Health
SEPPO S. SANTAVIRTA, 1945 – 2005 Helsinki University	Heart attack	Health
WILLIAM C. KOLLER, 1945 – 2005* University of North Carolina at Chapel Hill	Sudden cardiac problems	Health
WAYNE A. HENING, 1945 – 2008 Rutgers University	Brief struggle with pulmonary fibrosis	Health
AXEL PERNECZKY, 1945 – 2009 University of Mainz	Died suddenly and unexpectedly	Health

(continued)

Personal Data	Excerpt from Obituary	Field
MASSIRAO CHIARIELLO, 1945 – 2010* University of Naples Federico II	Short struggle with cancer	Health
MARIO STEFANELLI, 1945 – 2010 University of Pavia	Haemorrhagic stroke	Health
ROBERT M. ADRIAN, 1946 – 2007 Georgetown University	Died suddenly	Health
JECKONIAH O. NDINYA-ACHOLA, 1946 – 2010 University of Nairobi	Sudden kidney failure	Health
JEFFERY M. ISNER, 1947 – 2001 Tufts University	Heart attack	Health
DAVID B. LARSON, 1947 – 2002* Duke University	Heart attack	Health
JOHN L. BEARD, 1947 – 2009 Pennsylvania State University	Died suddenly	Health
JOB J. BWAYO, 1948 – 2007 University of Nairobi	Murdered by carjackers	Health
WERNER A. BAUTZ, 1949 – 2008 University of Erlangen-Nuremberg	Heart attack	Health
GARY J. MILLER, 1950 – 2001 University of Colorado Medical Campus	Died suddenly while jogging	Health
DANIEL P. SCHUSTER, 1950 – 2007 Washington University in St. Louis	Died suddenly while playing racquetball	Health
GREG R. ALEXANDER, 1950 – 2007 University of South Florida	Heart failure	Health
ELIZABETH S. WILLIAMS, 1951 – 2004 University of Wyoming	Traffic accident	Health
HELMUT DREXLER, 1951 – 2009 Hannover Medical School	Accident during race biking	Health
GERD HAUSDORF, 1952 – 2001 University of Göttingen	Died unexpectedly	Health
HANS J. SCHWANITZ, 1952 – 2004 Osnabrück University	Died unexpectedly	Health
RICHARD L. WALKER, 1952 – 2008 University of California, Davis	Probable suicide	Health
HELMUT MAXEINER, 1952 – 2009 Charité – Berlin University of Medicine	Bicycle accident	Health
RICHARD W. SCHWARTZ, 1952 – 2010* University of Kentucky	Very brief battle with lung cancer	Health
BARRY M. KACINSKI, 1953 – 2003 Yale University	Heart attack	Health
TONY S. KELLER, 1955 – 2006* University of Vermont	Died of gunshots, apparent homicide	Health
FRANS W. J. ALBERS, 1955 – 2007 University of Groningen	Brief illness	Health
ALAN J. FLISHER, 1956 – 2010* University of Cape Town	Brief struggle with leukaemia	Health
ROBERT B. DUNCAN, 1957 – 2007 Virginia-Maryland College of Veterinary Medicine	Died suddenly	Health
JASON D. MORROW, 1957 – 2008* Vanderbilt University	Died suddenly	Health
JEFFREY W. TYLER, 1957 – 2009 University of Missouri-Columbia	Died unexpectedly	Health

(continued)

Personal Data	Excerpt from Obituary	Field
JAE-YOUNG RHO, 1958 – 2002 University of Memphis	Heart attack	Health
WALTER J. MUIR, 1958 – 2009 University of Edinburgh	Died suddenly and unexpectedly	Health
BERNIE J. O'BRIEN, 1959 – 2004 University of Edinburgh	Died tragically while jogging	Health
HERMAN T. YEE, 1959 – 2011 New York University	Died suddenly	Health
KEVIN P. GRANATA, 1961 – 2007 Virginia Polytechnic Institute and State University	Victim of university campus shooting	Health
JEFFREY W. BERGER, 1963 – 2001 University of Pennsylvania	Stomach cancer, died two weeks after diagnosis	Health
SERGIO VIDAL, 1966 – 2003 University of Santiago de Compostela	Sudden illness	Health
JAN KWIECINSKI, 1938 – 2003 Polish Academy of Sciences	Died suddenly during a cycling trip	Physical
JAMES R. HOLTON, 1938 – 2004 University of Washington	Stroke and heart attack during a mid-day run	Physical
LORENZ KRAMER, 1941 – 2005 University of Bayreuth	Died unexpectedly	Physical
DAVID J. FAULKNER, 1942 – 2002 University of California, San Diego	Complications after heart surgery	Physical
JIN AU KONG, 1942 – 2008 Massachusetts Institute of Technology	Complications from pneumonia	Physical
PAUL GRANGE, 1943 – 2003 University of Louvain	Heart attack	Physical
JÜRGEN O. BESENHARD, 1944 – 2006 Graz University of Technology	Stroke while returning from conference	Physical
ANDREI YAKOVLEV, 1944 – 2008 University of Rochester	Heart attack	Physical
REX E. SHEPHERD, 1945 – 2003 University of Pittsburgh	Heart attack	Physical
TADEUSZ PAKULA, 1945 – 2005 Max Planck Institute for Polymer Research	Short and severe illness	Physical
ROBERT F. DENNO, 1945 – 2008 University of Maryland	Heart attack	Physical
STEPHEN H. SCHNEIDER, 1945 – 2010 Stanford University	Heart attack	Physical
ROBERT A. SCHOMMER, 1946 – 2001 Cerro Tololo Inter American Observatory	Suicide	Physical
RICHARD E. EWING, 1946 – 2007 Texas A&M University	Heart attack	Physical
MICHAEL J. WEAVER, 1947 – 2002 Purdue University	Died unexpectedly	Physical
HANS J. RATH, 1947 – 2012 University of Bremen	Short and severe illness	Physical
YORAM J. KAUFMAN, 1948 – 2006 NASA Goddard Space Flight Center	Bicycle accident	Physical
PAUL G. SILVER, 1948 – 2009 Carnegie Institution of Washington	Car accident	Physical
CHARLES E. HOYLE, 1948 – 2009 University of Southern Mississippi	Died unexpectedly	Physical

(continued)

Personal Data	Excerpt from Obituary	Field
JOHN P. HUCHRA, 1948 – 2010 Harvard University	Heart attack	Physical
PHILIPPE FLAJOLET, 1948 – 2011 INRIA at Rocquencourt	Died suddenly and unexpectedly	Physical
IOANNIS VARDOULAKIS, 1949 – 2009 National Technical University of Athens	Gardening accident	Physical
ULRICH M. GÖSELE, 1949 – 2009 Max Planck Institute of Microstructure Physics	Found dead in his apartment	Physical
ISAAC GOLDHIRSCH, 1949 – 2010 Tel Aviv University	Died unexpectedly	Physical
GERHARD H. JIRKA, 1944 – 2010 Karlsruhe Institute of Technology	Heart attack	Physical
HASSAN AREF, 1950 – 2011 Virginia Polytechnic Institute and State University	Aortic dissection	Physical
SHENG YU, 1950 – 2012 Western University	Unexpectedly	Physical
JAAP G. SNIJDERS, 1951 – 2003 University of Groningen	Died unexpectedly due to short-term illness	Physical
JEAN-PIERRE MAELFAIT, 1951 – 2003 University of Ghent	Died suddenly and unexpectedly	Physical
PAUL F. BARBARA, 1953 – 2010 University of Texas at Austin	Complications following cardiac arrest	Physical
IAN P. ROTHWELL, 1955 – 2004 Purdue University	Car accident	Physical
STRATIS V. SOTIRCHOS, 1956 – 2004 University of Rochester	Car accident	Physical
RICHARD C. PLAYLE, 1956 – 2005 Wilfrid Laurier University	Heart failure after brief illness	Physical
STEPHEN P. HOPKIN, 1956 – 2006 University of Reading	Car accident	Physical
ZLATKO B. TEŠANOVIĆ, 1956 – 2012 Johns Hopkins University	Heart attack	Physical
IAN I. KOGAN, 1958 – 2003 University of Oxford	Heart attack	Physical
ADOLFO PARMALIANA, 1958 – 2008 University of Messina	Suicide	Physical
PETER G. DUYNKERKE, 1959 – 2002 Utrecht University	Tragic accident	Physical
LEOPOLDO P. FRANCA, 1959 – 2012 University of Colorado Denver	Heart attack	Physical
IAN H. LANGFORD, 1961 – 2002* University of East Anglia	Suicide or home accident	Physical
WILLIAM D. ARMSTRONG, 1961 – 2006 University of Wyoming	Plane crash	Physical
TIL AACH, 1961 – 2012 RWTH Aachen University	Died unexpectedly	Physical
WERNER S. WEIGLHOFER, 1962 – 2003 University of Glasgow	Struck by an avalanche	Physical
ALEXANDER E. FARRELL, 1962 –2008 University of California, Berkeley	Died unexpectedly	Physical
RAJEEV MOTWANI, 1962 – 2009 Stanford University	Accidental drowning	Physical

(continued)

Personal Data	Excerpt from Obituary	Field
MANUEL FORESTINI, 1963 – 2003 University of Grenoble	Heart attack	Physical
ROBERT HEITZ, 1964 – 2003 Technical University of Berlin	Cardiac aneurysm	Physical
EDOARDO CAPELLO, 1965 – 2009* Polytechnic University of Milan	Heart attack while skiing	Physical
FEMKE OLYSLAGER, 1966 – 2009 University of Ghent	Died unexpectedly	Physical
LUIS SERRANO-ANDRÉS, 1966 – 2010* University of Valencia	Died unexpectedly	Physical
JOAKIM H. PETERSSON, 1968 – 2002 Linköping University	Died suddenly and unexpectedly	Physical
KEITH FAGNOU, 1971 – 2009 University of Ottawa	Complications from influenza	Physical
SAM T. ROWEIS, 1972 – 2010 New York University	Suicide	Physical
KEVIN E. STRECKER, 1974 – 2012 Rice University	Heart attack	Physical
FRANS M. DIELEMAN, 1942 – 2005 Utrecht University	Died suddenly and unexpectedly	Social
DENNIS A. RONDINELLI, 1943 – 2007 Duke University	Died unexpectedly[+]	Social
ROB KLING, 1944 – 2003 Indiana University	Unexpectedly due to cardiovascular disease	Social
VICTOR FLORIAN, 1945 – 2002 Bar-Ilan University	Died unexpectedly[+]	Social
DICK R. WITTINK, 1945 – 2005 Yale University	Diabetic seizure while swimming in his pool	Social
MICHAEL W. PFAU, 1945 – 2009 University of Oklahoma	Brief illness	Social
KENNETH A. KAVALE, 1946 – 2008 Regent University	Died unexpectedly	Social
PETER GOLDIE, 1946 – 2011 University of Manchester	Brief illness	Social
PHILLIP L. WALKER, 1947 – 2009 University of California, Santa Barbara	Died unexpectedly	Social
IVAN MERVIELDE, 1947 – 2011 University of Ghent	Short illness	Social
PETER W. JUSCZYK, 1948 – 2001 Johns Hopkins University	Heart attack	Social
SUMANTRA GHOSHAL, 1948 – 2004 London Business School	Brain haemorrhage	Social
LYNDA L. KAID, 1948 – 2011 University of Florida	Died unexpectedly	Social
M. THEA SINCLAIR, 1950 – 2006 University of Nottingham	Riding accident	Social
MARK S. JOHNSON, 1950 – 2007* Montclair State University	Died suddenly	Social
GEORGE M. ZINKHAN, 1952 – 2009 University of Georgia	Suicide after being prime suspect in a triple homicide	Social
PETER LIPTON, 1954 – 2007 University of Cambridge	Collapsed after a squash game	Social

(continued)

Personal Data	Excerpt from Obituary	Field
BRIAN D. MULLEN, 1955 – 2006 Syracuse University	Died unexpectedly	Social
STEVEN C. POE, 1960 – 2007 University of North Texas	Heart attack	Social
JEAN O. LANJOUW, 1962 – 2005 University of California, Berkeley	Renal cancer, died three months after first symptoms	Social
JÖRG SCHUMACHER, 1962 – 2010 University of Leipzig	Died unexpectedly	Social
RODNEY CLARK, 1967 – 2006 Wayne State University	Died unexpectedly	Social
ALASDAIR CROCKETT, 1968 – 2006 University of Essex	Suicide	Social
STEPHEN O. GYIMAH, 1968 – 2012* Queen's University	Unexpectedly due to brief illness	Social

Tab. A.2: LIST OF TREATMENT STARS

Notes: The list comprises 162 outstanding scientists whose active careers ended abruptly between 2001 and 2012 due to unexpected death at a maximum age of 65 years. Asterisks indicate that the year of birth could not be ascertained and was instead estimated based on the year of death and the reported death age. Plus signs indicate that the death cause was verified after personal consultation with former colleagues. Affiliations are selected as of the last held job position.

A.3 Detailed Matching Procedure

Technically, the matching procedure is performed at the star-collaborator dyad level. This implies that although the majority of covariates refers to the collaborator, characteristics of the star scientist and features of their collaboration are equally important to account for. The overall goal is to construct a control group that mirrors the treatment group in these three dimensions and thus determines the hypothetical outcome path for the latter group had they not experienced the unexpected star death. We proceed by detailing our matching algorithm in three main steps.

Step 1: Identifying treatment and potential control dyads

Our bibliometric data cover substantial network information. We apply several constraints to identify treatment and control dyads therein. First, we only consider established star collaborations. Thus, as of the year of death, stars must have fulfilled a star criterion (as defined in Section 2.2.2) and collaborations must have emerged through jointly published articles.[55] Second, we require collaborators to be research-active at the time of death. We implement this constraint by confining the matching sample to dyads where collaborators are below 40 career years and have not ended their publication activities prior to the year of death. Third, we focus on established collaborators. This leads us to exclude collaborators with less than five career years and collaborators who simultaneously began their careers and star collaborations. Fourth and lastly, we remove collaborators that died, regardless of cause, as documented by our treatment case search.

These general constraints are common to both treatment and control dyads. In order to draw the distinction between the two groups, we lean on the star scientists. Treatment stars died unexpectedly at a maximum age of 65 years. We impose two additional constraints to infer that they were engaged in research activities at the time of their death. First, their obituaries do not indicate that they entered any kind of retirement phase. Second, they published at least one article over the two years preceding the year of death. Control stars, in contrast, must not die. Deceased stars, as disclosed through our treatment search, are therefore not eligible to be part of control dyads. From the remaining pool of potential control dyads, we first remove stars with career ages of over 35 years, which resembles the age threshold applied to treatment stars on the assumption that scientific careers start at the age of 30.[56] Second, we restrict the control pool to stars who continued publishing for a

[55] This essentially excludes a small number of treatment dyads that are solely verifiable through delayed publications, i.e., after the year of death. In these cases, it remains unclear if collaboration actually took place or if the co-author possibly served as a replacement for the deceased star.

[56] Jones (2010) points out that the age at which eminent scientists and inventors become research-active has notably increased over the past. He documents a mean age of 31 years for the end of the 20th century and further shows that age patterns are very similar across scientific fields.

minimum of five years after the considered death year. We expect these stars to be alive besides being involved in further research activities. Third and analogous to the treatment case, control stars are required to have published one article over the past two years.

Together, these constraints allow us to determine treatment dyads and to narrow down potential control dyads. As for the former group, we make a final adjustment by excluding collaborators that experience more than one treatment event. These cases are relatively rare,[57] but still problematic from a methodological standpoint since it would be hardly possible to isolate the individual treatment effects.

Step 2: Matching field-by-field and year-by-year

This step aims to identify counterparts for each treatment dyad. We begin by splitting the confined sample into four distinct groups according to the stars' field classification (life, health, physical, and social sciences). Within each field, we iterate over the years from 2001 to 2012. Every treatment year is hereby associated with a disjoint subgroup of the full set of treatment dyads. Potential control dyads, in contrast, can be linked to more than one year. To be clear, a dyad pictures a collaboration over time. Control dyads thus serve as feasible matches in any given year, in which they meet the criteria stated above. We proceed by matching treatment and control dyads field-by-field and year-by-year. Note that treatment dyads can initially match with multiple control dyads as we postpone the implementation of the one-to-one and without replacement features to Step 3.

Dyads form a match if they belong to the exact same stratum derived from partitioning the support of the joint distribution of the following covariates, with optional percentile cut-offs in parentheses:

- DYAD LEVEL: no. of joint articles (50th; 85th), years since last joint article (25th; 75th)
- STAR LEVEL: no. of received citations (25th; 75th), field classification[58]
- COLLABORATOR LEVEL: no. of distinct co-authors (25th; 75th), adjusted forward citations in each of the last five years and aggregated over a max. 5-year period prior to that (50th; 80th; 98th), career age in 5-year intervals

The selection of cut-offs is strongly guided by distributional features and therefore hard to determine a priori. For instance, we discover that casual dyads that collaborate once or twice are very common,[59] which allows us to choose a relatively high first cut-off for the number of joint articles, i.e., the 50th percentile. This decision is motivated on theoretical

[57] They account for 3% of the treated collaborator sample. Among them, an unfortunate group of eight collaborators is exposed to the maximum of three death events.

[58] The stars' field constitutes an implicit covariate since we opt to match dyads field-by-field.

[59] This finding is indeed not new. For instance, Azoulay et al. (2010) depict a very similar distribution of co-authorship intensity for US life scientists.

grounds, as we attempt to separate casual dyads from (more) regular ones, but is also reasonable on pragmatic grounds, as broad stratifications usually improve the chances of successful matches. Moreover, the assessment of collaboration intensity likely changes depending on the time horizon. To illustrate this point, in 2001, one might label a dyad frequent if more than five collaborations occurred from 1996 onwards. Yet this absolute threshold is probably not suitable to classify dyads as of 2010 given that our data (still) span until 1996. Relative cut-off points are hence preferred and eventually determined in an iterative process that intends to maximise both the degree of covariate balance and the overall matching rate.

Step 3: Selecting final control dyads

In line with the related literature (Azoulay et al., 2010; Jaravel et al., 2018), we employ a one-to-one matching without replacement. More specifically, control collaborators can only be matched once irrespective of multiple occurrences within different control dyads. In slight deviation to previous studies that have pursued a purely chronological approach, we select control dyads by first considering all allocation possibilities. This offers two advantages. First, it allows executing each matching year simultaneously and can thus lead to substantial time savings if parallel computing resources are available. Second, it gives us the opportunity to inspect if some collaborators, who are allocated more than once, constitute mandatory matches for certain treatment dyads (in possibly later years).

We begin by locking mandatory matches, i.e., we assign control dyads that are without alternatives. In case that control collaborators are part of multiple mandatory dyads, we prioritise earlier treatment years and, if necessary, break ties at random. After removing all other occurrences of these collaborators, we proceed chronologically. In other words, we restrict all remaining control collaborators to their first match, again breaking ties at random. At this stage, matching is realised without replacement. In order to implement the one-to-one feature, we lastly select control dyads randomly in the event that treatment dyads are presented with multiple options. In sum, we manage to find a definite match for 93.6% of all treatment dyads, hereby employing 8,406 distinct strata.

A.4 Additional Summary Statistics

Variable	Life Sciences		Health Sciences		Physical Sciences		Social Sciences	
	Treated	Control	Treated	Control	Treated	Control	Treated	Control
Career age	18.52	18.44	18.57	18.54	17.42	17.33	16.94	16.96
Female prediction	0.268	0.291	0.273	0.283	0.140	0.163	0.293	0.279
U.S. affiliated	0.459	0.424	0.422	0.414	0.375	0381	0.630	0.468
Star status	0.243	0.253	0.265	0.251	0.269	0.253	0.175	0.104
No. of distinct co-authors	60.09	60.09	77.95	74.71	54.95	58.10	21.54	18.55
No. of articles	13.26	12.53	15.26	14.56	15.25	15.72	7.71	7.33
No. of citations	524.7	522.5	531.9	519.4	505.6	475.7	166.5	145.3
No. of collaborations	2.03	1.93	2.49	2.40	2.31	2.32	1.63	1.68
Years since last collaboration	4.40	4.30	3.77	3.67	4.07	4.01	4.20	4.27
No. of citations (star)	1,621.6	1,660.3	1,643.2	1,393.4	1,712.6	1,601.2	209.0	213.3
No. of collaborators	5,598		7,950		4,738		308	

Tab. A.3: SUMMARY STATISTICS ON MATCHED COLLABORATORS BY SCIENTIFIC FIELD

Notes: The table reports a breakdown of mean values by scientific field and treatment status. All time-varying variables refer to the year preceding the (inherited) year of star death. Article, citation, and distinct co-author numbers are aggregated over a prior 5-year span. Gender information are inferred through name and country data and are available for 85.3% of the sample.

A.5 Robustness Checks

	Main Model	Variant 1	Variant 2	Variant 3	Variant 4	Variant 5	Variant 6
Overall Sample)	*Article count as dependent variable*						
After death × treated	-0.043*	-0.045*	-0.043*	-0.043*	-0.043**	-0.073**	n/a
	(0.022)	(0.021)	(0.022)	(0.022)	(0.013)	(0.028)	n/a
Log pseudo-likelihood	-189,139	-189,135	-189,227	-189,138	-189,780	-67,906	n/a
No. of observations	275,344	275,344	275,344	275,344	275,344	90,600	n/a
No. of dyads	18,542	18,542	18,542	18,542	18,542	6,040	n/a
Overall Sample)	*Forward citations as dependent variable*						
After death × treated	-0.081**	-0.087**	-0.081**	-0.081*	-0.080**	-0.087**	-0.081**
	(0.028)	(0.029)	(0.028)	(0.028)	(0.019)	(0.034)	(0.030)
Log pseudo-likelihood	-2,800,261	-2,799,841	-2,804,363	-2,799,577	-2,819,493	-964,785	-2,873,838
No. of observations	275,166	275,166	275,166	275,166	275,166	90,585	275,166
No. of dyads	18,527	18,527	18,527	18,527	18,527	6,039	18,527
Life Sciences)	*Article count as dependent variable*						
After death × treated	-0.066*	-0.065*	-0.065*	-0.066*	-0.063**	-0.079*	n/a
	(0.030)	(0.030)	(0.030)	(0.030)	(0.024)	(0.036)	n/a
Log pseudo-likelihood	-52,754	-52,755	-52,776	-52,752	-52,948	-23,952	n/a
No. of observations	83,541	83,541	83,541	83,541	83,541	34,770	n/a
No. of dyads	5,585	5,585	5,585	5,585	5,585	2,318	n/a
Life Sciences)	*Forward citations as dependent variable*						
After death × treated	-0.114**	-0.119**	-0.113**	-0.115**	-0.114**	-0.107*	-0.122*
	(0.041)	(0.044)	(0.041)	(0.041)	(0.034)	(0.046)	(0.053)
Log pseudo-likelihood	-809,005	-808,823	-810,045	-808,285	-813,795	-337,475	-835,832
No. of observations	83,526	83,526	83,526	83,526	83,526	34,770	83,526
No. of dyads	5,584	5,584	5,584	5,584	5,584	2,318	5,584

(continued)

	Main Model	Variant 1	Variant 2	Variant 3	Variant 4	Variant 5	Variant 6
Health Sciences)	*Article count as dependent variable*						
After death × treated	-0.044	-0.042	-0.044	-0.044	-0.047*	-0.065	n/a
	(0.034)	(0.034)	(0.034)	(0.034)	(0.021)	(0.046)	n/a
Log pseudo-likelihood	-79,330	-79,327	-79,369	-79,327	-79,603	-27,978	n/a
No. of observations	118,212	118,212	118,212	118,212	118,212	37,125	n/a
No. of dyads	7,940	7,940	7,940	7,940	7,940	2,475	n/a
Health Sciences)	*Forward citations as dependent variable*						
After death × treated	-0.043	-0.039	-0.044	-0.043	-0.047	-0.050	-0.030
	(0.038)	(0.040)	(0.038)	(0.038)	(0.027)	(0.043)	(0.038)
Log pseudo-likelihood	-1,154,988	-1,154,805	-1,156,912	-1,154,597	-1,164,565	-390,776	-1,176,750
No. of observations	118,148	118,148	118,148	118,148	118,148	37,110	118,148
No. of dyads	7,934	7,934	7,934	7,934	7,934	2,474	7,934
Physical Sciences)	*Article count as dependent variable*						
After death × treated	-0.026	-0.033	-0.026	-0.026	-0.027	-0.089*	n/a
	(0.030)	(0.029)	(0.030)	(0.030)	(0.025)	(0.043)	n/a
Log pseudo-likelihood	-53,894	-53,892	-53,936	-53,890	-54,089	-15,207	n/a
No. of observations	69,107	69,107	69,107	69,107	69,107	17,580	n/a
No. of dyads	4,711	4,711	4,711	4,711	4,711	1,172	n/a
Physical Sciences)	*Forward citations as dependent variable*						
After death × treated	-0.104*	-0.121**	-0.106*	-0.104*	-0.102**	-0.121*	-0.115**
	(0.043)	(0.040)	(0.042)	(0.043)	(0.038)	(0.060)	(0.044)
Log pseudo-likelihood	-784,720	-784,690	-787,455	-784,029	-791,293	-218,597	-808,717
No. of observations	69,008	69,008	69,008	69,008	69,008	17,580	69,008
No. of dyads	4,703	4,703	4,703	4,703	4,703	1,172	4,703

(continued)

	Main Model	Variant 1	Variant 2	Variant 3	Variant 4	Variant 5	Variant 6
Social Sciences)	*Article count as dependent variable*						
After death × treated	-0.062	-0.133	-0.061	-0.063	-0.051	0.043	n/a
	(0.151)	(0.157)	(0.153)	(0.151)	(0.146)	(0.211)	n/a
Log pseudo-likelihood	-3,021	-3,020	-3,034	-3,017	-3,042	-706	n/a
No. of observations	4,484	4,484	4,484	4,484	4,484	1,125	n/a
No. of dyads	306	306	306	306	306	75	n/a
Social Sciences)	*Forward citations as dependent variable*						
After death × treated	-0.015	-0.137	-0.013	-0.012	-0.017	0.206	-0.015
	(0.205)	(0.222)	(0.205)	(0.206)	(0.187)	(0.356)	(0.205)
Log pseudo-likelihood	-40,122	-40,192	-40,612	-39,909	-40,837	-12,103	-40,127
No. of observations	4,484	4,484	4,484	4,484	4,484	1,125	4,484
No. of dyads	306	306	306	306	306	75	306

Tab. A.4: ROBUSTNESS CHECKS

Notes: The table reports the results of a series of robustness checks that probe our main model stated in Equation (2.2). In Variant 1, we prolong the pre-treatment period to include the death event, thus delaying the start of the post-treatment period to the first full calendar year after the stars' passing. In Variant 2, we switch to an alternative career age specification and employ 5-year brackets to capture life cycle effects. In Variant 3, we include interacted calendar year and career age fixed effects instead of inserting them separately. In Variant 4, we shift the level of standard error clustering from the star to the collaborator level. In Variant 5, we estimate effects based on a balanced panel of collaborators that are traceable for exactly seven years before and after their respective death year. In Variant 6, we refrain from winsorizing forward citations. Any further estimation features and variable definitions are maintained from the main model. Robust standard errors are in parentheses.

* $p < 0.05$, ** $p < 0.01$, and *** $p < 0.001$.

A.6 Supplementary Estimations

This section provides further estimations that are useful for probing the robustness of the findings presented in Section 2.4.2. Methodologically, we rely on Equation (2.3) or on a slightly modified version thereof, which contains either multiple three-way interactions or continuous interaction terms.

Basic Interaction Measures. Frequency, timing, and length of a collaborative relationship provide intuitive starting points for exploring treatment effect heterogeneity. However, frequency was not found to be a relevant factor. We draw similar conclusions with regard to the other two interaction measures, as reported in Tables A.5 and A.6. As for timing, we distinguish recent collaborations, who published a joint article either in the year of death or in the year before, from older collaborations. The former group comprises slightly over 30% of both the treated and control sample. In the absence of any statistically significant interaction terms (see Table A.5), we find no reliable link between recency and treatment effect levels, thus indicating that the disruption of ongoing research projects plays a negligible role. As for collaboration length, we separate collaborations that published joint articles over multiple years (30% in both samples) from one-year collaboration. As can be seen from Table A.6, treatment effect differences between these groups are statistically insignificant.

Horizontal Spillovers & Intra-US Effects. We address the concern that the effects estimated for star-star dyads and US-US dyads might be entangled by combining both interactions (together with their common terms) in the same specification. As shown in Table A.7, results hardly change under this scenario. More concretely, both effect channels stay statistically significant in life sciences despite slight reductions in absolute point estimates. In case of health sciences, we see a minor increase in both effect sizes and precision, which causes the differential quality effect for US-US dyads to become statistically significant at the five-percent level (former p-value was 0.064).

Intra-US Effects & Other Physical Proximity Effects. Our measures of physical proximity at the city-level (co-located) and at the country-level (US-US) partly overlap. We thus test the potential dependence of these channels by estimating their differential treatment effects within one specification. In view of Table A.8, we find the concern to be unfounded given that the field-specific results see only marginal changes. In addition to that, we explore the possibility that intra-US effects could just be a reflection of intra-country effects. However, as can be inferred from Table A.9, this is unlikely to be true. While intra-country effects take up some part of the productivity effect in health sciences, they point in the opposite direction of the quality effect in life sciences, thereby enhancing it.

Subject Space & Topic Space. Distance in subject space represents a key predictor for the treatment effect sizes in health and physical sciences. This finding suggests that the visible decline in research output supposedly results from the loss of complementary scientific resources that were provided by the deceased star. However, it remains relatively vague, which concrete form or combination of resources plays a decisive role. In order to improve our understanding in this regard, we distinguish between distances in subject and topic space. Methodologically, the calculation of topic space distance follows the exact same steps as outlined for subject distance with the exception of utilising keywords instead of journal categories. Moreover, keywords are cleaned and Porter-stemmed to mitigate the risk of misclassification, which could arise from Scopus using a non-standardised keyword pool. Overall, this leads us to differentiate between 162,791 keywords. However, it should be noted that keywords are not available for every publication, which causes a roughly 10% reduction in sample size (topic distance could not be determined for these dyads in absence of keywords). Combining subject and topic space metrics in one estimation gives rise to the results in Table A.10. As can be seen, focus on divergent research matters is not a central driver of the treatment effects, as none of the topic space interactions become statistically significant. On the contrary, the relevance of the subject space channel in health and physical sciences remains largely unaffected (especially in the quality dimension). In unreported estimations, we additionally tested if proximity in subject and topic space, or a combination of both, might explain treatment effect outcomes, but could not determine any reliable link.

Continuous Interactions. For the ease of interpretation, we solely used dummy variables to investigate effect heterogeneities in Section 2.4.2. While most variables naturally allow for a binary classification (e.g., co-location, star status, or intra-US collaborations), we applied a cut-off between the second and third tertile in some instances, most notably regarding the subject distance measure. To allay the concern that this specific cut-off may be pivotal for our results, we present additional results derived from continuous interaction effects. Technically, we allow for a non-linear relationship between the treatment effect and the distance measure by inserting two interaction terms, one regular (*After death* $\times$ *treated* $\times$ *subject distance*) and one squared (*After death* $\times$ *treated* $\times$ *squared subject distance*), in addition to the standard treatment term (*After death* $\times$ *treated).* Models that include multiple continuous interactions become hardly interpretable from estimated coefficients alone. We thus present a graphical illustration in Figure A.1 that depicts how the (overall) treatment impact varies along the subject distance range. As can be seen, higher distances are reliably linked to a higher treatment magnitude in case of the overall sample, health sciences, and the quality sphere of physical sciences, which is in line with our field-specific findings.

	Overall Sample	Life Sciences	Health Sciences	Physical Sciences	Social Sciences
	Article count as dependent variable				
After death × treated	-0.051 *	-0.086 **	-0.052	-0.014	-0.142
	(0.021)	(0.033)	(0.034)	(0.038)	(0.174)
After death × treated × recent dyad	0.018	0.065	0.026	-0.041	0.281
	(0.028)	(0.048)	(0.038)	(0.048)	(0.320)
Log pseudo-likelihood	-189,024	-52,745	-79,239	-53,873	-3,020
No. of observations	275,344	83,541	118,212	69,107	4,484
No. of dyads	18,542	5,585	7,940	4,711	306
	Forward citations as dependent variable				
After death × treated	-0.091 **	-0.092 *	-0.056	-0.134 *	-0.158
	(0.028)	(0.041)	(0.038)	(0.052)	(0.257)
After death × treated × recent dyad	0.018	-0.071	0.025	0.067	0.461
	(0.044)	(0.075)	(0.061)	(0.084)	(0.396)
Log pseudo-likelihood	-2,791,626	-808,089	-1,149,979	-782,563	-39,999
No. of observations	275,166	83,526	118,148	69,008	4,484
No. of dyads	18,527	5,584	7,934	4,703	306

Tab. A.5: EFFECT HETEROGENEITY BY COLLABORATION RECENCY

Notes: The overall sample consists of 9,297 pairs of treated and control collaborators confined to a nine-year window around the (inherited) year of star death. The panel is equally unbalanced for treated and control collaborators. Both output measures are co-author adjusted. Forward citations are winsorized at the 99.9th percentile. Collaborators without variation in output over the observation period are dropped by the estimation routine. Field delineation is based on the stars' publication profile or, if available, derived from the classification of the NAS. Robust standard errors are in parentheses, clustered at the level of the star.

* $p < 0.05$, ** $p < 0.01$, and *** $p < 0.001$.

	Overall Sample	Life Sciences	Health Sciences	Physical Sciences	Social Sciences
	Article count as dependent variable				
After death × treated	-0.039	-0.057	-0.037	-0.020	-0.150
	(0.023)	(0.031)	(0.036)	(0.037)	(0.180)
After death × treated × multi-year dyad	-0.012	-0.031	-0.019	-0.014	0.261
	(0.028)	(0.053)	(0.041)	(0.050)	(0.281)
Log pseudo-likelihood	-189,223	-52,752	-79,329	-53,892	-3,020
No. of observations	275,344	83,541	118,212	69,107	4,484
No. of dyads	18,542	5,585	7,940	4,711	306
	Forward citations as dependent variable				
After death × treated	-0.076 **	-0.084	-0.069	-0.074	-0.154
	(0.029)	(0.044)	(0.040)	(0.052)	(0.246)
After death × treated × multi-year dyad	-0.015	-0.105	0.069	-0.077	0.422
	(0.045)	(0.073)	(0.058)	(0.092)	(0.405)
Log pseudo-likelihood	-2,800,185	-808,858	-1,154,909	-784,615	-40,064
No. of observations	275,166	83,526	118,148	69,008	4,484
No. of dyads	18,527	5,584	7,934	4,703	306

Tab. A.6: EFFECT HETEROGENEITY BY COLLABORATION LENGTH

Notes: The overall sample consists of 9,297 pairs of treated and control collaborators confined to a nine-year window around the (inherited) year of star death. The panel is equally unbalanced for treated and control collaborators. Both output measures are co-author adjusted. Forward citations are winsorized at the 99.9th percentile. Collaborators without variation in output over the observation period are dropped by the estimation routine. Field delineation is based on the stars' publication profile or, if available, derived from the classification of the NAS. Robust standard errors are in parentheses, clustered at the level of the star.

* $p < 0.05$, ** $p < 0.01$, and *** $p < 0.001$.

	Overall Sample	Life Sciences	Health Sciences	Physical Sciences	Social Sciences
	Article count as dependent variable				
After death × treated	0.012	0.015	0.002	0.025	-0.153
	(0.030)	(0.050)	(0.047)	(0.045)	(0.266)
After death × treated × star-star dyad	-0.042	-0.096 *	-0.006	-0.073	0.082
	(0.029)	(0.043)	(0.045)	(0.048)	(0.252)
After death × treated × US-US dyad	-0.090 *	-0.063	-0.131 *	-0.018	0.230
	(0.037)	(0.053)	(0.056)	(0.062)	(0.297)
Log pseudo-likelihood	-188,769	-52,666	-79,135	-53,830	-2,984
No. of observations	274,210	83,314	117,559	68,907	4,430
No. of dyads	18,461	5,569	7,892	4,698	302
	Forward citations as dependent variable				
After death × treated	0.006	0.057	-0.010	-0.034	-0.122
	(0.033)	(0.053)	(0.047)	(0.060)	(0.354)
After death × treated × star-star dyad	-0.049	-0.144 *	0.017	-0.076	0.132
	(0.043)	(0.059)	(0.056)	(0.074)	(0.394)
After death × treated × US-US dyad	-0.146 **	-0.173 *	-0.130 *	-0.049	0.223
	(0.048)	(0.076)	(0.063)	(0.082)	(0.416)
Log pseudo-likelihood	-2,785,534	-804,699	-1,147,343	-782,330	-39,282
No. of observations	274,032	83,299	117,495	68,808	4,430
No. of dyads	18,446	5,568	7,886	4,690	302

Tab. A.7: EFFECT HETEROGENEITY VIA COLLABORATOR STATUS AND US CHANNEL

Notes: The overall sample consists of 9,297 pairs of treated and control collaborators confined to a nine-year window around the (inherited) year of star death. The panel is equally unbalanced for treated and control collaborators. Both output measures are co-author adjusted. Forward citations are winsorized at the 99.9th percentile. Collaborators without variation in output over the observation period are dropped by the estimation routine. Field delineation is based on the stars' publication profile or, if available, derived from the classification of the NAS. Robust standard errors are in parentheses, clustered at the level of the star.

* $p < 0.05$, ** $p < 0.01$, and *** $p < 0.001$.

	Overall Sample	Life Sciences	Health Sciences	Physical Sciences	Social Sciences
	Article count as dependent variable				
After death × treated	-0.003	-0.038	0.000	-0.001	-0.077
	(0.028)	(0.039)	(0.043)	(0.034)	(0.268)
After death × treated × co-located	-0.058	0.022	-0.030	-0.135 *	-0.425
	(0.034)	(0.072)	(0.046)	(0.067)	(0.299)
After death × treated × US-US dyad	-0.084 *	-0.083	-0.122 *	-0.006	0.263
	(0.037)	(0.054)	(0.054)	(0.064)	(0.308)
Log pseudo-likelihood	-188,780	-52,673	-79,135	-53,832	-2,983
No. of observations	274,210	83,314	117,559	68,907	4,430
No. of dyads	18,461	5,569	7,892	4,698	302
	Forward citations as dependent variable				
After death × treated	-0.025	-0.047	0.003	-0.073	0.052
	(0.032)	(0.046)	(0.045)	(0.051)	(0.336)
After death × treated × co-located	-0.028	0.138	-0.062	-0.131	-0.632
	(0.053)	(0.082)	(0.082)	(0.109)	(0.382)
After death × treated × US-US dyad	-0.137 **	-0.210 **	-0.109	-0.027	0.239
	(0.047)	(0.077)	(0.061)	(0.084)	(0.406)
Log pseudo-likelihood	-2,789,319	-805,419	-1,149,316	-782,330	-39,180
No. of observations	274,032	83,299	117,495	68,808	4,430
No. of dyads	18,446	5,568	7,886	4,690	302

Tab. A.8: EFFECT HETEROGENEITY VIA CO-LOCATION AND US CHANNEL

Notes: The overall sample consists of 9,297 pairs of treated and control collaborators confined to a nine-year window around the (inherited) year of star death. The panel is equally unbalanced for treated and control collaborators. Both output measures are co-author adjusted. Forward citations are winsorized at the 99.9th percentile. Collaborators without variation in output over the observation period are dropped by the estimation routine. Field delineation is based on the stars' publication profile or, if available, derived from the classification of the NAS. Robust standard errors are in parentheses, clustered at the level of the star.

* $p < 0.05$, ** $p < 0.01$, and *** $p < 0.001$.

	Overall Sample	Life Sciences	Health Sciences	Physical Sciences	Social Sciences
	Article count as dependent variable				
After death × treated	0.004	-0.017	0.015	-0.012	-0.247
	(0.031)	(0.040)	(0.046)	(0.034)	(0.417)
After death × treated × same country	-0.038	-0.038	-0.045	-0.008	0.113
	(0.043)	(0.092)	(0.063)	(0.071)	(0.479)
After death × treated × US-US dyad	-0.069	-0.058	-0.101	-0.021	0.223
	(0.046)	(0.088)	(0.067)	(0.085)	(0.279)
Log pseudo-likelihood	-188,770	-52,672	-79,120	-53,834	-2,986
No. of observations	274,210	83,314	117,559	68,907	4,430
No. of dyads	18,461	5,569	7,892	4,698	302
	Forward citations as dependent variable				
After death × treated	-0.016	-0.060	-0.003	-0.048	0.186
	(0.038)	(0.058)	(0.054)	(0.054)	(0.510)
After death × treated × same country	-0.035	0.081	-0.022	-0.135	-0.533
	(0.051)	(0.091)	(0.068)	(0.110)	(0.524)
After death × treated × US-US dyad	-0.119 *	-0.238 **	-0.099	0.051	0.478
	(0.057)	(0.092)	(0.070)	(0.122)	(0.340)
Log pseudo-likelihood	-2,789,309	-805,386	-1,149,019	-783,525	-39,304
No. of observations	274,032	83,299	117,495	68,808	4,430
No. of dyads	18,446	5,568	7,886	4,690	302

Tab. A.9: EFFECT HETEROGENEITY VIA COUNTRY CHANNEL AND US CHANNEL

Notes: The overall sample consists of 9,297 pairs of treated and control collaborators confined to a nine-year window around the (inherited) year of star death. The panel is equally unbalanced for treated and control collaborators. Both output measures are co-author adjusted. Forward citations are winsorized at the 99.9th percentile. Collaborators without variation in output over the observation period are dropped by the estimation routine. Field delineation is based on the stars' publication profile or, if available, derived from the classification of the NAS. Robust standard errors are in parentheses, clustered at the level of the star.

* $p < 0.05$, ** $p < 0.01$, and *** $p < 0.001$.

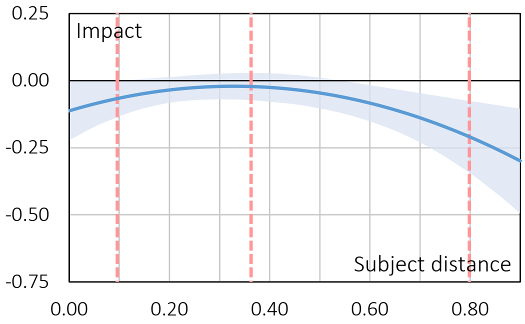
	Overall Sample	Life Sciences	Health Sciences	Physical Sciences	Social Sciences
	Article count as dependent variable				
After death × treated	-0.029	-0.071 *	-0.008	-0.015	-0.105
	(0.026)	(0.029)	(0.040)	(0.035)	(0.150)
After death × treated × subject distance in 3. tertile	-0.058	-0.025	-0.098	-0.031	-0.282
	(0.038)	(0.075)	(0.054)	(0.067)	(0.523)
After death × treated × topic distance in 3. tertile	0.013	0.090	-0.043	-0.006	0.597
	(0.040)	(0.072)	(0.058)	(0.063)	(0.384)
Log pseudo-likelihood	-183,169	-50,859	-77,644	-51,820	-2,685
No. of observations	251,756	75,993	109,817	62,350	3,596
No. of dyads	16,802	5,038	7,311	4,208	245
	Forward citations as dependent variable				
After death × treated	-0.050	-0.125 **	0.017	-0.070	-0.094
	(0.035)	(0.043)	(0.048)	(0.052)	(0.211)
After death × treated × subject distance in 3. tertile	-0.121 *	0.020	-0.161 *	-0.204 *	-0.353
	(0.048)	(0.079)	(0.074)	(0.090)	(0.637)
After death × treated × topic distance in 3. tertile	-0.010	-0.023	-0.119	0.092	0.865
	(0.053)	(0.095)	(0.068)	(0.092)	(0.585)
Log pseudo-likelihood	-2,663,073	-761,129	-1,110,334	-746,954	-32,757
No. of observations	251,736	75,993	109,797	62,350	3,596
No. of dyads	16,800	5,038	7,309	4,208	245

Tab. A.10: EFFECT HETEROGENEITY IN SUBJECT AND TOPIC SPACE

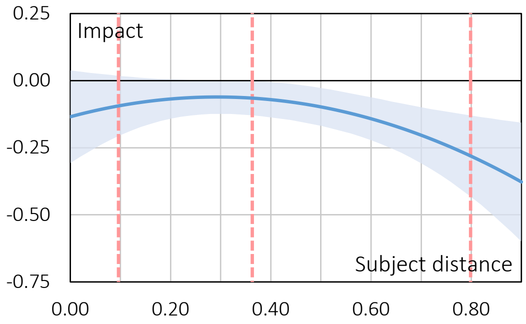
Notes: The overall sample consists of 9,297 pairs of treated and control collaborators confined to a nine-year window around the (inherited) year of star death. The panel is equally unbalanced for treated and control collaborators. Both output measures are co-author adjusted. Forward citations are winsorized at the 99.9th percentile. Collaborators without variation in output over the observation period are dropped by the estimation routine. Field delineation is based on the stars' publication profile or, if available, derived from the classification of the NAS. Robust standard errors are in parentheses, clustered at the level of the star.

* $p < 0.05$, ** $p < 0.01$, and *** $p < 0.001$.

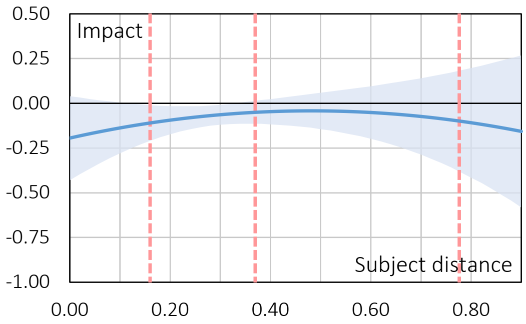
Overall: Article Count



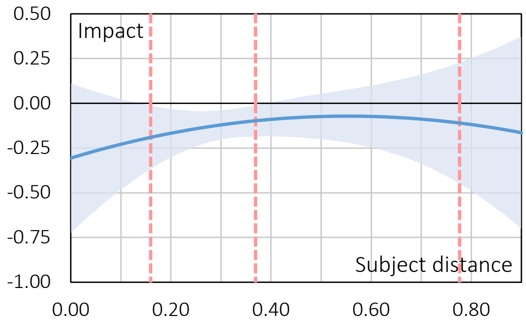
Overall: Forward Citations



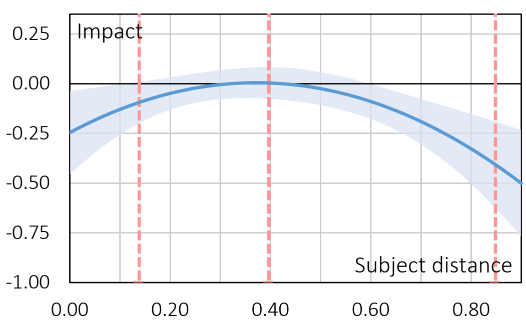
Life: Article Count



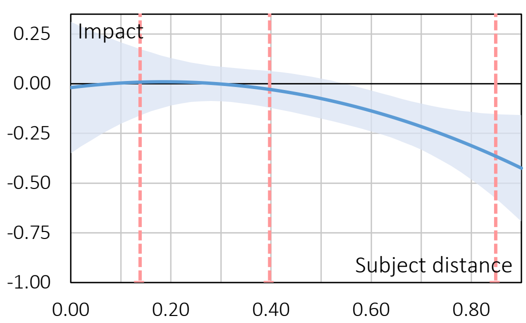
Life: Forward Citations



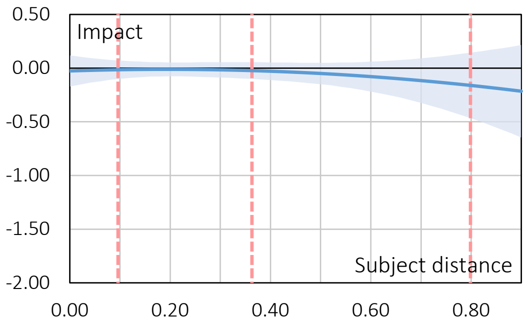
Health: Article Count



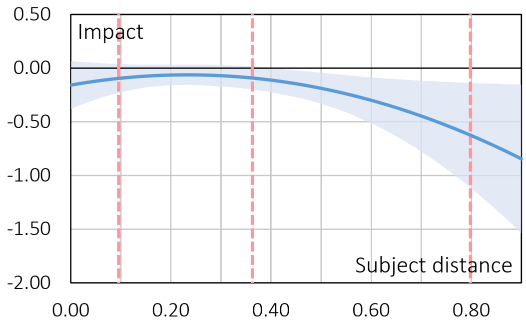
Health: Forward Citations



Physical: Article Count



Physical: Forward Citations



(continued)

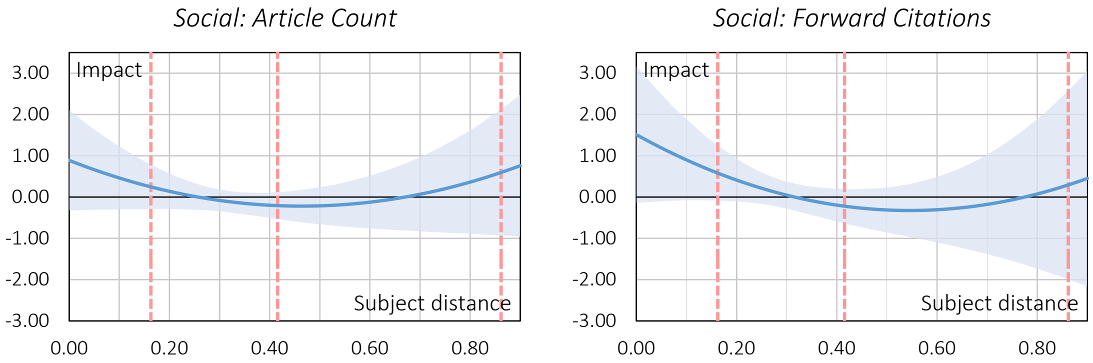


Fig. A.1: EFFECT HETEROGENEITY IN SUBJECT SPACE, CONTINUOUS INTERACTION

Notes: The panels plot the estimated treatment impact over the subject distance range from 0-0.9. In theory, subject distance can reach values up to 1.41 (square root of 2). In practice, however, most distributions are characterised by a thin right tale. The 2.5th, 50th, and 97.5th percentiles are marked by dotted lines to provide reference points. Analogous to all former estimations, treatment impacts result from a Poisson model and thus require transformation to be interpreted as a percentage change (i.e., exponentiating and decreasing by one). Point estimates are depicted by solid blue lines and 95% confidence intervals are pictured as light blue areas. Any further estimation features and variable definitions are maintained from the main model.

Appendix B

B.1 Geographical Depiction

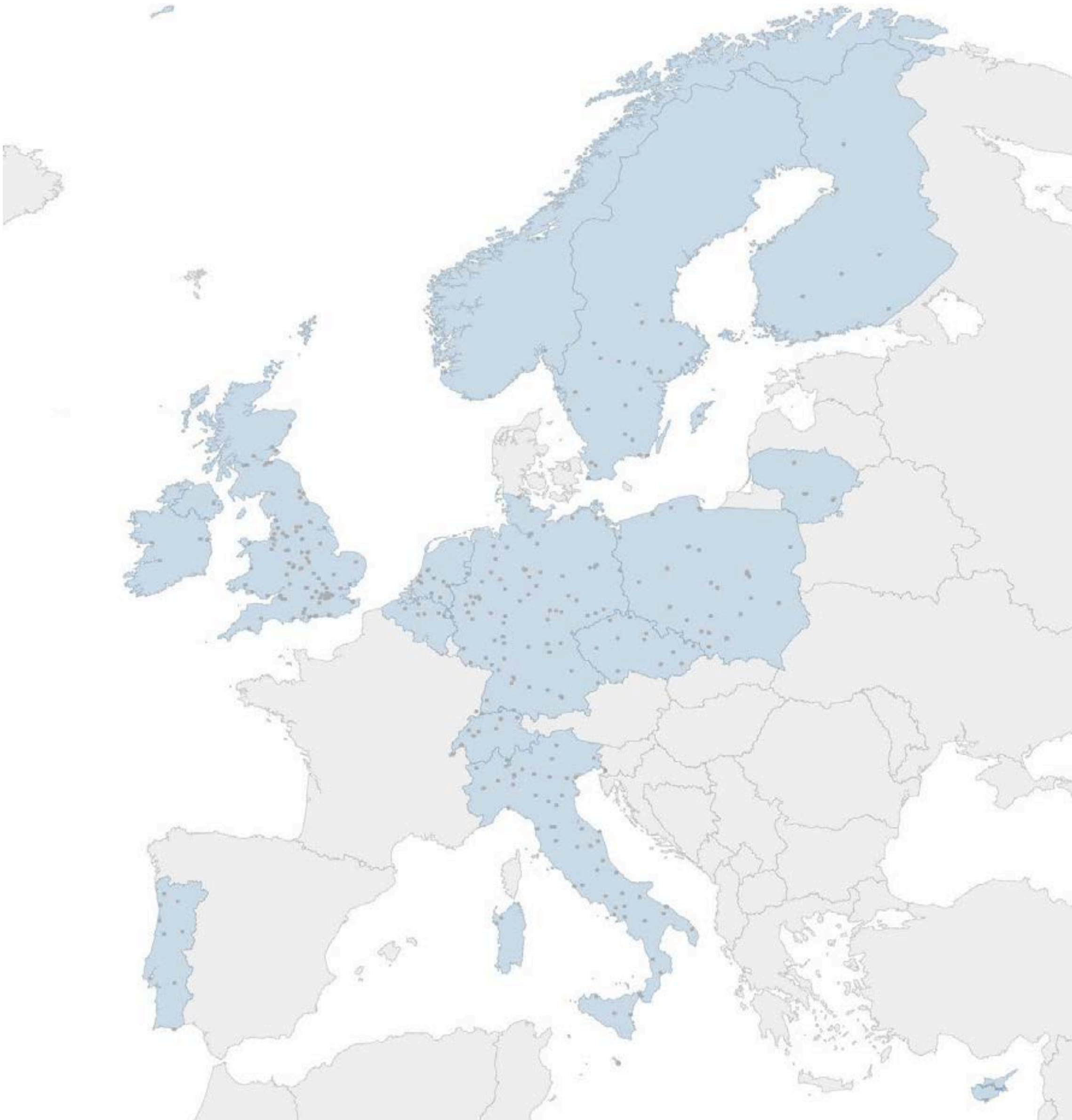


Fig. B.1: MAP OF UNIVERSITY AND COUNTRY COVERAGE

Notes: University locations are marked by dots; their respective countries are highlighted in blue. Special territories of the European Union are omitted from display.

B.2 Cluster Number Evaluation

In order to test our decision for four clusters, we consider a second heuristic, i.e., the Caliński-Harabasz index (also termed pseudo-F). According to a comparative study by Milligan and Cooper (1985), this index performed best among 30 stopping rules and has since become a standard tool in clustering analysis. Technically, it combines compactness and separation in its formula, where the former term refers to similarity within clusters and the latter term to deviation between clusters.

Cluster No.	2	3	4	5	6	7	8	9	10	11	12
Pseudo-F	319.6	351.8	380.1	366.1	356.7	356.3	359.5	362.1	366.3	364.9	363.0
Best count	0	0	874	2	2	12	0	0	100	6	4

Tab. B.1: CLUSTER NUMBER EVALUATION BASED ON CALIŃSKI-HARABASZ APPROACH
Notes: Pseudo-F values are averaged over 1,000 replications of K-means with random starting centres. "Best count" indicates how often a clustering solution was selected by this criterion.

Consistent with our methodology outlined in Section 3.3.2, we ran the K-means algorithm 1,000 times based on the same sequence of random starting centres. The results reported in Table B.1 favour four clusters given that the average Caliński-Harabasz index reaches a maximum for this configuration. In total, we observe a four-cluster solution being selected in 87.4% of our replications. On these grounds, our initial choice can be confirmed.

B.3 Citation Window Analysis

Within our model specification, we employ citations as an indicator of research impact. Since our data cover a citation window of three years, it is worth discussing if this time span complies with long-term impact. For this purpose, we traced our universities back to 1996 and compiled a new dataset consisting of 184,766 publications.

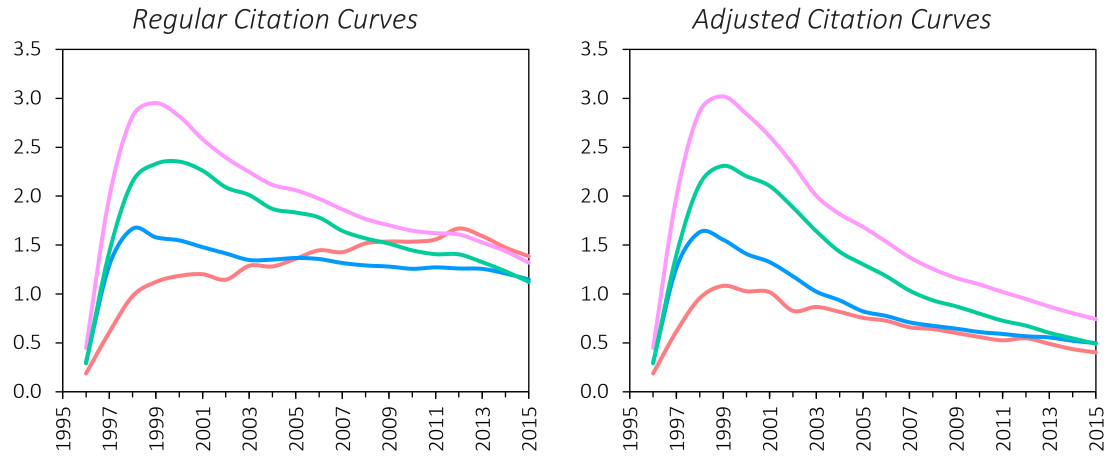


Fig. B.2: REGULAR AND ADJUSTED CITATION CURVES FOR PUBLICATIONS FROM 1996 BY FIELD
Notes: Colouring refers to life (rose), health (green), physical (blue), and social (red) sciences.

As a first step, we calculated citation curves, i.e., average annual citation counts, for the subset of publications with at least one citation over the 20-year period from 1996 to 2015. This subset comprises 86.18% of our initial data. As can be seen in the left-hand panel of Figure B.2, the curves of life, health, and physical sciences follow a similar shape, peaking between 1998 and 2000 followed by a steady decline. Citation counts in social sciences, on the contrary, continue to rise until 2012. From a theoretical standpoint, one might argue that this difference is primarily due to a slower pace of theoretical development (Nederhof, 2006). Interestingly, however, we find this pattern to be largely attributable to a higher growth rate of social sciences within the Scopus database. Once we deflate citation counts by field-specific growth rates, social sciences clearly becomes less of an exemption as illustrated by the right-hand panel of Figure B.2.[60]

Our graphical depiction further reveals citations not only to differ in absolute terms but also regarding the way they mature. This finding could potentially raise concerns about the accuracy of using short-term citations as a predictor for long-term citations. In order to test this relation, we examine correlations between cumulative citation counts over

[60] Following Aizenman and Kletzer (2011), citations counts are divided by a time-varying index, defined as the number of publications in a given year relative to the number of publications in our base year 1996. Of course, indices are calculated separately for each field. Moreover, it should be noted that our adjustment is not based on the full Scopus database but on a comprehensive subset of 15.6 million publications.

increasing time spans, each starting in 1996, and total citation counts. Wang (2013) rightly points out that citation counts are far from being normally distributed, so that Spearman correlations are expected to be most reliable. Yet we also report Pearson correlations to allow comparison with previous studies, e.g., by Adams (2005) or Waltman et al. (2011).

Year	Life Sciences		Social Sciences		Physical Sciences		Health Sciences	
	Pearson	Spearman	Pearson	Spearman	Pearson	Spearman	Pearson	Spearman
1996	0.345	0.350	0.313	0.280	0.141	0.321	0.354	0.333
1997	0.550	0.657	0.510	0.533	0.261	0.616	0.561	0.643
1998	0.647	0.790	0.652	0.693	0.341	0.746	0.655	0.789
1999	0.703	0.854	0.722	0.781	0.403	0.813	0.715	0.859
2000	0.745	0.891	0.772	0.835	0.465	0.857	0.763	0.898
2001	0.781	0.915	0.813	0.869	0.532	0.887	0.805	0.923
2002	0.813	0.932	0.840	0.893	0.601	0.909	0.840	0.940
2003	0.841	0.946	0.870	0.914	0.667	0.927	0.869	0.953
2004	0.864	0.957	0.895	0.931	0.726	0.941	0.894	0.963
2005	0.887	0.966	0.916	0.945	0.782	0.953	0.917	0.971
2006	0.907	0.973	0.936	0.956	0.832	0.963	0.936	0.978
2007	0.924	0.980	0.951	0.965	0.875	0.971	0.950	0.983
2008	0.940	0.985	0.964	0.973	0.909	0.978	0.963	0.987
2009	0.956	0.989	0.975	0.980	0.939	0.983	0.974	0.991
2010	0.969	0.992	0.983	0.985	0.960	0.988	0.983	0.993

Tab. B.2: CORRELATION BETWEEN CUMULATIVE AND TOTAL CITATION COUNTS BY FIELD
Notes: Cumulative citation counts span the period from 1996 up to and including the year given in the first column. Total citation counts cover 20 years (1996-2015).

From Table B.2, we can infer that short-term citations vary in their accuracy as a proxy for long-term citations. In social sciences, for instance, it would require eight years to exceed a Spearman correlation of 0.9, whereas six years would suffice in life or health sciences. Of course, it is hard to define an acceptable level of correlation. However, we might be in a position to circumvent this question. In fact, our research design is not overly concerned about correlations on the publication level, given that we take an institutional perspective. Once we aggregate citations over universities, we discover a considerable increase in the degree of dependence between initial and overall citations (see Table B.3). Apparently, variation is largely cancelled out as becomes evident by almost perfect correlations in all four clusters. This result then leads us to conclude that relatively short citation windows indeed provide a reliable basis for our study.

Year	Social Cluster		Physical Cluster		Health Cluster		General Cluster	
	Pearson	Spearman	Pearson	Spearman	Pearson	Spearman	Pearson	Spearman
1996	0.994	0.858	0.974	0.964	0.983	0.972	0.977	0.985
1997	0.994	0.958	0.977	0.979	0.990	0.981	0.983	0.989
1998	0.994	0.975	0.982	0.984	0.991	0.988	0.986	0.991
1999	0.995	0.984	0.984	0.987	0.992	0.990	0.987	0.992
2000	0.995	0.984	0.987	0.988	0.993	0.993	0.989	0.993

Tab. B.3: CORRELATION BETWEEN CUMULATIVE AND TOTAL CITATION COUNTS BY CLUSTER

Notes: Cumulative citation counts span the period from 1996 up to and including the year given in the first column. Total citation counts cover 20 years (1996-2015). Citations are aggregated by institutions.

B.4 Additional Density Plots

UK, Germany, and Italy constitute the three largest countries in our dataset with 120, 79, and 64 universities, respectively. Given these sample sizes, these countries are best suited for a cluster-specific comparison of density estimates. As illustrated in Figure B.3, UK is characterised by high efficiency levels across all clusters. With the exception of the social cluster, Italy also performs well, quite closely resembling UK's distributions. Germany, in contrast, features consistently lower efficiency levels. Interestingly, Germany's efficiency estimates are rather evenly distributed, thereby suggesting that the national landscape appears very heterogeneous from an efficiency standpoint.

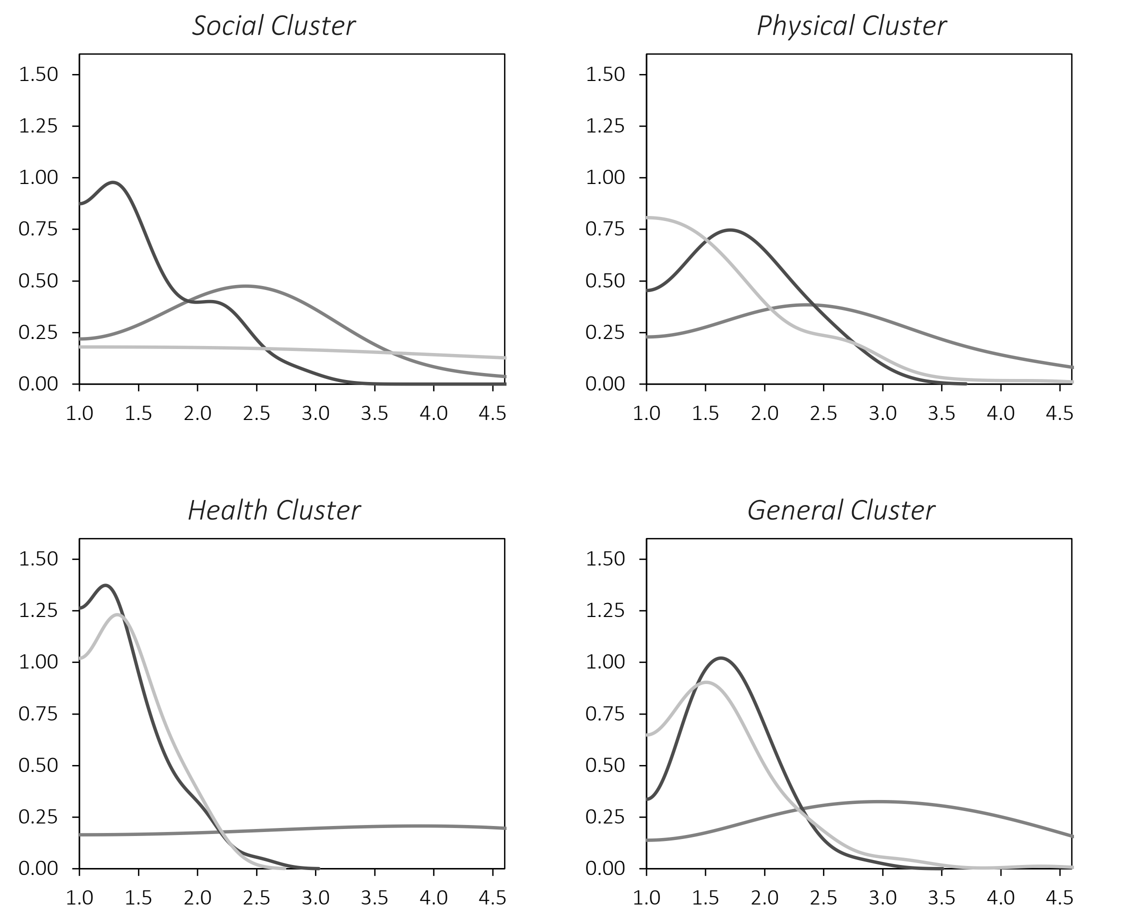


Fig. B.3: DENSITY ESTIMATES OF BIAS-CORRECTED EFFICIENCY SCORES BY CLUSTER AND COUNTRY
Notes: United Kingdom (dark), Germany (medium), and Italy (light) are delimited by greyscale.

B.5 Silhouette Plots

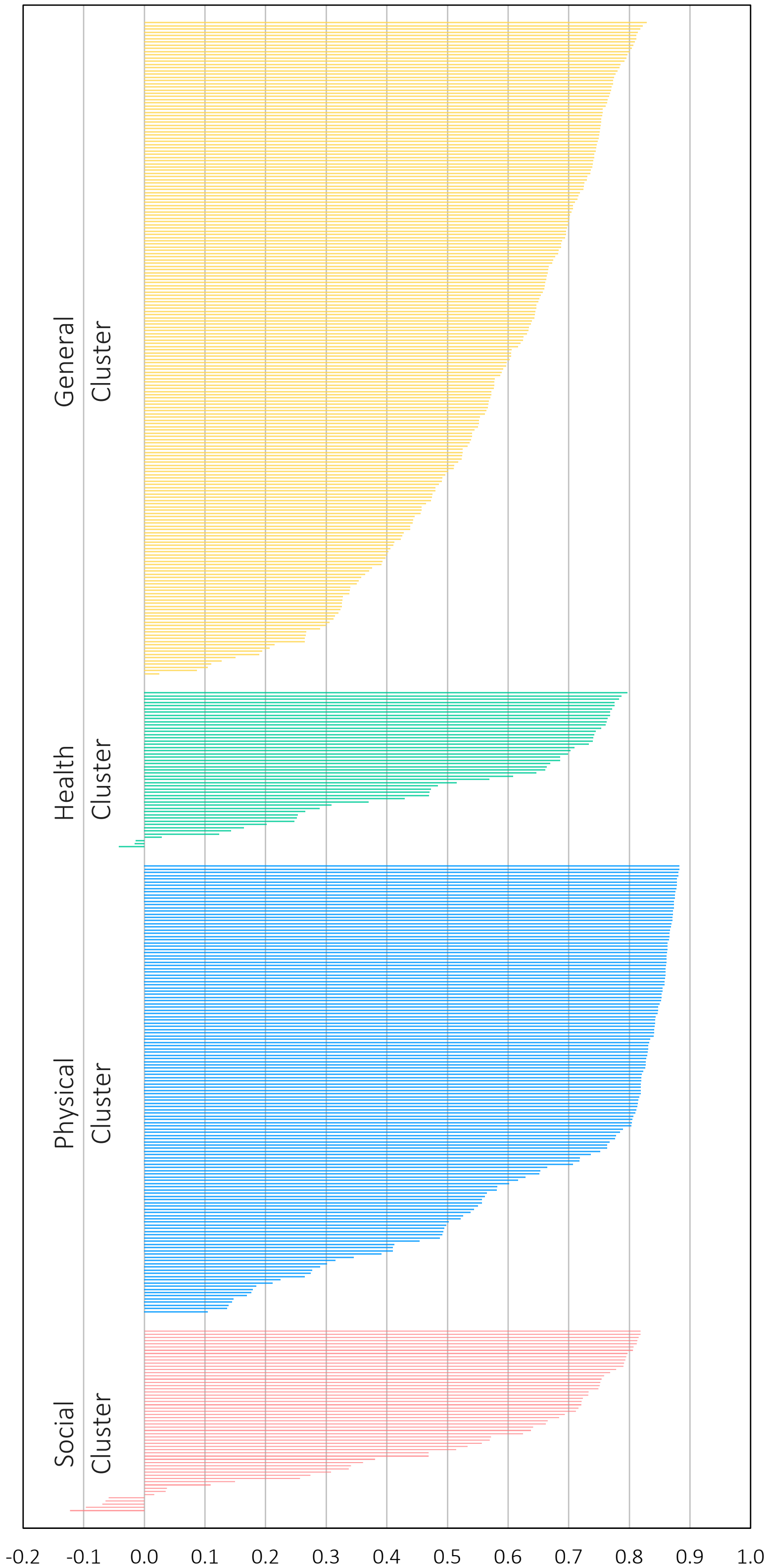


Fig. B.4: SILHOUETTE PLOTS BY CLUSTER

B.6 Time-Lag Regression Design

Reverse causality is well known to hinder clear inference. In the present context, it may arise if universities become more successful in competing for third-party funds as a result of increased efficiency. We attempt to avoid this problem by using time-lagged variables. The idea behind this approach is that, within a given year, funding structures could be affected by efficiency; however, it is unlikely for past funding structures to be subject to the same problem. As this reasoning can also be applied to other variables, we universally employ a time lag of one year. Results of both the clustering and nearest neighbourhood approach are reported below. Overall, our main results are largely persistent.

Variable	Social Cluster	Physical Cluster	Health Cluster	General Cluster
Natural logarithm of bias-corrected efficiency score as dependent variable				
ln(GDP)	- 0.1559 **	0.1396 *	0.0691	- 0.0679
	(0.0518)	(0.0687)	(0.0897)	(0.0395)
Multisite	0.0492	0.1124 *	0.2030	0.0082
	(0.0594)	(0.0479)	(0.1401)	(0.0234)
Hospital	- 0.2850 *	- 0.0458	0.2840 ***	0.1085 ***
	(0.1333)	(0.0581)	(0.0784)	(0.0255)
ln(Size)	- 0.2105 ***	- 0.3369 ***	- 0.1014	- 0.1621 ***
	(0.0536)	(0.0255)	(0.0599)	(0.0208)
Herfindahl	0.0050 *	- 0.0026 **	- 0.0025	- 0.0041
	(0.0024)	(0.0010)	(0.0019)	(0.0030)
Prof	- 0.0209 *	- 0.0269 **	- 0.0619 ***	- 0.0106
	(0.0086)	(0.0086)	(0.0170)	(0.0056)
Female	0.0074 **	0.0051 *	0.0142 ***	- 0.0041 *
	(0.0024)	(0.0023)	(0.0038)	(0.0016)
International	0.0031	0.0023	- 0.0209 *	0.0001
	(0.0024)	(0.0031)	(0.0087)	(0.0021)
Thirdparty	- 0.0146 ***	- 0.0076 ***	- 0.0088 *	- 0.0075 ***
	(0.0038)	(0.0016)	(0.0041)	(0.0016)
Fees	0.0066 **	- 0.0041 *	- 0.0308 ***	- 0.0030 *
	(0.0024)	(0.0018)	(0.0045)	(0.0013)
No. of observations	134	225	106	487

Tab. B.4: REGRESSION RESULTS, CLUSTERING APPROACH WITH TIME LAG

Notes: Results are obtained from 1,000 bootstrap repetitions. Constants as well as time and country dummies are included but not reported. The model employs a one-year time lag, i.e., efficiency scores from year t are regressed on explanatory variables from year t-1. Bootstrap standard errors are in parentheses. * $p < 0.05$, ** $p < 0.01$, and *** $p < 0.001$.

Variable	Social Cluster	Physical Cluster	Health Cluster	General Cluster
Natural logarithm of bias-corrected efficiency score as dependent variable				
ln(GDP)	- 0.0661	0.1415 *	- 0.1244	- 0.0652
	(0.0618)	(0.0713)	(0.0798)	(0.0363)
Multisite	0.1631 *	0.1097 *	0.2766 *	0.0076
	(0.0720)	(0.0486)	(0.1286)	(0.0223)
Hospital	- 0.3313	0.0704	0.3884 ***	0.0870 ***
	(0.1708)	(0.0647)	(0.0705)	(0.0249)
ln(Size)	- 0.1798 **	- 0.3396 ***	- 0.0404	- 0.2133 ***
	(0.0638)	(0.0278)	(0.0556)	(0.0208)
Herfindahl	0.0010	- 0.0027 *	0.0003	- 0.0050
	(0.0028)	(0.0011)	(0.0016)	(0.0030)
Prof	- 0.0133	- 0.0349 ***	- 0.0371 *	- 0.0104
	(0.0110)	(0.0088)	(0.0151)	(0.0054)
Female	- 0.0029	0.0071 **	0.0149 ***	- 0.0046 **
	(0.0028)	(0.0025)	(0.0033)	(0.0016)
International	- 0.0012	0.0034	- 0.0240 **	0.0000
	(0.0028)	(0.0033)	(0.0078)	(0.0019)
Thirdparty	- 0.0134 **	- 0.0083 ***	- 0.0044	- 0.0073 ***
	(0.0046)	(0.0018)	(0.0034)	(0.0015)
Fees	0.0106 ***	- 0.0043 *	- 0.0231 ***	- 0.0024
	(0.0028)	(0.0020)	(0.0036)	(0.0013)
No. of observations	134	225	106	487

Tab. B.5: REGRESSION RESULTS, NEAREST NEIGHBOURHOOD APPROACH WITH TIME LAG

Notes: Results are obtained from 1,000 bootstrap repetitions. Constants as well as time and country dummies are included but not reported. The model employs a one-year time lag, i.e., efficiency scores from year t are regressed on explanatory variables from year t-1. Bootstrap standard errors are in parentheses. * p < 0.05, ** p < 0.01, and *** p < 0.001.

Appendix C

C.1 Robustness Checks

Variable	PROBIT	HETLOGIT	LOGIT 2010
QPT single project funding as binary dependent variable			
ln(Thirdparty)	0.4363 ***	0.8617 **	0.5658 **
	(0.1046)	(0.2908)	(0.1736)
STEM	0.9645	2.3626	1.5210
	(0.5024)	(1.3854)	(0.8146)
ln(Fees)	-0.3249	-0.6086	-0.6070
	(0.6107)	(1.3843)	(0.8960)
Abolition	-0.8251	-1.7564	-1.2559
	(0.4690)	(1.1987)	(0.7526)
ln(Fees) × Abolition	1.2659	2.4749	2.2093
	(1.0133)	(2.2651)	(1.5686)
West	-0.1144	-0.2481	-0.2966
	(0.2990)	(0.6454)	(0.5021)
University	-0.5537	-0.8157	-0.3265
	(0.3999)	(0.8694)	(0.6047)
Education	1.8096 **	3.4925 *	2.6667 *
	(0.6944)	(1.6836)	(1.2407)
Music	0.6752	1.6671	1.1100
	(0.4535)	(1.1687)	(0.7584)
Excellence	-0.2689	-0.2319	-0.2950
	(0.4180)	(1.1228)	(0.7368)
ln(Size)	0.3919 **	0.8681 *	0.6991 ***
	(0.1308)	(0.3966)	(0.2103)
ln(Teachratio)	-0.4174	-0.6566	-0.8461
	(0.3395)	(0.7350)	(0.5699)
ln(Basic)	0.2797	0.8333	0.2178
	(0.2824)	(0.7891)	(0.4398)
Pseudo R^2	0.2570	0.2599	0.2325
F test / Wald test	0.0000	0.0128	0.0000
No. of observations	250	250	250

Tab. C.1: RESULTS OF REGRESSION VARIANTS

Notes: The PROBIT column refers to a probit model. The HETLOGIT column refers to a logit model that permits the residual variance to differ according to $ln(Thirdparty)$. The LOGIT 2010 column refers to a logit model that utilises data from 2010. All remaining features are adopted from the main model. Moreover, constants are included but not reported. Pseudo R^2 is computed according to McFadden (1973). Robust standard errors are in parentheses.
* $p < 0.05$, ** $p < 0.01$, and *** $p < 0.001$.

Variable	PROBIT	HETLOGIT	LOGIT 2010
QPT single project funding as binary dependent variable			
ln(Thirdparty)	0.1206 ***	0.0997 **	0.0949 ***
	(0.0270)	(0.0330)	(0.0277)
STEM	0.2665*	0.2988*	0.2551
	(0.1356)	(0.1382)	(0.1326)
ln(Fees)	0.0401	-0.0770	0.0363
	(0.1355)	(0.1809)	(0.1237)
Abolition	-0.1218	-0.2222	-0.0977
	(0.0723)	(0.1312)	(0.0744)
ln(Fees) × Abolition	0.3679	0.3130	0.3932
	(0.3337)	(0.2862)	(0.3380)
West	-0.0313	-0.0314	-0.0490
	(0.0809)	(0.0848)	(0.0809)
University	-0.1459	-0.1032	-0.0543
	(0.0979)	(0.1212)	(0.0994)
Education	0.3186 ***	0.4417 *	0.2940 ***
	(0.0625)	(0.2035)	(0.0737)
Music	0.1605	0.2109	0.1581
	(0.0889)	(0.1230)	(0.0874)
Excellence	-0.0752	-0.0293	-0.0501
	(0.1177)	(0.1436)	(0.1265)
ln(Size)	0.1083 **	0.1098 **	0.1173 ***
	(0.0346)	(0.0354)	(0.0331)
ln(Teachratio)	-0.1154	-0.0830	-0.1419
	(0.0937)	(0.0995)	(0.0954)
ln(Basic)	0.0773	0.1054	0.0365
	(0.0775)	(0.0795)	(0.0737)
No. of observations	250	250	250

Tab. C.2: AVERAGE MARGINAL EFFECTS OF REGRESSION VARIANTS

Notes: The PROBIT column refers to a probit model. The HETLOGIT column refers to a logit model that permits the residual variance to differ according to $ln(Thirdparty)$. The LOGIT 2010 column refers to a logit model that utilises data from 2010. All remaining features are adopted from the main model. Robust standard errors are in parentheses.

* p < 0.05, ** p < 0.01, and *** p < 0.001.