Hochschulschriften

Alexander Kaiser

On the Influence of Institutional Division of Labor and Specialization on Scientific Productivity

metropolis

Alexander Kaiser

On the Influence of Institutional Division of Labor and Specialization on Scientific Productivity

Hochschulschriften

Band 171

Alexander Kaiser

On the Influence of Institutional Division of Labor and Specialization on Scientific Productivity

Metropolis-Verlag
Marburg 2025

Bibliografische Information Der Deutschen Nationalbibliothek
Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über <https://portal.dnb.de> abrufbar.

Metropolis-Verlag für Ökonomie, Gesellschaft und Politik GmbH
https://www.metropolis-verlag.de

ISBN 978-3-7316-1599-6

Contents

A. List of Abbreviations

DoL	Division of Labor
Spec.	Specialization
CPM	Citation Productivity Model
Den.	Denominations
PPM	Publication Productivity Model
Spec. Conc.	Specialization Concentration
Spec. Grav.	Specialization Gravity
Task Coord.	Task Coordination
Task Div.	Task Division
AUC	The University of Auckland
CAL / Caltech	California Institute of Technology
COL	Columbia University in the City of New York
ETH	Eidgenössische Technische Hochschule Zürich
GAU	Georg-August Universität Göttingen
HAR	Harvard University
LEE	University of Leeds
LMU	Ludwig-Maximilians-Universität (München)
MIT	Massachusetts Institute of Technology
RFW	Rheinische Friedrich-Wilhelms-Universität Bonn
STA	(Leland) Stanford (Junior) University
UCB	University of California, Berkeley
UCD	University of California, Davis
UCL	University of California, Los Angeles
UCS	University of California, San Diego
UoC	University of California

UOS	The University of Sydney
UOW	University of Washington (at Seattle)
UPP	Uppsala Universitet
UTA	The University of Texas at Austin
UZH	Universität Zürich

Attr.	Attributable
CWTS	Centre for Science and Technology Studies
DEA	Data Envelopment Analysis
Destatis	Statistisches Bundesamt
ETER	European Tertiary Education Register
FDH	Free Disposal Hull
GPR	Gaussian Processes Regression
HEI	Higher Education Institution
KMO	Kaiser-Meyer-Olkin
LSCV	Least Squares Cross Validation
Misc.	Miscellaneous
MSE	Mean Squared Error
NSDAP	Nationalsozialistische Deutsche Arbeiterpartei
OECD	Organisation for Economic Cooperation and Development
PLoS	Public Library of Science
RMSE	Rooted Mean Squared Error
THE	Times Higher Education
WoS	Web of Science

B. List of Tables

C. List of Figures

1. Introduction

As of late, evidence is accumulating that scientific progress has come to a halt (Bloom et al. 2020; Chu and Evans 2021; Cui et al. 2022). Park et al. (2023) just recently published an article in *Nature*, providing convincing empirical evidence that combinatorial novelty of publications is declining, making publication of atypical papers less likely and that science is becoming less disruptive across all major fields. Apart from these qualitative concerns, researchers are even debating potentially declining (quantitative) productivity levels (given a continuously growing scientific community) contributing towards regressing progressiveness (Abramo and Angelo 2023; Cauwels and Sornette 2022; Shkliarevsky 2022)

Explanations for all kinds of pathologies regarding scientific inquiry are debated in the field of science studies. One example is 'Newton's shoulders of giants' argument. Future scientific progress is believed to build on the previously accumulated knowledge stock, which may be pictured as seeing only farther because we are standing on the shoulders of our predecessors' seminal works. Yet as the accumulated knowledge stock grows and most 'low-hanging fruits' are yielded, it gets harder to climb the shoulders of the giant (or to catch up with the contemporary state of research). Over time, achieving scientific progress thus gets more resource-intensive and complex, which over the past decades provoked a shift towards research increasingly being conducted in larger teams. (Furman and Stern 2011; Park et al. 2023)

Jones (2009) described this as the 'growing knowledge burden' which eventually led to 'the death of the renaissance man'. Consequently, contemporary philosophy of science promotes a perspective of social epistemology (rather than individual epistemology) and evaluates the epistemic consequences of relations among collaborating scientists and the institutional arrangements they are confronted with (Goldman and Blanchard 2016)'. Central topics in this area are 'testimony', 'peer disagreement', 'group belief and justification' as well as formal modelling of interactions within the epistemic community (Goldman and Blanchard 2016).

From an economist's point of view, a different aspect in context of institutional arrangements and systemic relations comes to mind given the above-described change of how science is conducted. The shift from a science centered around an individual (or polyhistor) to research in large teams implicates increasing costs of coordination caused by an enforced division of labor and specialization. The larger the body of accumulated knowledge becomes, the narrower the scope of research that a single researcher can oversee if he successfully wants to 'climb the giant's shoulders'. The narrower the scopes of researcher's specialties become though, the more they need to be institutionalized in delineated tasks and coordinated (both in research and teaching). Today's science thus requires for an enforced and differently organized collaboration among scientists and specialties with an enforced institutional division of labor. In economic theory, division of labor, specialization and the institutional arrangements regarding the management of coordination costs are important determinants of productivity levels and (technological) progress. It seems reasonable to suppose that this is at least to some extent applicable to scientific institutions as well. Potentially even, some of the observed pathologies could be explained by anomies in division of labor and specialization within scientific institutions.

Surprisingly, however, a thorough analysis of the effect of institutional division of labor on any form of epistemic outcome (qualitative or quantitative) is missing so far. Given the good evidence of the effect of related topics like e.g., interdisciplinary research on scientific production and the inconsistencies of the rationality-based theory on an efficient cognitive division of labor with empirical evidence in scientometrics' studies, this research gap needs to be closed. (see ch. 3) Thus, this work seeks to answer the research question if whether division of labor and specialization are determinants of epistemic outcomes that unjustly have been neglected in science studies and bear the potential to explain pathologies in the scientific production process?

The latter research question can be divided into two separate objectives. For one, it will be examined if institutional division of labor and specialization are indeed determinants of epistemic outcomes. This will be achieved by operationalizing the two phenomena using a new dataset and by conducting a thorough descriptive and quantitative analysis to identify path dependencies created by initial configurations of DoL and Spec. Secondly, if the latter is the case and DoL and Spec. indeed create path dependencies, it will be assessed if their influence on scientific productivity is necessarily efficient (as the existing paradigm within science studies suggests, which supports the idea that the self-governing scientific community allocates its cognitive labor efficiently). If this is not the case and structural effects of DoL and Spec. on efficiency can be derived, it is concluded that they are neglected determinants of epistemic outcomes, which can explain part of the pathologies in science observed.

This work is organized as follows. In chapter 2, the theoretical line of thought is presented, reviewing acknowledged determinants of epistemic outcomes in the science studies and motivating that DoL and Spec. need to be accounted for, because of their potential to explain pathologies in science. In chapter 3, a new dataset providing microdata of 20 renowned and highly ranked universities for the period 1890 to 2020 is introduced to identify (university types and) path dependencies created by DoL and Spec. In chapter 4, a state-of-the-art nonparametric conditional framework is employed to examine the (functional form of the) effect of DoL and Spec. on scientific productivity of the latter universities. Finally, the work closes with a discussion of results and conclusions in chapter 6 and 7.

2. On the Neglection of Institutional DoL and Spec. in the Science Studies

2.1 *Literature Review (Science Studies)*

2.1.1 *Epistemic Outcomes*

Before DoL and Spec. can be established as neglected determinants of epistemic outcomes, a brief introduction into the factors that are acknowledged and accounted for in the science studies is inadmissible (2.1.1). Further, all topics which are to some extent connected to division of labor and specialization in context of scientific inquiry e.g., studies on cognitive diversity among researchers, are introduced to outline the differences of existing studies to this work's scope and perspective (2.1.2). Once the current state of research is outlined, theories of influential thinkers like Adam Smith and Emile Durkheim and their thoughts on DoL and Spec. in context of scientific inquiry will be revisited to motivate the institutional perspective supported in this work (2.2.1). After establishing that the latter is not accounted for in the existing literature (2.2.2), it will be argued that DoL and Spec. nonetheless have the potential to explain some of the pathologies (outlined in the introductory chapter) and observed in today's science (2.2.3). Consequently, the research question can be based on the idea of DoL and Spec. as wrongfully neglected determinants of epistemic outcomes.

The science studies comprise the branches of philosophy, sociology and economics of science, as well as the scientometrics (and bibliometrics) literature, which are dedicated towards understanding researchers' motives and behavior, institutional arrangements, communication and collaboration (networks) as well as social norms of scientific inquiry and how to operationalize them in empirical studies. (Goldman and O'Connor 2021)

In contemporary philosophy of science, the literature on individual epistemology has been superseded by the perspective of social epistemology (Mayo-Wilson 2011). The latter is concerned with the assessment of the impact of institutional arrangements on epistemic outcomes. Those relate to topics such as testimony, peer disagreement, as well as group belief and justification (De Ridder 2014). Formal models proposed for explanation are the credit economy, where scientists are characterized as utility maximizers or are concerned with modelling diversity in epistemic communities (e.g., Singer 2019; Hong and Page 2004; De Langhe 2010). (Goldman and O'Conner 2021)

The sociology of science, based on the pioneering work of Robert K. Merton, is concerned with the relationship between the elements of the social and the normative structure of science (Merton 1973e). Initially, this concerned such institutions as the peer-review system and the role of editors (Merton and Zuckerman 1973a; Storer 1973a), behavior patterns and reward systems (e.g., Nobel prizes) (Merton 1973a; Storer 1973), priority in discovery (e.g., Robert Hooke vs. Isaac Newton) (Merton 1973), age structure (Merton 1973a), the importance of multiples (discovery of equal epistemic outcomes in different locations at the same point in time) (Merton 1973b, c), as well as the *Matthew effect* of cumulative advantages (Merton 1973d). Most of the topics brought up by Merton are still important in today's sociology of science. Jones and Weinberg (2011) for example modeled the age-creativity relationship in science, Kwiek (2019) analyzed the *Matthew effect* for the case of project funding in the EU and Romero (2020) just recently dealt with the issue of multiples in his work on the replicability crisis in the biomedical sciences (Bechtel 1993), which he connects to pathologies between the normative and the reward system of science (Storer 1973).

Another line of seminal works are the *laboratory studies* conducted by Latour and Woolgar (1986) as well as Knorr-Cetina (1984), which moved the abstract perspective from (individual epistemology) promoting the genius (polyhistor) pushing the frontier of science on his own (from within the ivory tower), to a more realistic[1] perspective on science by observing the daily work 'done by a scientist located firmly at his laboratory bench (Latour and Woolgar: 27)'. By imitating the methodology of anthropologists, they showed the dependence of scien-

[1] At least more realistic for (post-)modern scientific inquiry outside of the humanities.

tific facts and norms on the profanity of social interactions and institutional arrangements (e.g., when local regulations on employment law predetermine the available procedures for conservation processes (Knorr-Cetina 1984: 72)). The most crucial innovation of the laboratory studies though, supposedly lies in the rational description of scientific inquiry as a production process, where scientists interact in the production plant of the laboratory to recombine materials and prefabricated publications to fabricate facts and produce new publication output, which is in demand among the scientific community. This way, the already established ideas on rewards and recognition could be smoothly integrated in what has today developed into science studies, which are dominated by economic thought, methodology and terminology.[2] (Knorr-Cetina 1984; Latour and Woolgar 1986)

Indeed, in both sociology of science and social epistemology, scientists are characterized as credit maximizers (for the sake of their own career path), which is constitutive for the normative structure of science and for an ongoing production of epistemic outcomes. According to Latour and Woolgar (1986) we need to differentiate here between credit as reward and credit as credibility. The former means the recognition received by peers for previous works, whereas the latter denotes the ability of a scientist to actually produce valuable epistemic outcomes. The authors convincingly argue, that if only reward was considered as being constitutive for scientific norms and the production of epistemic outcomes, we could not understand the efforts scientists put into education, skills, moving in location or position as well as networking, which often come without recognition at all or delayed rewards in later career stages. From a sociological point of view, credibility is thus the factor that allows to relate external factors like rewards and recognition from institutions to the 'substance of scientific production (of facts) (Latour and Woolgar 1986: 198)'.[3]

On the level of the individual researcher, determinants of epistemic outcomes are often embedded in context of the discussion on scientific

[2] See Polanyi et al. (2000) for a treatise on economic theory in the 'republic of science' and Zollman (2018) for a discussion of economic rationality in science.

[3] Or as Paul Samuelson famously stated in a presidential address to fellow economists: 'In the long run, the economic scholar works for the only coin worth having – our own applause. (Merton 1973a: 339)'

progress. According to Thomas Kuhn (1976), scientific inquiry may be divided into two phases, (1) normal science dominated by puzzle-solving activities and (2) phases of scientific revolutions, where existing paradigms are substituted by structurally different ones, which create a new set of puzzle-solving activities. Lin et al. (2022) have argued that since this dichotomy creates a tension, where one must be traded for the other, scientists are confronted with different choice situations, possibilities to pursue research and career outcomes. The characterization of types of researchers is not exclusive to the sociology of science but is an important part of the debate on models of diversity in epistemic communities promoted in social epistemology (See e.g., the characterization of scientists as 'Mavericks' and 'Followers' in Weisberg and Muldoon (2009) or 'explorer' and 'extractor' types in Thoma (2015)). Another issue in context of an individual researcher's career path is the probability of being involved in different scientific communities according to age, how this relates to recognition and the relationship with academic supervisors (Zeng et al. 2019). The latter concerns in particular order of author positioning in papers, which varies according to the prevailing culture of a particular field (Stephan 2012).

A considerable share of the literature in the sociology of science is concerned with publications, as the primary mean of scientific communication and output. Prestige and authority of journals, the boundaries they define and how the latter are connected e.g., in interdisciplinary research, is debated here (Bechtel 1993). Other fringe topics are, e.g., the issue of so-called *sleeping beauties*, meaning publications, which initially receive no or very low attention from peers, yet later receive disproportionate amounts of recognition (Lin et al. 2022), as well as analysis of publication and citation patterns of Nobel laureates (Li et al. 2019). Studies on communication patterns within the scientific community are often operationalized by the so-called *bibliometric hypothesis* (Zitt et al. 2019), which denotes the idea that communication and collaboration networks can be examined based on information retrieved from publications (databases) such as coauthorship, text relations and citations. The latter paradigm is dominating the *Scientometrics* literature, which deals with quantitative analysis of scientific indicators, like for example the Science Citation Index (see e.g., Van Raan (2019) for a brief introduction). Another important notion in this context is that of so-called *invisible colleges*, which means the (not institutionalized) networks of communication of scientists, constituting a paradigm of a research domain (Clark 1970). Constitutive elements of such paradigms are the formation of disciplinary boundaries in the sense of the social and communicative isolation of a part of the scientific community. (Weingart 1974)

Normative structure, publication process and invisible colleges are embedded in institutional contexts and spark the creation of new ones (Clark 1970; Weingart 1974). Scientists operate within such institutions as the university, academic societies and the institutionalized structures within them (e.g., academic departments or the laboratory). The latter influence opportunities and distribution of funding resources, the hierarchies within academic staff (full tenured professor vs. non-tenured research associate) and with students (Bourdieu 1975), as well as trade-offs in the value attached to teaching responsibilities, as opposed to gaining recognition and reputation for the affiliated institution. (Bourdieu 1975; Reif 1961)

Finally, whereas economic thought plays a remarkably important role in all of science studies (scientometrics, philosophy and sociology of science), a separate economics of science branch exists, which is mainly focused on competitiveness and transdisciplinary aspects of the scientific production process such as e.g., knowledge transfer and technological progress, industry collaboration and competition with non-traditional research organizations, as well as funding and cost efficiency related aspects like R&D spendings (Eaton and Stevens 2020; Stephan 2012; Petrella 1992).

2.1.2 Topics related to DoL and Spec.

In this subsection, topics of the science studies will be addressed that come closest towards a consideration of DoL and Spec. as determinants of epistemic outcomes or are somehow related. One example for where division of labor is directly referenced is in the debate on the so-called *cognitive division of labor*. Originally introduced by philosopher of science Phillip Kitcher (1990), the evolving literature stream promotes a rational choice-based theory, which claims to explain the (assumed) efficient allocation of researchers across research issues and methods. Core idea is that there are (well-documented) instances in (the history of) science, where researchers reasonably disagree on the importance of research issues, the methodologies that should be employed, or the theories developed to resolve them. Kitcher argues that in such cases (of peer disagreement), it might be desirable and also, more importantly, it is de facto the case, that researchers do work on all relevant research issues and pursue all promising methodological and theoretical paths. Simply put, the scientific community divides its cognitive labor efficiently and Kitcher develops a model that allows to explain how the latter comes about. Congruent with the economic paradigm in the science studies as a whole, Kitcher proposes a rational theory-based model, where scientists are characterized as credit maximizers. Here, credit above all means recognition and rewards in forms of promotion, grants and citations linked to the successful pursuit of a particular research issue. Given a research issue is promising, in the sense of a good availability of methods that will allow to achieve the desired epistemic outcome and the latter being valuable to the scientific community (or beyond), the probability for the latter to be pursued by a (truth-seeking) researcher is higher than for a research issue that is less relevant and for which the methods available are less promising. Assuming rational behavior, Kitcher argues that the credit maximizing scientist will pursue the latter research issue when the expected credit is higher than that of the former research issue. In conjunction with the priority rule (Streven 2003), which states that the expected credit is negatively correlated to the number of scientists working on the same research issue, scientists will allocate themselves over all issues, methods and theories efficiently. In plain language, Kitcher provided a model that validates the idea of the scientist as a credit maximizer e.g., promoted by the authors of the laboratory studies, as opposed to the (outdated individual epistemological) idea of the scientist as merely interested in truth-seeking. (Kaiser 2023; Kitcher 1990)

Kitcher's work started a flourishing literature stream and provoked proposals of more sophisticated models, where scientists act as rational agents making choices given fixed rules, examining the way cognitive labor is distributed. While most of the latter studies focused on the evaluation and modelling of cognitive diversity, assessing how distinct evaluative attitudes lead to different strategies to solve a problem (e.g., Alexander et al. 2014; D'Agostino 2009; Muldoon and Weisberg 2011; Muldoon 2013; Pöyhönen 2017; Thoma 2015; Weisberg and Muldoon 2009), studies also treated the issue if Kitcher's model relies on a particular understanding of the history of science (Viola 2015), cases where project selection is subject to imprecise parameters (Zhang et al. 2014), the varying objectives of funding practices potentially hindering the assumed self-governing of scientists (Bedessem 2019), as well as the development of a fragmented knowledge account, dealing with how individuals' beliefs can be reasonably aggregated (Habgood-Coote 2019). While those studies represent a lot of interesting and valuable puzzle-solving activities linked to the initial model, with a particular emphasis on how cognitive diversity ensures an efficient division of labor within the scientific community, it is quite surprising that no one ever challenged the assumption of Kitcher, that cognitive labor is indeed efficiently allocated in science. Until today, this is still an uncontested idea (paradigm) in social epistemology.

Diversity also plays an important role in analyses of specialization and interdisciplinarity. Liu et al. (2021) for example argue that the knowledge burden, aggravated by the ongoing accumulation of knowledge, forces researchers to focus on their own field, favoring the isolation of fields from one another resulting in highly specialized disciplinary silos[4]. By reconnecting the latter, interdisciplinary collaboration is believed to improve the chances to successfully tackle complicated research issues (Liu et al. 2021). On the contrary, increased efforts necessary for knowledge integration lowering productivity, as well as a lack of suited publication venues (, which are mainly disciplinary) are considered as potential penalties for conducting interdisciplinary research (Leahey et al. 2017). The latter is empirically evaluated by examining joint publication of authors from different disciplines, disciplinary belongings of cited references (e.g., Leahey et al. 2017; Lu et al. 2020) or concentration of publications on research domains (e.g., López-Illescas 2011), which is operationalized among others by employing diversity indices (e.g., Zitt et al. 2019).

Differences in specialization according to discipline are examined as well. Leahey and Reikowsky (2008) assessed for the example of

[4] Even though seldom, some authors point at disciplines not being as siloed as commonly believed when analyzing changes in formation of disciplines over the course of extended periods of modern science (e.g., Jacobs 2017).

sociology potential trade-offs in between specialization and collaboration. Their results suggest that within sociology, specialization in depth is more common than coordination of research with sociologists working in the same or different areas of expertise. Further, Pierce (1992) for example showed that, as disciplines mature, bibliometric features of research articles become increasingly similar, indicating a sort of convergence in disciplinary language due to specialization. Interdisciplinarity is also approached from a historical case-study perspective on the formation of a discipline by recombining institutionally separated research domains (see e.g., Bechtel (1993) for the emergence of cell biology, Ben-David and Collins (1974) for psychology, Latour and Woolgar (1986) for neuroendocrinology or Mullins (1974) for molecular biology).

Aspects of division of labor and specialization are further implicitly addressed, for example in the *laboratory studies*. Frequently, remarks refer to some form of task division and coordination within (the micro-level) of a group, project or team. Walsh and Lee (2015) for example pointed out that science is increasingly becoming a team activity, where the size of teams in some fields may be compared to that of medium-sized firms. They further argue that their 'quasi-firm'-like state comes alongside more bureaucratic structuring within the team, such as task division, hierarchy, decentralization and standardization. These changes in scientific inquiries from an individual endeavor to large group and project-based activities is well-reflected by the ongoing rise of numbers of authors per paper over the past decades. Also, the interdependence within teams varies according to research field, where the group size in science and engineering domains is greater than e.g., in social sciences and humanities. (Walsh and Lee 2015) Studies on division of labor within (micro-level) teams were conducted e.g., by Shibayama et al. (2015) for life science research labs, Adams et al. (2005) to assess the relationship of team size and received awards, and by Walsh et al. (2019) to analyze the relationship of group size with retractions of papers.

Alternatively, team DoL has been assessed based on the *bibliometric hypothesis* using author contributorship statements as e.g., in Jabbehdari and Walsh (2017), Larivière et al. (2016), Macaluso et al. (2016) and Wuchty et al. (2007) (See section 3.1.2.2 for details). Here, the term division of labor is explicitly used for differentiating scientists' involvement in the tasks of writing, data collection and performance of the actual empirical analysis. The studies suggest that DoL within teams is higher in the medical sciences when compared to physics and the social sciences for example. Further, there seems to be a difference according to seniority (serving the traditional Mertonian topic of the age structure in the science studies) with senior researchers performing more conceptual tasks and younger researchers more technical ones. Regarding the reward system in science, the studies suggest that first and last-mentioned authors indeed contribute to more tasks than middle authors (and thus rightfully receive more credit by the scientific community).

Finally, Häussler and Sauermann (2020), as well as Cummings and Kiesler (2014) analyze the interaction of DoL in teams with interdisciplinarity. These studies suggest that team size (, which is correlated with higher DoL) is positively correlated with interdisciplinarity of its members among others because of emerging topics, such as e.g., 'computational biology', requiring for an integration of expertise and knowledge from different research domains.

2.2 Why DoL and Spec. are neglected determinants of epistemic outcomes

2.2.1 Adam Smith, Emile Durkheim and other thinkers revisited

The topics outlined in the previous section consider very particular aspects of division of labor and specialization. In the literature on *cognitive division of labor* it is discussed whether Kitcher's model indeed concerns division of labor as originally popularized by Adam Smith. For one thing, Alexander et al. (2014) explicitly motivate their work with the example of the pin factory, then again Muldoon (2013) and D'Agostino (2009) claim that the two are entirely different phenomena. Habgood-Coote (2019) in turn, states that cognitive division of labor refers to 'a Smithian phenomenon of scientific competition (925p)' at least assuming a connection between the two. Therefore, in this section, the theories of outstanding thinkers on division of labor and specialization will be revisited (, with a special emphasis on what they had to say about division of labor in scientific contexts) to clarify what DoL and Spec. in context of the scientific production process truly means and to identify potential research gaps in the science studies.

Interestingly, theories on division of labor are as old as scientific inquiry itself. Plato's conception of the ideal state in the *Politeia* is based on an optimally designed division of labor (Schumpeter 1994: 51p). In fact, division of labor is of major concern in a lot of utopian thought, be it in an ideal conception of it in Thomas More's 'Utopia' and Bacon's 'Republic of Science' (Bacon 1902) or in advocating for its complete abolishment based on the 'alienation theory' promoted in Marx's 'Das Kapital'. The notion of division of labor may sound outdated, but it has always been relevant in debates on urbanization (Gibbs and Martin 1962), industrial productivity (Smith 1978; West 1999), management and organizational design (Kamijo and Nakama 2023; Meier et al. 2019; Raveendran et al. 2016) or even biology (De Oliveira and Campos 2019; Rueffler et al. 2012). Another interesting question, in particular given the scope of this work, is the one of priority of discovery when it comes to our understanding of the phenomenon. Cases are made i.e. for the *Muqaddimah* of Ibn Khaldun (Al-Hamdi 2006; Boulakia 1971) and Adam Ferguson's (1767) *An Essay on the History of Civil Society*, whose teachings have influenced Adam Smith (Bücher 1922; Merton 1973c).

Regardless of whether the first modern account of DoL should be attributed to the work of Ibn Khaldun or not, Adam Smith's case of the pin factory is undoubtedly the most influential illustration of division of labor. This might partly be explained by its positioning in the first chapter of the *Wealth of Nations* (1804), as paradigmatic publication shaping the domains of modern economics, partly by its intuitive appeal for understanding the relationship of division of labor with productivity. What is often ignored in this context is that Smith differentiated the notion of division of labor from specialization. In the former he saw first and foremost the institutional and organizational differentiation of tasks, whereas the latter is the (anthropologically motivated) natural tendency of individuals to specialize in tasks in which they have comparative advantages. In the sociological realm, Emile Durkheim's The *Division of Labor in Society* (1966) is the most influential work on DoL. Inspired by Darwin's theory, Durkheim's main concern was to regress solidarity in advanced societies on the interdependencies created by division of labor. His main contribution regarding the effect of division of labor on productivity are certainly his examples for abnormal division of labor like e.g., anomic division of labor. Durkheim's abnormal forms were coordinated with Smith's theory by Becker and Murphy (1992), which introduced the notion of the coordination costs. While their contribution is certainly not comparable in importance with the seminal works of Smith and Durkheim, it is the most original one since and seminal insofar as it reintroduced the (philosophically derived) old theories in modern formalized economic terms.[5] One thinker that should certainly be mentioned here as well is economist Karl Bücher (1922), who provided the most thorough and detailed conceptualization of division of labor, retracing its development historically, categorizing and classifying all its types (e.g., according to ownership of the productional means within a production process (or supply chain)) and integrating division of labor on societal, firm and household level continuously.

That Smith and Durkheim put the analysis of division of labor at the forefront of their works, making it the basis of their theories, illustrates the importance of the phenomena to explain any form of (modern) human interaction quite well,[6] be it economic or social (if one sees fit

[5] Following the above discussed perspective on recognition from peers, it can be noted that their article was published in the Quarterly Journal of Economics and according to google scholar received over two thousand citations up until today. Given the latter, it should be fair to say that their contribution belongs to the most important publications on division of labor and specialization in the last decades.

[6] Schumpeter (1994) described it in his *History of Economic Thought* as follows: 'The first three chapters of Book I deal with Division of Labor. We are in the oldest part of the building, the part already completed in the Draft. […] Though, as we know, there is nothing original about it, one feature must be mentioned that has not received the attention it deserves: nobody, either before or after A. Smith, ever thought of putting such a burden upon division of labor. With A. Smith it is practically the only factor in economic progress. Alone it accounts 'for the superior affluence, […] technological progress 'invention of all those machines' – and even investments – is induced by it and is, in fact, just an incident of it. […] Division of labor itself attributed to an inborn propensity to truck and its development to the gradual expansion of markets – the extent of the market at any point of time determining how far it can go (ch.3) It thus appears and grows as an entirely impersonal force, and since it is the great motor of progress, this progress too is depersonalized. (182)'

to differentiate the two). On a more ironic note, their works contributed to the division of labor in between economics and sociology as separated disciplines. Both thinkers, as well as Becker and Murphy (1992) applied their ideas to the scientific production context. As opposed to the above introduced consideration of the phenomena in today's science studies, those applications concerned above all the scientific production process as a whole and the division of it into the basic institutions of science. In this context, Smith (1804) wrote that:

> 'In the progress of society, philosophy or speculation becomes, like every other employment, the principal or sole trade and occupation of a particular class of citizens. Like every other employment too, it is subdivided into a great number of different branches, each of which affords occupation to a peculiar tribe or class of philosophers; and this subdivision of employment in philosophy, as well as in every other business, improves dexterity, and saves time. Each individual becomes more expert in his own peculiar branch, more work is done upon the whole, and the quantity of science is considerably increased by it. (17)'

Durkheim (1966) equally considered this institutional perspective yet from a more pessimistic point of view:

> 'Another illustration of the same phenomenon has often been observed in the history of sciences. Until very recent times, science, not being very divided, could be cultivated almost entirely by one and the same person. Thus, one had a very lively sense of its unity. The particular truths which composed it were neither so numerous nor so heterogeneous that one could not easily see the tie which bound them in one and the same system. Methods, being themselves very general, were little different from one another and one could perceive the common trunk from which they imperceptibly diverged. But, as specialization is introduced into scientific work, each scholar becomes more and more enclosed, not only in a particular science, but in a special order of problems. (356)'

Finally, Becker and Murphy (1992) claim that specialization is constitutive of knowledge growth and illustrate the latter by the differentiation of journals:

> 'The first three economic journals started in the United States were general purpose journals – the Quarterly Journal of Economics in 1886, the Journal of Political Economy in 1892, and the American Economic Review in 1911 – whereas most of the many journals established in recent years are highly specialized: the Journal of Applied Econometrics, the Journal of Legal Studies, and the Journal of Economic Demography are a few examples. [...] The [...] examples illustrate that much of the growth in specialization over time has been due to an extraordinary growth in knowledge. (1145)'

Clearly, as opposed to the perspective of today's science studies,[7] Smith and Durkheim see the main effects of division of labor and specialization less in collaboration within micro-entities like e.g., a project or a laboratory (, which of course in this form did not even exist back then), but much rather in the institutionalization of different 'classes' of philosophers and the common 'trunk' of science. Their point of departure is science (or philosophy) as a whole and how it is institutionally divided by occupational differentiation (, not to say 'professions'). Considering the institution of the university, which was originally founded based on the idea of supporting all faculties (in the European sense of the term), constituting a single scientific production process,[8] the latter remarks suggest that division of labor and specialization in science should (at least also) be thought from an institutional perspective or in terms of the differentiation of professions within science (, as opposed to assuming the existence of pre-existing knowledge domains and how they might or might not be connected to one another).

[7] Notable exception here is the father of the sociology of science, Robert K. Merton, who considered this perspective to be relevant. In his analysis of behavior patterns of scientists, he claimed that the increase in: 'number of scientists has been accompanied by more and more specialization of research along the lines of both Spencerian and Durkheimian theories of role differentiation (330)'.

[8] Eaton and Stevens (2000) argue for this institutional perspective too, when they claim that universities have complex production functions supporting the division of labor in between cosmologists, economists and neuroscientists for example.

2.2.2 Demarcating institutional DoL and Spec. from the topics reviewed in 2.1.2

Based on the theories of Adam Smith, Emile Durkheim and Becker and Murphy, division of labor and specialization can be broken down into components (Ervin 1987). Initially, division of labor was understood solely as task division. The latter refers to the division of the whole production process into coherent parts. In Smith's pin factory for example, this meant that 'one draws out the wire, another straights it, a third cuts it, a fourth points it (Smith 1804: 13)' and so on. In most cases the institutionalization will provoke some sort of organizational separation of the tasks from one another, represented for example in different workstations. But this is not necessarily the case. Even in the straightforward industrial context of the pin factory, the tasks denote the differentiation of functions, which taken together constitute the production process. It is the function of the 'cutting' and 'straightening' that is institutionalized (in practice, carried out by different workers and, or at separated workstations), not e.g., a workstation of 'cutting' or department of 'straightening'.

Given this functional perspective, task division is not limited to industrial contexts either. Much rather, it is a continuous phenomenon, occurring equally on the aggregated level of societal production processes e.g., in occupational differentiation (e.g., priest vs. farmer), in context of the firm (operational vs. strategic department) or in micro-level productional contexts like the household (e.g., care work vs. provision of income). Common trait is that the institutionalization of a task requires a certain conceptual severability of tasks from one another, a common agreement on the scope of the task and how to perform it.[9] (Bücher 1922: 326-334; Durkheim 1966: 111-131; Gibbs and Poston 1975)

Task coordination in turn means the (re-)integration of divided tasks into a coherent production process. The effort linked to the latter are in economic terms described by coordination costs, which are posi-

[9] Bücher (1922: 309p) categorized the different forms of division of labor and (unsuccesfully) tried to establish the usage of different terms like 'Produktionsteilung' ('division of production'), 'Arbeitszerlegung' ('partitioning of labor') or 'Berufsspaltung' (division of occupation) according to the specific context in which division of labor occurs.

tively correlated to the degree of task division. The idea of coordination costs is based on works on principal-agent conflicts, free-riding and difficulties in exchange of information and communication, indicating that high degrees of task division can set incentives e.g., to shirk or hold-up other members involved in the production process. As opposed to the initial assumption of Smith that task division is exclusively limited by the extent of the market, Becker and Murphy (1992) claimed that coordination costs will effectively always limit the extent of differentiation since e.g., decision-makers in charge of institutionalizing separated tasks, will only promote task division to the extent, where the coordination costs level out the productivity gains induced by specialization on the task. Whenever this assumption is violated, we are confronted with abnormal forms of division of labor. (Becker and Murphy 1992)

The above-mentioned productivity gains through specialization can also be separated in two differentiable aspects. Productivity gains may arise through the concentration on a task. This concentration can either refer to the choice of an individual or an institution to specialize in a particular task. For the individual, this choice may best be represented by the choice of an occupation or field of study, whereas an institution e.g., a company might concentrate the resources over which it deposes on producing goods for a particular market segment. It is assumed and supposed that individuals and institutes will concentrate on tasks in which they have comparative advantages in comparison with peers, e.g., because of intrinsic interest, talent, or better basic prerequisites. Specialization concentration is thus a natural mechanism, by which individuals and institutions allocate themselves to tasks where they have comparative advantages, which results in productivity gains as opposed to a state where individuals and institutions do not concentrate on tasks (or cannot concentrate, because tasks are not divided). (Becker and Murphy 1992; Bücher 1922: 319-320; Smith 1804: 18-21)

Finally, specialization will also enhance productivity through what could be described as the gravitational force of specialization. The focus on a narrowed task (, as opposed to the whole production process) frees resources e.g., time to improve skills, develop a deeper understanding of the task and come up with more productive processes and technologies to perform it. The latter is believed to induce additional productivity gains and in the Smithian rationale technological progress in general (Schumpeter 1994: 182). Also, by isolating individuals or institutions within the scope of a narrowed task, they are forced to project the freed resources along the vertical dimension of a production process. Therefore, a mechanism is instated where concentration on a task provokes specializing in depth, which in turn sets an incentive for further task division. Since the latter of course will enable further specialization concentration, I will refer to this phenomenon as specialization gravity, that is the gravitational force of specialization pulling individuals and institutions towards narrower and narrower tasks. (Becker and Murphy 1992; Bücher 1922: 313-314, 322-323; Smith 1804: 18-21)

In conjunction with the remarks in the previous sections on science theory, we can derive the equivalents of task division and coordination, as well as specialization concentration and gravity for the case of the scientific production process. Of course, task division occurs on different levels of aggregation in science. Task division occurs on the micro-level of the lab or team (e.g., functional differentiation in author contributorship) and the macro-level of scientific communication (e.g., functional differentiation of journals or academic societies), which were previously introduced and are sufficiently considered in the science studies. Incorporating a historical perspective, institutionalized differentiation of tasks though mainly occurred on the intra-institutional (meso-)level of the university. The university professor represents the institutionalization of functional task differentiation by demarcating domains of research expertise and teaching responsibilities. The latter differentiation of professions or professorships (and their ongoing subdivisions, likely reflected by the denomination of professorial chairs and academic departments) is the stable driver of institutional division of labor in the scientific production process as described for example by Smith, when he thinks of classes of philosophers and Durkheim's perspective on the scholar enclosed in his special order of problems (see section 2.2.1). It existed long before science was conducted in large teams or its (every day) communication became globalized and up until today determines the way curricula are shaped and areas of research covered.

In accordance with the remarks of Becker and Murphy (1992), this functional differentiation provokes rising coordination costs. Given that an area of research covered by a professor is institutionally divided by his successors, research issues located at the boundary of those subtasks do not disappear. Much rather the latter requires for coordination in form of collaboration, knowledge exchange and integration. One example for such integration efforts in science are academic societies and the organization of scientific conferences, which facilitate such boundary transactions (Cohen 2021). Scientific task coordination thus refers to all efforts dedicated towards the coordination of the functionally differentiated professions and the integration of their knowledge domains (either in research or teaching). Task coordination may occur on the macro-level of communication in form of interdisciplinary journals or collections of papers, as well as on the micro-level of the lab, where research requires the coordination of tasks like conducting experiments, writing or operational procurement. On the meso-level, institutionalization of projects, institutes, chairs or departments dedicated to boundary-spanning research, linking institutionally separated research areas to one another, document efforts for task coordination.

Specialization concentration in turn denotes the allocation of scientists across functionally differentiated areas of research. On the macro-level this means the numbers of publications according to journals, as well as the number of journals according to subject area, research field and discipline. On the micro-level, this concerns the relationship of number of technicians vs. doctors in a lab. On the meso-level, concentration means the allocation of academic staff across different fields of research. On all levels, this allocation process is influenced by scientists' interests, their prior choice in field of study or the focal point of their PhD. Further, this as well concerns the concentration of academic staff resources according to academic departments e.g., when polytechnic universities focus on areas of research within engineering fields for example.

Finally, specialization gravity denotes the mechanism resulting from concentration on research domains provoking the demarcation of increasingly narrower scopes of research in newly institutionalized tasks. On the macro-level, this is reflected by the increasingly narrower domains of journals as documented by Becker and Murphy (1992) (see section 2.2.1), whereas the shift towards team-based science is itself an example for specialization on narrowed (, in this case technical) tasks. On the meso-level, the latter is reflected by the increasing number of departments and chairs established within disciplines, dedicated to narrower and narrower subject areas over the course of time. Finally, an overview of the latter application of the components of DoL and Spec. on the scientific production process is provided in table 1.

Tab. 1: Theory of division of labor and specialization applied to scientific production

	Division of Labor		**Specialization**	
	Task Division	Task Coordination	Concentration	Gravity
Regular (economic, societal)	Institution-alization of functional differentia-tion of tasks	Coordination of institutionally differentiated tasks through commu-nication, collab-oration and integration of information	Specialization in a task according to interest, talent and opportunity	Deepened understanding, improved skill, and develop-ment of tech-nologies by focus on narrowed task
Scientific	Research domains (and technical tasks) institu-tionalized	Coordination of research domains through boundary-spanning	Concentration (by choice of field of study, PhD program) of academic staff according to research domain	Pushing the knowledge frontier by focus on narrowed research
Macro	*in e.g., academic societies, journals*			
Meso	*In e.g., departments, professorial chairs*			
Micro	*within e.g., teams, laboratories*			

Assigning the topics covered in the science studies to this scheme, it can be concluded that a thorough consideration of DoL and Spec. on the meso-level of the institution as determinant of epistemic outcomes is missing thus far. The author contributorship literature for example concerns task division within teams on the micro-level, whereas the literature on diversity and interdisciplinarity is concerned with task coordination (and to some extent concentration of resources) on the macro-level of scientific communication. The debate on cognitive division of labor in turn concerns the specialization concentration holistically from an abstract and exclusively theoretical point of view. To summarize, the existing studies are concerned either with the aggregated macro-level of scientific communication, specific micro-productional contexts like interaction within a team or a researcher's choice of method or research issue. In conjunction with the statements of Smith, Durkheim, as well as Becker and Murphy, we conclude that a consideration of division of labor and specialization on the level of the institution is missing, and that the latter poses a research gap in the science studies.

2.2.3 Potential of institutional DoL and Spec. to explain pathologies in science

First argument in favor of closing this research gap and accounting for institutional DoL and Spec. as determinant of epistemic outcomes, is its proximity to the topics, theories and methodologies employed in the science studies (Kaiser 2023). Indeed, it is odd that the terminology of division of labor plays an important role in the debate on cognitive division of labor, suggesting that there is at least a tacit knowledge on its importance, yet the institutional perspective by thinkers like Smith or Durkheim is not accounted for (Kaiser 2023). Another peculiar circumstance is that the whole science studies (, not only the branch of economics of science) are to an astonishing extent permeated by economic thought. In context of the sociology of science, Jansen (1995) highlighted the underlying assumption of the autonomy and self-government of science, which is expressed by the employment of functional or quasi-economic models of science. Now while it may be argued that Kitcher's rational choice-based approach is as much typically philosophical as it is economic, it was already lined out in the section above, that authors like Polanyi et al. (2000) and Zollman (2018) agree on the importance of economic rationality and theory in science. De Langhe and Greiff (2009) as well as Mäki (2005) called these approaches 'invisible hand models (290)', directly referring to Smith (yet not to DoL and Spec). Goldman and Shaked (1991) even explicitly argued in favor of extending 'the economics paradigm to certain problems in epistemology and the philosophy of science (31)'. (Kaiser 2023)

Also, studies in scientometrics frequently involve some kind of performance measurement e.g., in university rankings, supporting an efficiency paradigm employing econometric methods. Clearly, institutional division of labor and specialization would fit in there quite well as a further determinant of epistemic outcomes. Judging by the different topical focal points in philosophy, sociology and economics of science one could make a case that the neglection of institutional DoL and Spec. is caused by institutional DoL and Spec. itself. Even though economic thought is dominating the discussions, knowledge integration does apparently not go far enough for philosophers of science to acknowledge that what they are dealing with in Kitcher's model is not division of labor in an original Smithian sense. Further, insights from scientometrics like e.g. skewed distribution of citations, and sociology of science e.g., Merton's *Matthew effect*, are on closer inspection good examples for 'market failures' not accounted for by Kitcher's rationality-based model. Yet they play no role in the cognitive division of labor literature. It can thus be concluded that institutional DoL and Spec. deserve a closer look, simply because of their topical and theoretical proximity to the science studies' paradigms. (Kaiser 2023; Weingart 1974)

Second argument in favor of adopting an institutional perspective concerns the nature of knowledge coordination. The latter is not limited to knowledge integration within or in between publications (on the output side), but also concerns collaboration of researchers within or in between institutions (on the input-side). Certainly, communication networks in science, where authors cite documents from distinct disciplines or fields and form 'invisible colleges' are one legitimate, acknowledged and valuable form of research coordination (Zitt et al. 2019) When it comes to knowledge coordination from (allegedly) distant research domains, institutional coordination of researchers from different fields might be required (e.g., within a joint department, laboratory or project) rather than just recombining content from different disciplinary publications. (Crow and Dabars 2017; Moschini et al. 2020; Walsh and Lee 2015)

The literature on interdisciplinarity might thus as well benefit from the consideration of meso-level institutional division of labor. Calls for the promotion of interdisciplinarity may fall short whenever institutional knowledge coordination is not sufficiently accounted for and the lack of research coordination is determined on grounds of e.g., low bibliometric diversity. Interdisciplinary departments need not necessarily publish only on boundary-spanning topics or in interdisciplinary journals but may on some occasions produce highly disciplinary publication output. On the flip side, in extreme cases, publications classified as interdisciplinary (e.g., according to categorization schemes of publication databases) could possibly stem from researchers affiliated with highly disciplinary departments working on highly disciplinary research issues loosely connected by a paper's interdisciplinary topic (Bammer 2017).

Apart from complementing the bibliometric perspective, thinking about scientific collaboration in terms of meso-level division of labor and specialization might also be much more concrete than thinking about it in terms of disciplinarity. It is easier to implement and operationalize than interdisciplinarity with its vague concepts, which are known to be quite fuzzy and its terminology (inter-, multi-, trans-, pluridisciplinarity) often used inconsistently (Jacobs 2017). In addition, the definition of disciplines is mostly arbitrary (bibliometric based) and not granular enough to get a grasp on the development of task differentiation and coordination as well as specialization depth and how the latter have evolved over the course of time. Schumpeter (1994: 20) noted in this context that division of labor has created large numbers of specialties and applied fields, which appear and disappear, vary in importance and have overlapping boundaries. Such changes, as well as the dynamics over the course of time, may best be evaluated employing an institutional perspective (Bechtel 1993; Walsh et al. 2019). Given the arguments outlined above, it might thus be reasonable to complement the rich literature focusing on collaboration and coordination on the output side with an analysis of coordination on the input side of the scientific production process.

Finally, apart from the aggregated level of the discipline, the existing literature is concentrated on the micro-level analysis of coordination within units like teams or laboratories. Since DoL and Spec. were, in the previous section, defined as continuous phenomena, they are simultaneously applicable to coordination at different levels of granularity. Accounting for the meso-level of research institutions by integrating research integration and coordination within universities or the academic department allows for a synthesis by integrating the topics of interdisciplinarity and diversity with the division of labor in teams, building a bridge between micro and macro-level, as well as input and output perspective.

Third and final argument in favor of accounting for the institutional perspective is that the theory on DoL and Spec. provided in the previous section bears the potential for explaining pathologies in science e.g., the declining progress of research in some fields addressed in the introductory sections (Morris 2020; Walsh et al. 2019). From a theoretical point of view, we know that division of labor and specialization are important determinants of productivity of any production process. Further we know that those productivity gains are counterbalanced by increasing efforts necessary for the coordination of institutionally differentiated tasks. Again, according to Becker and Murphy (1992) the decision-makers will only permit labor to be divided to the extent where its productivity gains are leveled out by the coordination costs. In a way, this is compatible with the idea promoted in the science studies of an efficient cognitive division of labor. It may be reasonably questioned though, whether in all scientific institutions, e.g., universities, such a decision-maker, which actively monitors and controls for coordination costs, really exists. Indeed, in particular for the European, *Humboldtian* university model, which grew organically over the course of the last centuries, it might be questionable to assume such a centralized decision-making entity. Finally, given the insights from sociology of science, which describe all sorts of 'market failures' applicable to the rationality-based models promoted to explain the scientific production process, assuming an optimal extent of task division by default, without actual empirical validation, is simply not satisfying.

Another issue is that the causes of specialization concentration suggested by economic theory are incompatible with the assumptions of the model on cognitive division of labor. If individuals indeed tend to specialize in tasks, in which they are particularly gifted or interested in, this should equally apply to the choice of research issues and methodologies. One could even make the case that this form of specialization (concentration) sets in long before a scientist is even confronted with the choice situation described in Kitcher's model. Supposing that information on expected credit when pursuing a particular research question or method is available, one might nonetheless decide against the one that maximizes credit, because it simply not suits his interests, or he lacks the necessary skills to employ the methodology (Merton and Zuckerman 1973a). One can easily imagine a junior faculty member, who e.g., specialized as a postgraduate student in qualitative methods, now confronted with the choice of pursuing a research question using either a quantitative or a qualitative approach. Even though in his area of research, qualitative methods might be the most frequently used and in this particular instance more credit could be gained from choosing the quantitative approach, he might nonetheless stick with the qualitative one. Even without validating the latter empirically, this line of thought has a certain intuitive appeal. Such path dependencies are completely ignored in the literature on cognitive division of labor. Equally, scientists' choices are also bound to the scope of research of the academic department or professorial chair they are affiliated with, their equipment or even their location when research projects require access to local facilities or natural sites (Knorr-Cetina 1984; Latour and Woolgar 1986). The mobility in between those institutional units is also known to be limited and depends on factors like endowment with resources, which are completely independent of expected credit. (Ben-David and Collins 1974)

Also, the theoretical insights on the gravitational force of specialization might provide an interesting perspective on the issue of the increasing knowledge burden. Simply put, one may ask if it is really just the accumulation of knowledge that forces researchers to focus on narrower research domains to ensure innovative research, or if the rationale is not actually the other way around and it is the gravitational force of specialization (, which adds towards) increasing the knowledge burden. Research output need not necessarily be specialized in depth to be disruptive, but could also be innovative due to horizontal knowledge integration from different research domains.[10] Taking into

[10] Taking into account the literature on groundbreaking discoveries and formation of new research fields in the history of science, combinatorial novelty is a successful strategy for creating disruptive, innovative research (Bechtel 1993). On the contrary, Schumpeter (1994) pointed out that 'cross-fertilization might easily result in cross-sterilization (24)' whenever the latter is pursued at the cost of narrow specialization where needed. Arguably though, in today's science there is no shortage of institutions and (academic staff) resources and such a trade-off need not necessarily arise.

account that the gravitational force of specialization promotes the former and increases the coordination costs linked to the latter (through increased institutionally differentiated tasks), one could argue that the knowledge burden (is not a destiny, because all low-hanging fruits are already yielded, but) to some extent poses a manageable problem within our sphere of influence.

Even though this is not the focus in this work, it might be worthwhile to turn to teaching activities in scientific institutions for a moment. Assuming that at least partly (over the course of the last decades), scientists are working in more differentiated tasks which are more specialized in depth, this should to some extent be reflected by the content of courses, curricula and programs also becoming increasingly fragmented and specialized in depth. Here too the question arises if knowledge coordination in between the courses is controlled for by scientific institutions or if this relies entirely on students' shoulders. If the latter is the case, then the 'knowledge burden' on students' shoulders would increase over time and this is not entirely due to accumulation of the knowledge stock but also due to the institutional setting.

Clearly also, this can be reconnected to the issue of interdisciplinarity. Interdisciplinary programs have become highly popular over the last years, yet their value is still regularly doubted (Häussler and Sauermann 2020). Applying the institutional perspective of DoL and Spec. here as well, we might want to question if interdisciplinary programs are coordinated in the same way as publications are (from the output side) or if they are constructed inherently interdisciplinary by academic staff integrating their knowledge domains (from the input side) to provide a coherent set of courses constituting a kind of coherent 'teaching production process'. Supposing the latter is not always the case, the knowledge burden on students' shoulders should in interdisciplinary programs be even higher than in disciplinary ones.

All-things-considered, institutional (meso-level) DoL and Spec. should be examined as potential determinants of epistemic outcomes, because they are well-suited for explaining pathologies in science, bear the potential to complement and connect existing (bibliometric) perspectives on diversity and interdisciplinarity, as well as their proximity to the economic thought permeating the science studies.

2.3 Research Question and Outline of the Empirical Analysis

To summarize, the introductory section served to motivate the accumulating empirical evidence pointing at a declining quality of epistemic outcomes, represented for example in decreasing disruptiveness and novelty of publications, likely caused by increasing coordination costs ('increasing knowledge burden'). Based on a literature review on the science studies, acknowledged determinants of epistemic outcomes were introduced, which are known to affect the scientific production process (in general) and might be able to explain the observed pathologies. The latter revealed that an economic rationale and here in particular rationality-based theories dominate the discussion, which characterize science as a self-governed sphere, constituted by scientists' motive to receive reward and gain credit. In the cognitive division of labor debate for example, it is presupposed that scientists anticipate the expected reward of pursuing a research issue or employing a method and since they seek to maximize the latter, an efficient allocation and coordination of scientists and research is secured. Since the empirically observed pathologies of the scientific production process conflict with the rational theory-based paradigm in the science studies, the need for considering alternative explanations was outlined. Since empirical evidence points at rising costs in coordinating specialized research and given the economic rationale permeating the science studies, it was natural to consider the economic theory on division of labor and specialization to analyze pathologies of the scientific production process. Even though the latter bears interesting implications for understanding allocation and coordination of scientists and research and the terminology (as in cognitive division of labor) as well as related topics such as diversity and interdisciplinarity are to some extent part of the science studies' literature, a thorough discussion of division of labor and specialization in its original sense as popularized by thinkers like Adam Smith, Emile Durkheim, Karl Bücher or Becker and Murphy is missing thus far. This particularly concerns division of labor and specialization on the meso-level of the institution, which in comparison to considerations of specialization on the aggregate macro-level of scientific communication and task allocation within teams on the micro-level, has thus far not been accounted for at all.

This work seeks to close the above motivated research gap by accounting for institutional DoL and Spec. and pursuing the following research question:

> **RQ:** Are institutional DoL and Spec. determinants of epistemic outcomes which can explain pathologies in the scientific production process?

To answer this research question, an empirical validation of the theoretical arguments provided in favor of considering DoL and Spec. as important determinants of epistemic outcomes, is needed. Thus, the first research objective in this work is to measure DoL and Spec. (input-based) on the meso-level of the institution. To achieve this, a new data set, which allows to operationalize the four components of DoL and Spec., based on institutional data collected for 20 excellent universities over the period 1890 to 1920, is introduced (see ch. 3). In case institutional DoL and Spec. indeed influence the scientific production process, the theoretically introduced mechanisms linked to the interaction of task division, specialization concentration and gravity need to be confirmed. This will at first be evaluated by a thorough descriptive analysis, exploring the rich insights provided by the new data set over the course of time and on different levels of granularity within individual institutions. To affirm the latter quantitatively, a hierarchical cluster analysis is employed to segregate different university types and a correlation analysis is used to assess whether significant path dependencies are induced within the latter by initial configurations of DoL and Spec.

The second research objective builds on the former and seeks to evaluate the rationality-based idea of an efficient allocation of cognitive labor as suggested in the science studies. Therefore, in chapter 4, a conditional nonparametric efficiency framework is employed (see ch. 4), which allows to measure efficiency and model the relationship of the latter with the four components of DoL and Spec. In case that sufficient evidence is found that DoL or Spec. create path-dependencies within institutions, and it can further be shown that the latter influence efficiency, which can be functionally modelled in accordance with theoretical expectations, it follows that DoL and Spec. are determinants of epistemic outcomes. Further, it can be concluded that since the degree of DoL and Spec. is not necessarily efficient, the rationality-based paradigm in the science studies focused on the motives of reward and credit need to be augmented with a more solid perspective accounting for institutional prerequisites.

3. Introducing a New Dataset to Measure Institutional DoL and Spec.

(Dataset Construction and
Descriptive Empirical Analysis)

3.1 Empirical operationalization of the institutional DoL and Spec. concept

3.1.1 The need to limit the empirical analysis to engineering and natural sciences

Given the aim of this work is to establish division of labor and specialization as determinants of epistemic outcomes by empirically analyzing their impact on efficiency, a few conceptual notes are in order. An empirical analysis of the influence of DoL and Spec. on efficiency requires a quantitative perspective on epistemic outcomes, meaning that here a logically necessary connection between quantity of outputs of the scientific production process (e.g., publications) and epistemic outcomes in general is assumed.

From a layman's perspective, epistemic outcomes might intuitively be linked to some sort of notion of scientific progress. Epistemic outcomes could indeed be understood as the products of scientific inquiries, expected to be transmittable to advancements in society, health or technology (Jansen 1995). Such naive accounts of scientific progress often build on the idea that (as a result of the scientific production process) true theories are proposed and accepted. A more sophisticated version of the latter can be rediscovered in Popper's truthlikeness account, where scientific progress is defined as the substitution of propositions with propositions that are more truthlike (Popper 2002). Although this is still an influential account of scientific progress today, alternative or complementary accounts are available. The noetic account for example suggests that scientific progress occurs whenever our ability to predict or understand a particular phenomenon is improved. Another account was proposed by Bird (2007), who argues that scientific progress is the accumulation of new knowledge to what is already known, with the notion of knowledge being among others defined by internal and external epistemic justification like e.g., reliability and sensitivity of results. (Dellsen et al. 2015)

The most influential alternative to the account of Popper is the anti-realist account proposed by Thomas Kuhn. According to the latter, scientific problems can be divided into empirical and conceptual problems, which are defined by the dominant research tradition at a certain time within a given discipline. Scientific progress is fulfilled when number and importance of problems within a research tradition decreases (Kuhn 1976). Here, the truthlikeness of those problems is of no concern. Indeed, even if it later turned out that due to a wrong theory, a discipline worked on a spurious problem, they (believed to have) solved, this would still constitute scientific progress according to the latter account. (Dellsen et al. 2015)

While it is unclear whether a necessary connection of scientific progress (as a quality of epistemic outcomes) and quantity of scientific output exists in case of the noetic and truthlikeness account,[11] a connection between the latter and the accounts of Bird and Kuhn is clearly given. Even if not every single paper's content must contribute towards knowledge accumulation, higher quantities of publications are expected to contribute towards epistemic justification like reliability of results e.g., when reproductions of experiments are published. In the problem-solving account, the connection is even clearer. Assuming that on average, a publication within a part of the scientific community does not contribute more new problems to a field than it contributes towards solving a field's problems, scientific progress according to this account is logically connected to the quantities of scientific output like for example the number of publications.

Since all four accounts are acknowledged explanations for scientific progress and for two of them, a necessary connection between quantity of scientific outputs and progress exists, it is uncritical to

[11] Even though there might be no necessary condition of quantity of scientific outputs and progress in those two accounts, a higher number of e.g., publications should in general contribute towards substitution of theories with more truthlike ones or better understanding and modeling of phenomena of interest respectively.

assume that the impact of DoL and Spec. on (quantitatively assessed) efficiency in producing scientific outputs is sufficient for establishing an influence of the latter on epistemic outcomes in general.

A distinct issue though is if there might not be a different goodness of fit according to epistemic branches of science. Arguably, a reduction in importance and numbers of problems is well-proxied by an accumulation of publications in e.g., natural sciences, engineering as well as medical and life sciences.[12] The low rates of rejection by journals in these disciplines (see Merton and Zuckerman 1973) support the idea that in the latter disciplines increased quantities are more valuable than in other epistemic branches. On the contrary, in the humanities one could object to the above made assumption that a publication on average contributes more towards solving a field's problems. For one, it is unclear whether in the different subbranches of humanities a finite field of problems indeed exists (or if the latter are rather infinite). Second, it is not trivial to come up with problems exclusively dealt with in the humanities, which can be conclusively solved at all. In accordance with the line of thought in this work, one could very well argue that a result of the institutionalization of division of labor over the course of modern science is that solvable and unsolvable problems originally jointly treated within a philosophical faculty (European) were segregated first into a mathematical-natural and a philosophical section of the same faculty and today are completely institutionally isolated in separated faculties and departments. Consequently, the idea of contributing continuously (through Kuhnian-like puzzle-solving activities) towards a better understanding of phenomena or reduction of limited problems might simply not be applicable at all to the humanities, since all the problems which could be tackled this way, were outsourced to other epistemic branches.

A different case could be made for the numbers of publications produced in formal sciences like e.g., mathematics, statistics and

[12] There are also more pragmatic considerations regarding the availability of consistent data for and comparability of scientific outputs. Publication databases like e.g., Scopus and Web of Science have only just recently made an effort to integrate more books and book chapters, which are an important publication type next to the traditional peer-reviewed journal article in the social sciences and humanities. (Bonaccorsi et al. 2017; López-Illescas 2011; Van Raan 2019)

logics. Instead of contributing towards problem-solving within their discipline, their institutionalized task is above all realized in collaborations with other empirical sciences (Rousseau et al. 2019), and less with boundary transactions (Cohen 2021). Finally, among the empirical sciences, the social sciences might take the position of a hybrid in between humanities and other empirical sciences when it comes to definition of a set of problems and extent of solvability. While e.g., in economics, puzzle-solving activities exist like e.g., economic forecasting, which may be better understood and improved with increasing number of publications on the topic, all social sciences are confronted with renewing sets of problems and a quite limited longevity of paradigms since the social environment is continuously evolving (Stephan 2012).[13] It might thus be impossible to confidently state that any absolute number of publications produced in the social sciences are 'enough' to reduce importance and number of problems at hand. We thus follow the line of thought in e.g., Stephan (2012) and Bornmann and Mutz (2014), which equally argued against a comparability of social sciences with other empirical sciences.

Consequently, when comparing scientific institutions' performance based on quantitative inputs and outputs it is deemed a better fit to limit the scope of the latter to the empirical sciences (, excluding social sciences) instead of limiting the latter to one of the other sciences or examining performance based on all sciences, whose production processes are hardly comparable. It is believed that this limitation produces the least conceptual and methodological trouble. The dataset introduced in this chapter thus exclusively considers disciplines in the natural sciences, medical and life sciences as well as engineering. Humanities, formal and social sciences are not considered.

[13] Or as Schumpeter described it in the History of Economic Analysis (1994): 'Scientific analysis is not simply a logically consistent process that starts with some primitive notions and then adds to the stock in a straight-line fashion. It is not simply progressive discovery of an objective reality – as is, for example, discovery in the basin of the Congo. Rather it is an incessant struggle with creations of our own and our predecessors' minds and it 'progresses' if at all, in a criss-cross fashion, not as logic, but as the impact of new ideas or observations or needs, and also as the bents and temperaments of new men, dictate (3).'

3.1.2 Available perspectives and data sets

3.1.2.1 Aggregate institutional level (input-based)

Before the construction of the data set is explained in the upcoming sections, a summary of available perspectives and data sets is presented to motivate the need for a new data set. While it was established in the previous chapter that division of labor and specialization are neglected determinants of epistemic outcomes in the science studies, more or less closely related topics exist (and were also addressed), which might use datasets that are suited for pursuing the research issue of this work. The brief literature review provided in this section is thus supposed to convincingly show that data and analyzes thus far performed on an aggregated institutional level are not granular enough to allow for an operationalization of the components of institutional division of labor and specialization conceptually defined in the previous chapter, nor do they deal with the two phenomena as understood in this work. In the upcoming section, this will be repeated for the case of datasets and analyzes based on the 'bibliometric hypothesis' with special emphasis on why it is not adequate to examine DoL and Spec. from an output-based perspective.

As already addressed, division of labor within teams has been analyzed before. Walsh and Lee (2015) for example employed a survey-based approach to examine the interdependence of tasks in teams and analyzed to what extent team members had one person jobs like e.g. specialized technicians in laboratories. They find significant field differences with higher task interdependency in physics than e.g., biology with a higher probability for internal division of labor with increasing project size.

A quantitative, institutional-data based approach was for example employed by Lepori et al. (2019), which analyzed subject composition of universities, calculating a Herfindahl-index for the distribution of students by field, employing the field of education and training classification. The analysis was based on micro-data from the ETER project, which contains information on a set of 3,474 European HEIs over the period 2011 to 2020. The latter contains numbers of academic, instructional, research and public service staff, information on funding as well as share of students in education, humanities and arts and social sciences and is the only data set that contains information on the micro level for European HEI (Bruni et al. 2021). Catalano et al. (2017)

also used the ETER data to study heterogeneity in between different countries' higher education systems by analyzing their subject mix based on student numbers.

Daraio et al. (2015a) related the subject mix to universities' efficiency, adopting a cross-country perspective for modeling production trade-offs between the Humboldt (teaching and research) and sole research model using the EUMIDA dataset based on the Aquameth project (Bonaccorsi and Daraio 2007). The latter contains micro data for individual universities of 11 European countries (based on census information) for the period 1994 to 2005 (Daraio et al. 2011). According to the authors the analysis based on the latter provides a better generalizability of results when compared with the rich literature of other country-level studies either performed based on a small set of countries or observations according to the authors. They find a significant impact of scale and specialization (in terms of subject mix) on the Humboldt model yet not on the research model. Another country-level study employing micro-level data (from the ETER project) is provided by Bruni et al. (2021), which combine a cluster and efficiency analysis approach to examine the heterogeneity of national higher education systems. The results suggest that specialization either in research or teaching is more efficient than balancing both.

3.1.2.2 Bibliometric (output-based)

In this section, the studies are presented, which treat topics related to division of labor (in teams) and specialization based on the *bibliometric hypothesis*. As in the review on analyses incorporating institutional data, here as well accounting for specialization through the subject mix plays an important role. Pastor and Serrano (2016) for example, measured efficiency of universities on country-level for the period 2008 to 2012 and considered the disciplinary composition of institutions derived from information of the Scimago Journal and Country Rank database. The authors calculated Data Envelopment Analysis (DEA) based efficiency scores for a publication and a quality (citation) model for different disciplines e.g., humanities, medical and life sciences, natural sciences etc., to construct an aggregated efficiency measure. The study suggests that when disciplinary composition and quality of research is accounted for, the mean inefficiency obtained declines significantly. In context of performance measurement, López-Illescas et al. (2011) criticized university rankings for combining research activities from different disciplines and aggregate them to the whole university. They examine differences in disciplinary specialization for multiple bibliometric indicators for the case of 50 Spanish universities from 2003 to 2008 by calculating them according to the 27 classifications for publications in the Scopus database. The findings suggest that a university's overall score in those rankings can be dominated by the results of one or two disciplines only.

Liu et al. (2021) in turn analyzed if interdisciplinary collaboration research is more disruptive than monodisciplinary research. In their framework, interdisciplinary collaboration is measured by coauthorship of authors stemming from at least two disciplines, which was categorized based on the department and school affiliations provided in the analyzed papers. Leahey et al. (2017) targeted a similar research question, analyzing interdisciplinarity based on information retrieved from 32,000 articles of 900 research-center based scientists, whether spanning disciplines is credited or penalized by the scientific community, because it promises more disruptive results yet might also be confusing to place (in disciplinary journals). They find evidence for both and suggest characterizing interdisciplinary research as high-risk, high-reward endeavor. Bongioanni et al. (2014) provide a literature review for analysis of disciplinary structure of country-level scientific production and investigate the dynamics of the latter for European countries over the period 1996 to 2011 by examining multiple bibliometric indicators according to disciplinary classification in 27 Scopus subject categories. They find a convergence toward the average European disciplinary profile in general and departure of some countries from the latter, which initially constituted it.

Leydesdorff et al. (2019) provide an overview of operationalization of interdisciplinarity in the bibliometrics literature using diversity measures such as the Simpson diversity index (in economics more commonly known as Herfindahl-index) or the Rao-Stirling-index. The authors modify the latter to better account for variety, balance and disparity (to get a grasp on different aspects of interdisciplinary collaboration) and apply their new index to the aggregated citation patterns of 11,487 journals contained in the Journal Citation Report 2016 of the

Science Citation Index and the Social Sciences Citation Index. The authors suggest that their interdisciplinarity index provides improved results by differentiating knowledge diffusion (cited impact) and knowledge integration (citing references).

Daraio et al. (2018) adopt an actor network theory perspective, where they define papers as nodes and citations as links in scientific communication networks. The nodes in turn represent the scientific production process of a particular research domain and the links, the interdependencies between them. Their findings suggest strong ties of collaborations in between institutions belonging to a given geopolitical and cultural area (e.g., US, UK, Canada and Australia) and different results for cross-disciplinary interactions of different countries.

Another topic in scientometrics and bibliometric analysis is division of labor within teams. Larivière et al. (2016) used contributorship data extracted from PLOS one articles to study to what extent authors share responsibility of contributions. Simply put, the authors studied task division on the level of production of the individual paper, e.g., which author performed data collection and analyses and did all authors contribute towards writing and so on. The results suggest that differences for the latter exist according to fields. In biomedical research and clinical medicine for example, scientists rarely (in 10% of all cases) contribute to the five tasks defined, whereas in mathematics and physics for example in about 25% of all publications, authors contributed to all tasks. The authors suggest that this is only partly explained by the theoretical nature of the latter two favoring lower collaboration levels, since in physics a high division of labor is required. They then argue that the Mertonian ideal of scientific communalism might be better realized in physics with its egalitarian nature favoring the practice of indicating equal author contributions when specifying contributorship. Häussler and Sauermann (2020) also perform an analysis incorporating contributorship statements within scientific teams and find a correlation between the latter and interdisciplinarity measured based on domains of expertise on level of individual authors derived from article field classifications. Further, Lu et al. (2019) investigated the topic by analyzing co-contributorship based on performing one-mode projection of author-task bipartite networks obtained from 138,787 articles published in PLOS journals. Their findings suggest different types of contributors (specialists, team-players and versatiles), which can be assigned to certain characteristics of researchers, like e.g., academic positions. Finally, Walsh et al. (2019) analyzed interdisciplinarity of teams and suggest a positive relationship of interdisciplinarity of teams with retractions of publications.

A third major topic in bibliometrics and scientometrics studies is field normalization. Given that in most cases institutional data on differing input levels according to fields is not available, the heterogeneity introduced by judging performance based only on bibliometric information (output) is accounted for by normalization of citation counts. (See Waltman and Van Eck (2019) for an introduction, where they address the issue of defining and demarcating a field) Other fringe topics loosely connected with DoL and Spec. examined based on the *bibliometric hypothesis* are the conditions and probabilities for topic switching (e.g., Zeng et al. 2019), analysis of combinatorial novelty and disruptiveness based on semantic analysis (e.g., Lin et al. 2022) as well as science mapping. The latter as e.g., pursued by Leydesdorff and Milojevic (2012) is built on co-occurrence count data employed to map subject terms, documents and journals and is believed to be useful for a better understanding of the dynamics of science and for better informed decisions on funding allocation.

3.1.3 Justifying the need for a new data set on intra-institutional level (input-based)

The studies based on aggregated macro-level institutional data and the *bibliometric hypothesis* presented in the previous two sections deal with (related) topics of division of labor and specialization in different ways. Analyses taking the institutional perspective cover topics such as the survey-based analysis of division of labor within scientific teams, or quantitative analysis of subject mix (, operationalized with diversity indices) and HEI's efficiency by employing microdata from recently enforced collection initiatives such as the ETER or Aquameth project. While data from the latter refers to the individual institution, the majority of analyses of subject mix, which is conceptually comparable to the here promoted specialization concentration component, are conducted on aggregated country-level and concern the last two to three decades (, for which microdata was collected).

Studies performed based on the *bibliometric hypothesis* exist in larger quantities, are more diverse and not limited to the latter time frame. Division of labor within teams is based on authorship contributorship data, whereas subject mix is, analogously to the institutional perspective, operationalized by a set of acknowledged diversity measures. Here though, the subject mix is not determined based on student numbers, but on topics of publications, journals, author affiliations or classification schemes of popular bibliometric databases like e.g., Clarivate Analytics' *Web of Science* or Elsevier's *Scopus*. While some studies on aggregated country-level do exist, bibliometric studies operate on average on a higher level of granularity by examining journals, publications and the links between publications through citations. Major topics that deviate from the institutional perspective are the analysis of interdisciplinarity, comparison of production according to field, field-normalization of citation count and (network) analysis of scientific collaboration through communication.

In table 2, an overview of the empirical literature addressing topics related to division of labor and specialization in context of performance measurement in science is provided. A broad range of studies based on bibliometric data exists on the macro-level of the scientific community, the institutional level and the level of the individual or team. The dominating topics most closely linked to division of labor and specialization are the analysis of the impact of subject mix, disciplinarity and author contributorship on performance or production. Turning to the institutional perspective, we can obtain that a variety of studies exist on the level of the scientific community yet applications on institutional and team-level are rather scarce.

The very limited availability of studies based on intra-institutional data sets corresponds to the neglection of institutional DoL and Spec. established in the previous chapter. Indeed, the issue of the dominating output-perspective, basing analyses mostly on bibliometric information is well documented in the scientometrics literature. Bornmann et al. (2023) just recently emphasized that in performance measurement of universities the availability of institutional data is too restricted. The authors propose to derive a suitable input indicator based on unique author affiliations in publications and perform a two-stage DEA approach for 3,100 universities worldwide. Arguably though, the derived input indicator can hardly be considered as truly institutional since it still relies on compilation and quality of bibliometric data and thus too implicitly builds on the bibliometric hypothesis.

Tab. 2: Literature review on available perspectives and data sets – summary table

	institutional data	**bibliometric data**
Macro-level (country-level, communication and collaboration networks)	Subject mix: • Catalano et al. 2017 • Daraio et al. 2015a Heterogeneity of HEI systems: • Bruni et al. 2020	Interdisciplinarity: • Daraio et al. 2018; • Leydesdorff et al. 2019 Subject Mix: • Pastor and Serrano 2016 • Bongioanni et al. 2014 (see for further literature) Science Mapping: • Leydesdorff and Milojevic (2012) Semantic Analysis: • Lin et al. 2022
Meso-level (universities, research organization)	Subject mix: • Lepori et al. (2019)	Interdisciplinarity: • Liu et al. 2021 • Leahey et al. 2017 • Walsh et al. 2019 Subject Mix: • López-Illescas 2011 Field-normalization: • Waltman and van Eck 2019 (see for further literature)
Micro-level (team / individual)	Survey on DoL: • Walsh and Lee (2015) • Lee et al. (2015)	Author contributorship DoL: • Larivière et al. 2016 • Häussler and Sauermann 2020 • Lu et al. 2019 Topic Switching: • Zeng et al. 2019

The need for a strengthened intra-institutional perspective complementing the studies based on bibliometric information is further also due to the limits of bibliometric approaches in general and when dealing with aspects of DoL and Spec., be it in form of subject mix or interdisciplinarity. Regarding the former, Lepori et al. (2019) argued that bibliometric indicators cannot be used as reliable measures for allocation of resources, because effectively resources are never considered and suggested 'scale-free' indicators like mean-normalized citation scores are shown to be in reality size-dependent. The authors suggest that bibliometric based studies might provide robust results for identifying low performance but are less reliable when discrimination of top performance is of interest. Regarding the latter, Zitt et al. (2019) pointed out that the rich conceptualizations of disciplinarity in sociology do not bear information on how to operationalize field delineation. Interdisciplinarity measures based on bibliometric data though are highly dependent on the approach employed to separate disciplines or fields from one another (Rousseau et al. 2019).

Results of interdisciplinary collaboration, building on classification schemes which define disciplinarity based on information from (publication and journal) output might indeed not be very informative since they convey no information on collaboration occurring on the (institutional) input side. Simply put, if e.g., a professorial chair or department is installed that combines two areas of research recently isolated from one another, this might signal some kind of interdisciplinary coordination of research. The produced results could nonetheless be strictly disciplinary. In this context, López-Illescas (2011) highlighted that when assessing performance of different research groups, results of bibliometric analysis need to be treated with care since there is no necessary connection of a discipline's name and a department name. On the contrary, given the latter area of research grows big enough, it might eventually receive its own category in a bibliometric classification scheme, being considered disciplinary regardless of whether coordination of research within institutions is really disciplinary. So, thinking about collaboration and coordination from the output side will necessarily hide the dynamical changes in institutionalization of research domains (Leahey and Reikowsky 2008), and the instability of specialty fields over the course of time (Parsons and Platt 1990).

The need for the construction of a new institutional data set that allows the measurement of DoL and Spec. on the meso-level is thus founded on four arguments. For one, DoL and Spec. are understudied on the level of the institution and availability of microdata is restricted to a short time period, which will not suffice to get a grasp on major dynamics of DoL and Spec. Secondly, in the rare cases where an institutional perspective is adopted, it still relies on the bibliometric hypothesis instead of using empirically observed institutional data, which is a problem frequently addressed in the scientometrics literature. And finally, third, the latter either lack granularity when e.g., interdisciplinarity is analyzed or concern only the specialization concentration component when the subject mix is accounted for. Consequently, a new dataset was constructed that will be presented in the upcoming sections. To the best of our knowledge[14], this is the first dataset that allows for an operationalization (of all four components) of DoL and Spec. on intra-institutional level based on empirically observed institutional data.

3.1.4 The 'Denomination Hypothesis'

Before a dataset can be compiled, it needs to be clarified what kind of data conveys information on task division and coordination, as well as specialization concentration and depth on an intra-institutional level. In this section, the *denomination hypothesis* will be motivated, meaning that in the here introduced dataset the institutionalized research domains in form of professorial denominations, derived from designations of professorial chairs and academic departments, serve as informational nucleus to make judgments on DoL and Spec. Certainly, the reader may wonder what the potential benefit of substituting the above criticized *bibliometric hypothesis* with a different *denomination hypothesis* could be. Also, it could be doubted if such a denomination sufficiently demarcates a research domain granularly enough or if profes-

[14] Apart from own research on existing databases, we were in contact with Daniel Wagner-Schuster from the ETER project, Steven Brint and David Frank from the University of California, Stefan Brings from DESTATIS (among many others). All contacts affirmed that intra-institutional microdata of universities over a longer time period does not exist and encouraged the collection of a new dataset.

sorial denominations are not already sufficiently accounted for by bibliometric classification schemes. Some might even question if they convey meaningful information at all and argue that their only relevance is signaling expertise for laymen outside of academia.

Regarding the latter, rich evidence in the sociology of science exists, which either explicitly or implicitly assumes institutionalized denominations to be meaningful in differentiating research domains for people within academia. In their seminal work on the 'Laboratory Life', Latour and Woolgar (1986) describe the interactions observed in a laboratory from their perspective as sociologists:

> 'Although our observer shares the same broad cultural knowledge as scientists, he has never seen a laboratory before and has no knowledge of the particular field within which laboratory members are working. He is enough of an insider to understand the general purpose of walls, chairs, coats and so on, but not enough to know what terms like TRF, Hemoglobin, and "buffer" mean. Even without knowledge of these terms however, he can not fail to note the striking distinction between two areas of the laboratory. [...] Individuals referred to as doctors read and write in offices in section A while other staff, known as technicians, spend most of their time handling equipment in section B. Each of sections A and B can be further subdivided. Section B appears to comprise two quite separate wings: in the wing referred to by participants as the "physiology side" there are both animals and apparatus: in the "chemistry side" there are no animals. The people from one wing rarely go into the other. (45)'

Given their methodology, adopting the perspective of anthropologists, they (pretend to) empirically observe the laboratory with a bare minimum of knowledge about the research process in the natural sciences. Under those requirements, it is striking how their very first observations consider and differentiate the division of labor in teams as described in the sections above from the institutionalized task division promoted in this work. Further, for the description of the latter they are instantly falling back on the (denomination-like) terms 'physiology side' and 'chemistry side' to delineate two supposedly different institutionalized (and in this case also organizationally or spatially separated) tasks of the production process they observe in the laboratory.

Also, we too find evidence here that those two sides, which are delineated from one another by the denominations, 'physiology' and 'chemistry' and by the characteristic of a low interpersonal exchange, are in some form coordinated by their belonging to one laboratory. The latter is also described by a particular denomination, which the authors later describe as important tool for the defining the laboratory's research domain in institutional contexts:

> 'Our observer noticed that when asked by a total stranger, members of the laboratory replied that they worked (or were) "in neuroendocrinology" They went on to explain that neuroendocrinology was the result of a hybridization which had taken place in the 1940s between neurology, described as the science of the hormonal system. It occurred to our observer that such location in a 'field' facilitated the correspondence between a particular group, network, or laboratory and a complex mixture of beliefs, habits, systematized knowledge, exemplary achievements, experimental practices, oral traditions, and craft skills. Although referred to as the "culture" in anthropology, this latter set of attributes is commonly subsumed under the term paradigm when applied to people calling themselves scientists (54)'

Indeed, in his *Homo Academicus* (1988) Bourdieu argues that while the border between institutionalized and less objective characteristics (e.g., prestige) is relatively fluent, the degree of objectification in academia starts with the titles put forward when introducing one's own person. The latter need not be limited of course to a denomination defining a research domain but also include the affiliation to a (renowned) institution, positions of power (dean) or (specific) academic titles.

Assuming that professorial denominations are meaningful, the question remains why an analysis of e.g., subject mix or interdisciplinarity based on the latter should be considered an improvement in comparison to analysis based on the *bibliometric hypothesis*. Bechtel (1993) argues that when social institutions are ignored in science studies, operative units of science get confused with globally defined areas of knowledge such as physics or biology, whereas the latter is divided in a variety of smaller units, which come in different sizes and with different scopes of research. In principle, two biologists could share as little in social connection and cognitive labor as a literature professor and a biologist. The authors thus argue that when intra- and interdisciplinary work is to be assessed, a more granular perspective is necessary, and they suggest that the most specific unit features are institutionalized at the level of the laboratory or academic department.[15] Departments define the demarcation of a research domain through institutionalization of e.g., the awarding of degrees and defining the scope knowledge future researchers in this area must have. In terms of the designation of the latter, scientists identify themselves as e.g., 'anatomists', 'biochemist' or 'cell biologist'. (Bechtel 1993)

This process of institutionalization of a research domain goes beyond the mere choice of a topic that is interesting to do research on. Indeed, the institutionalization is decisive for the viability of scientific paradigms. Clark (1974) convincingly argued that the sociological paradigm of Emile Durkheim lacked the institutionalization in professorial chairs (in French universities) and thus vanished quickly because it depended exclusively on the continuation of the tradition by his scholars (or the viability of their invisible college). In this context, Clark explicitly highlights that the importance of institutionalization is linked to the differentiation following division of labor favored by growing numbers of scientists.[16]

In accordance with the latter, Cohen (2021) argues that e.g., a denomination like 'physics' denotes a category defining knowledge about particular aspects of nature parallelly to a group of people in academia called 'physicists', which have authority over the latter knowledge domain and are interested in maintaining both formal and informal boundaries of the category. One of the formal aspects of boundaries are e.g., job titles, which legitimize academic expertise in a certain area. The author argues that such boundaries are productive insofar as e.g., in 'physics' the development of specialties in which physical knowledge can advance and new 'physicists' can be trained. On the contrary, boundaries might hinder productivity whenever they only serve the purpose of consolidating existing power structures and impede the pursuit of promising collaboration forms. (Cohen 2021)

[15] In context of European institutions, the most specific unit (apart from the laboratory or institute in the natural sciences) is the professorial chair.

[16] Clark (1974) further states that this increased differentiation allows for more specialization, which is coordinated at e.g., conferences or through invisible colleges. This understanding of institutionalization of differentiation as division of labor is consistent with the notion developed in the theoretical part of this work.

A clear advantage of basing any form of DoL and Spec. measure on professorial denominations is that they reflect this institutionalization process whereas e.g., interdisciplinarity measures based on bibliometric data disregard the latter (by adopting an output-based perspective). Therefore, the latter could signal coordination of research in cases, where in reality no true knowledge integration happened. Rousseau et al. (2019) argue that an ideal situation for studying interdisciplinarity would be if scientists classified themselves in scientific fields (according to a classification scheme). Arguably, professorial denominations are exactly that, only without the limits imposed by arbitrarily predefining a classification scheme.

Further, denominations contain not only information on institutionalized task differentiation and coordination, but also on specialization depth. Comparing it to the categories defined e.g., in the Web of Science scheme, a denomination could be as specialized as an individual subject area, span over (parts of) a research field or even (in earlier times) a whole discipline. Also, and this is a major advantage of using professorial denominations instead of exogenously defined categories, they could also delineate a research domain that is narrower than the scope of a subject area. Additionally, and this is a particularly desirable property, denominations could shed light on novel combinations of existing denominations, thus providing information on the institutionalization of coordination of usually separated domains. As a consequence, professorial denominations might better reflect dynamics of task differentiation, coordination and the gravity of specialization over the course of time.

Historically, the institutionalization of a research domain was often linked to the patronage of some peerless scientist. As described by Merton (1973), the hierarchy of specialties can be expressed by denominations first installed by peerless scientists and is not necessarily limited to three levels (e.g., subject, field, discipline):

> 'Accordingly, these peerless scientists are typically included also in the next highest ranks of eponymy, in which they are credited with having a fathered a new science or a new branch of science. Of the illustrious Fathers of this or that science (or of this or that specialty), there is an end, but an end not easily reached. Consider only these few, culled from a list many times this length: Morgagni, the Father of Pathology, Cuvier, the Father of Paleontology, […] and of course

> Comte, the Father of Sociology. In a science as farflung and differentiated as chemistry, there is room for several paternities. If Robert Boyle is the undisputed father of Chemistry […] then Priestly is the Father of Pneumatic Chemistry, […] and the nonpareil Willard Gibbs, the Father of Physical Chemistry. […] Once established, this eponymous pattern is stepped up to extremes. Each new specialty has its own parent, whose identity is often known only to those at work within the specialty. (298 pp)'

According to Parsons and Platt (1990) this patronage was enforced by the European professorial chair system dominating the beginnings of modern science, where one professor was usually in charge (of the boundaries) of one research domain. The introduction of the department system is considered a key innovation of US institutions, which helped relieve this patronage, sparked renegotiations of entrenched boundaries and facilitated the institutionalization of coordination and collaboration within a research domain. They further argue that the establishment of the university profession coincided with a first preliminary systematization of denominations, which we might consider to be disciplines or even broader epistemic branches like e.g., humanities and natural sciences. The authors attribute the compartmentalization of the latter in differentiated tasks like e.g., 'physics', 'chemistry' and 'biology' as process of ongoing specialization. (Parsons and Platt 1990)

A detailed description of the institutionalization process can be found in the work of Mullins (1974). He describes the development of a paradigm into an institutionalized specialty for the example of 'molecular biology'. One key aspect of this process is the installation of new departments and the rebranding of existing 'biology' departments in 'molecular biology'. In accordance with the remarks of Parsons and Platt (1990), this process was facilitated by the decentralized and competitive organization of US universities.

Given the here presented statements of outstanding sociologists of science on the importance of institutionalization of research domains defined by denominations, it can securely be claimed that the professorial denomination (, derived from the designation of a professorial chair or academic department) conveys meaningful information on the differentiation of institutionalized tasks, their coordination, as well as their specialization depth in the scientific production process. Since specialization concentration is defined as the concentration of quantities of academic staff on the differentiated tasks, this component of DoL and Spec. can be operationalized when numbers of academic staff according to denominations are documented.

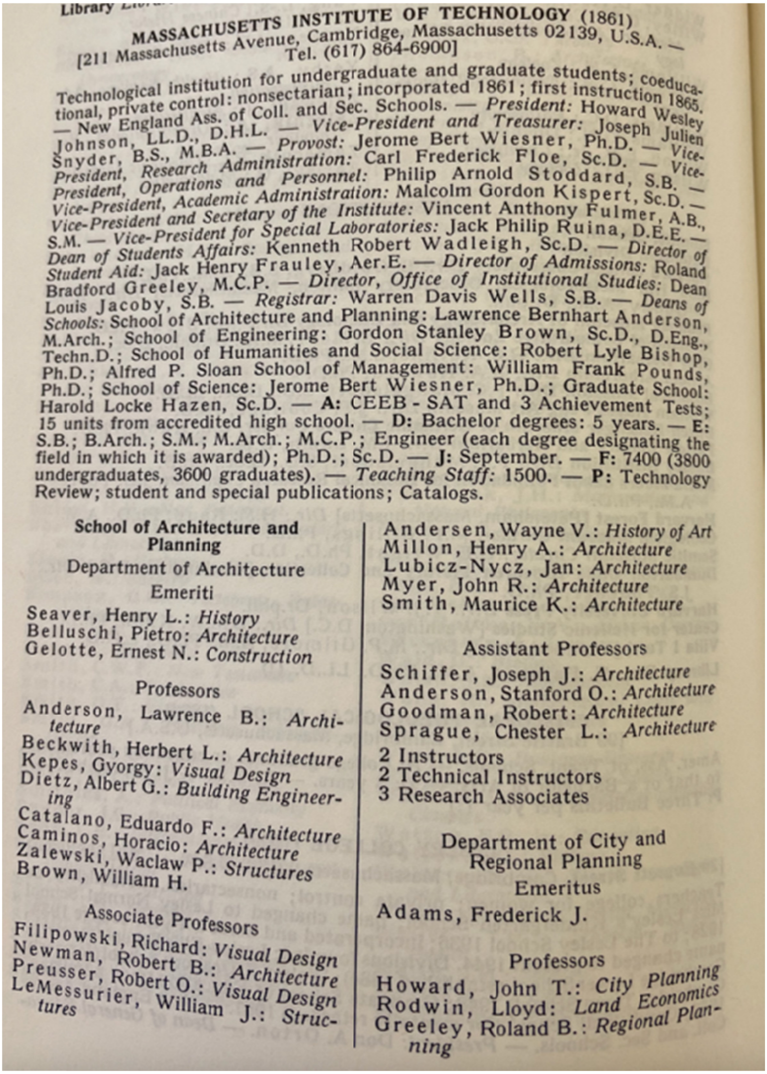
3.2 Sources, Dataset Compilation and Construction of Variables

3.2.1 Sources

Data from two different sources was used to operationalize the *denomination hypothesis* motivated above. The first source is the *Minerva – Jahrbuch der gelehrten Welt*, a collection of yearbooks listing general information and information on professorial staff, available at the Bavarian State Library (*Bayrische Staatsbibliothek*) in Munich (Kukula 1891-1930; Oestreich and Degener 1952; Schuder 1956-1969). The latter was issued first in 1891 and discontinued with final issue (nr. 35) in 1970. In the first years, the yearbook for the whole world was issued annually (up until 1921) and then appeared approximately every two years until 1928. The last issues (33-35) appeared in 1938, from 1952-1956 and 1966-1970 separately for European and non-European higher education institutions.

In figure 1, an example entry for two universities listed in the last issue is given. Each entry starts with a brief introduction into the institution's history and occasionally information on its organizational structure and curricula, people in charge of university management, staff and student numbers and funding are provided. Availability of the supplementary information varies largely across the issues and region. So, for example information on student numbers and funding was relatively regularly available for German institutions in the first issues, but in the last issues not available at all for US American universities. The latter equally concerns universities' organizational structure. Whereas in earlier periods, information on the assignments of professors according to schools, departments, institutes, clinics and laboratories was frequently completely available, this is rarely the case for later issues. The growing size and number of issues for one point in time and the change of frequency of publication indicate that collecting full information on all universities worldwide eventually overextended the capacity of the editors.

Fig. 1: 1960s-decade exemplary entries of MIT and LMU Munich in Schuder, W. (Ed.) (1966; 1969). Minerva. Jahrbuch der gelehrten Welt. 35 (1;2) De Gruyter



Library ...

MASSACHUSETTS INSTITUTE OF TECHNOLOGY (1861)
[211 Massachusetts Avenue, Cambridge, Massachusetts 02139, U.S.A. — Tel. (617) 864-6900]

Technological institution for undergraduate and graduate students; coeducational, private control: nonsectarian; incorporated 1861; first instruction 1865. — New England Ass. of Coll. and Sec. Schools. — *President:* Howard Wesley Johnson, LL.D., D.H.L. — *Vice-President and Treasurer:* Joseph Julien Snyder, B.S., M.B.A. — *Provost:* Jerome Bert Wiesner, Ph.D. — *Vice-President, Research Administration:* Carl Frederick Floe, Sc.D. — *Vice-President, Operations and Personnel:* Philip Arnold Stoddard, S.B. — *Vice-President, Academic Administration:* Malcolm Gordon Kispert, Sc.D. — *Vice-President and Secretary of the Institute:* Vincent Anthony Fulmer, A.B., S.M. — *Vice-President for Special Laboratories:* Jack Philip Ruina, D.E.E. — *Dean of Students Affairs:* Kenneth Robert Wadleigh, Sc.D. — *Director of Student Aid:* Jack Henry Frauley, Aer.E. — *Director of Admissions:* Roland Bradford Greeley, M.C.P. — *Director, Office of Institutional Studies:* Dean Louis Jacoby, S.B. — *Registrar:* Warren Davis Wells, S.B. — *Deans of Schools:* School of Architecture and Planning: Lawrence Bernhart Anderson, M.Arch.; School of Engineering: Gordon Stanley Brown, Sc.D., D.Eng., Techn.D.; School of Humanities and Social Science: Robert Lyle Bishop, Ph.D.; Alfred P. Sloan School of Management: William Frank Pounds, Ph.D.; School of Science: Jerome Bert Wiesner, Ph.D.; Graduate School: Harold Locke Hazen, Sc.D. — **A:** CEEB - SAT and 3 Achievement Tests; 15 units from accredited high school. — **D:** Bachelor degrees: 5 years. — **E:** S.B.; B.Arch.; S.M.; M.Arch.; M.C.P.; Engineer (each degree designating the field in which it is awarded); Ph.D.; Sc.D. — **J:** September. — **F:** 7400 (3800 undergraduates, 3600 graduates). — *Teaching Staff:* 1500. — **P:** Technology Review; student and special publications; Catalogs.

School of Architecture and Planning

Department of Architecture

Emeriti

Seaver, Henry L.: *History*
Belluschi, Pietro: *Architecture*
Gelotte, Ernest N.: *Construction*

Professors

Anderson, Lawrence B.: *Architecture*
Beckwith, Herbert L.: *Architecture*
Kepes, Gyorgy: *Visual Design*
Dietz, Albert G.: *Building Engineering*
Catalano, Eduardo F.: *Architecture*
Caminos, Horacio: *Architecture*
Zalewski, Waclaw P.: *Structures*
Brown, William H.

Associate Professors

Filipowski, Richard: *Visual Design*
Newman, Robert B.: *Architecture*
Preusser, Robert O.: *Visual Design*
LeMessurier, William J.: *Structures*
Andersen, Wayne V.: *History of Art*
Millon, Henry A.: *Architecture*
Lubicz-Nycz, Jan: *Architecture*
Myer, John R.: *Architecture*
Smith, Maurice K.: *Architecture*

Assistant Professors

Schiffer, Joseph J.: *Architecture*
Anderson, Stanford O.: *Architecture*
Goodman, Robert: *Architecture*
Sprague, Chester L.: *Architecture*

2 Instructors
2 Technical Instructors
3 Research Associates

Department of City and Regional Planning

Emeritus

Adams, Frederick J.

Professors

Howard, John T.: *City Planning*
Rodwin, Lloyd: *Land Economics*
Greeley, Roland B.: *Regional Planning*

München (Bayern, Deutschland BRD)

LUDWIG-MAXIMILIANS-UNIVERSITÄT (1472)

[8 München 22, Geschwister Schollplatz 1. — Tel. 22661]

Gegründet von Herzog Ludwig dem Reichen mit Genehmigung des Papstes Pius II. in Ingolstadt 1472; während der Reformation Zufluchtsstätte des Katholizismus, Ausgangspunkt der Gegenreformation; Sitz der Jesuiten von 1556—1772; Verlegung der Universität durch Kurfürst Max IV. Josef nach Landshut 1802; Namensumbenennung in Ludovica-Maximileana; 1826 erhielt die Universität ihren endgültigen Sitz in München; durch zunehmende Spezialisierung der Wissenszweige, Ausgestaltung von Lehrstühlen, Instituten und Seminaren, Bau großer Kliniken wachsende Bedeutung. — *Rektor:* Prof. Dr. Ludwig Kotter. — *Prorektor:* Prof. Dr. Ing. Dr. E. Wiberg. — *Syndikus:* Dr. Bruno Kadner. — **A:** Reifezeugnis; Zulassungsbeschränkungen bestehen bei der Überfüllung der Universität in den Fächern: Chemie, Medizin und Pharmazie. — **D:** durchschnittl. Studiendauer 8 Sem.; Medizin: 11 Sem.; tierärztl. und zahnärztliches Studium: 10 Sem.; Chemie: 12—14 Sem. — **E:** Dr. theol., Dr. jur., Dr. rer. pol., Dr. med., Dr. med. dent., Dr. med. vet., Dr. phil., Dr. rer. nat., Lizentiat der Theologie, Dipl.-Volkswirt, Dipl.-Kaufmann, Dipl.-Handelslehrer, Dipl.-Forstwirt, Dipl.-Psychologe, Dipl.-Chemiker, Dipl.-Geologe, Dipl.-Physiker, Dipl.-Mathematiker, Dipl.-Geophysiker, Dipl.-Meteorologe; sämtliche Abschlußprüfungen für Theologen, Juristen, Mediziner, Zahnmediziner, Tierärzte sowie die Lehramtsexamina. — **J:** April und Oktober. — **F:** ca 20 000. — *Lehrkörper:* 740.

Theologische Fakultät

Ordentliche öffentliche Professoren

Schmaus, Michael: *Dogmatik*
*Söhngen, Gottlieb: *Fundamentaltheologie und theologische Propädeutik*
*Schmid, Josef: *Neutestamentliche Exegese, biblische Hermeneutik*
Egenter, Richard: *Moraltheologie*
*Pascher, Joseph: *Liturgiewissenschaft und Pastoraltheologie*
Mörsdorf, Klaus: *Kirchenrecht*
Kampmann, Theoderich: *Pädagogik, Katechetik und Homiletik*
Ziegler, Adolf Wilhelm: *Kirchengeschichte des Altertums, Patrologie*
Kuss, Otto: *Neutestamentliche Exegese, biblische Hermeneutik*
Hamp, Vinzenz: *Alttestamentliche Einleitung, Exegese*
Keilbach, Wilhelm: *Systematische scholastische Philosophie und theologische Propädeutik*
Fries, Heinrich: *Fundamentaltheologie*
Tüchle, Hermann: *Kirchengeschichte des Mittelalters und der Neuzeit* (Dekan)
Weinzierl, Karl: *Kirchliche Rechtsgeschichte*
Dürig, Walter: *Liturgiewissenschaft und Pastoraltheologie*
Scheuermann, Audomar: *Kanonisches Prozeß- und Strafrecht*
Schwaiger, Georg: *Kirchliche Geschichte, Kunstgeschichte und Archäologie Bayerns*
Dettloff, Werner Rainer: *Geschichte des christlichen Glaubens*

Mit der Wahrnehmung einer oö. Professur betraut

Giers, Joachim: *Christliche Soziallehre, Allgemeine Religionssoziologie*

Planmäßiger außerordentlicher Professor

Brechter, Suso O.S.B.: *Missionswissenschaft*

Außerplanmäßiger Professor

*Hofmeister, Philipp O.S.B.: *Kanonisches Prozeß- und Strafrecht*

Honorar-Professoren

Lang, Hugo O.S.B.: *Enzyklopädie der Theologie*
Dambeck, Franz: *Geschichte der christlichen Kunst*

1 Privatdozent
4 Lehrbeauftragte

Juristische Fakultät

Ordentliche öffentliche Professoren

*Kaufmann, Erich: *Völkerrecht und Rechtsphilosophie*
*Apelt, Willibald: *Staats- und Verwaltungsrecht*
*Hueck, Alfred: *Bürgerliches Recht, Handelsrecht, Arbeitsrecht und Wirtschaftsrecht*
*Heckel, Johannes: *Öffentliches Recht, insbesondere Kirchenrecht, Deutsches Staats- und Verwaltungsrecht*
Kunkel, Wolfgang: *Römisches und Bürgerliches Recht*

Fig. 2: 1970s-decade exemplary entries of the MIT Bulletin 1975-76 (General Catalogue Issue 1975) and the LMU Munich 'Personen- und Vorlesungsverzeichnis 1975 (Sommersemester)'

Department of Ocean Engineering

Ira Dyer, Ph.D.
Professor of Ocean Engineering
Head of the Department

Professors

Martin Aaron Abkowitz, Ph.D.
Professor of Ocean Engineering

Alexander Douglas Carmichael, Ph.D.
Professor of Power Engineering

John Harvey Evans, B.Eng.
Professor of Naval Architecture

Ernst Gabriel Frankel, Mar. Mech E.
Professor of Marine Systems

Alfred Adolf Heinrich Keil, Dr. Rer. Nat.
Professor of Ocean Engineering
Dean of the School of Engineering

Justin Elliot Kerwin, Ph.D.
Professor of Naval Architecture

Patrick Leehey, Ph.D.
Professor of Naval Architecture
Professor of Applied Mechanics

Philip Mandel, B.S.
Professor of Naval Architecture

Koichi Masubuchi, D.Eng.
Professor of Ocean Engineering
Professor of Materials Science

John Nicholas Newman, Sc.D.
Professor of Naval Architecture

Kevin James O'Toole, Nav.E.
Professor of Naval Architecture
Professor of Naval Science (Visiting)
Director, Office of Naval Science

Shannon Curtis Powell, Dott.Ing.
Professor of Marine Engineering (Absent)

Associate Professors

Arthur Bernard Baggeroer, Sc.D.
Associate Professor of Ocean Engineering
Associate Professor of Electrical Engineering

Chryssostomos Chryssostomidis Ph.D.
Associate Professor of Naval Architecture

John William Devanney, III, Ph.D.
Associate Professor of Marine Systems

Clark Graham, Ph.D.
Associate Professor of Marine Systems

Wesley Leroy Harris, Sr., Ph.D.
Associate Professor of Ocean Engineering
Associate Professor of Aeronautics and Astronautics

Norman Jones, Ph.D.
Associate Professor of Ocean Engineering

Jerome H. Milgram, Ph.D.
Associate Professor of Naval Architecture

J. Daniel Nyhart, J.D.
Associate Professor of Management
Special Assistant to the Chancellor

Assistant Professors

Edward Clarence Kern, Ph.D.
Assistant Professor of Ocean Engineering

Judith Tegger Kildow, Ph.D.
Assistant Professor of Ocean Policy

Henry Stuart Marcus, D.B.A.
Assistant Professor of Marine Systems

Owen Horace Oakley, Jr., Ph.D.
Assistant Professor of Ocean Engineering

Ronald Wei-Chung Yeung, Ph.D.
Assistant Professor of Ocean Engineering

Senior Lecturers

Harry A. Jackson, B.S. (Visiting)
Miguel Chapero Junger, Sc.D. (Visiting)
William S. Pellini, B.S. (Visiting)
Willard Franklin Searle, Jr., Nav. E. (Visiting)

Lecturers

William Avery Baker, S.B.
Damon Ellis Cummings, Ph.D. (Visiting)
David A. Ross, Ph.D. (Visiting)

Instructor

Robert Joseph Stewart, S.M.

Administrative Officer

Keatinge Keays, Nav.E.

Research Associates

Kuoyu Itoga, Ph.D.
Johm Kim Vandiver, Ph.D.

Visiting Engineer

Oddvar Frydenlund

Professors Emeriti

Evers Burtner, S.B.
Associate Professor of Naval Architecture and Marine Engineering, Emeritus

Frank Mendell Lewis
Professor of Marine Engineering, Emeritus

Norman Judson Padelford, Ph.D., LL.D.
Professor of Political Science, Emeritus

Lehrkörper

Ordentliche Professoren:

***Frey Emil Karl** (22.12.30), Dr.med., Dr.rer.nat.h.c., Dr.rer.nat.h.c., für Chirurgie – liest nicht –, privat: M 27, Arberstraße 16 (48 07 46)

***Butenandt Adolf** (1933), Dr.phil., Dr.h.c.mult., Ehrenpräsident der Max-Planck-Gesellschaft, Direktor (em.) des Max-Planck-Instituts für Biochemie, 8033 Martinsried (8 58 53 64); für Physiologische Chemie, privat: M 60, Marsopstr.5 (88 54 90)

***Büngeler Walter** (Dez.1934), Dr.med., für Allgemeine Pathologie und Patholog.Anatomie – liest nicht –, privat: M 19, Schlagintweitstraße 15 (15 25 35)

***Wiskott Alfred** (1.1.38), Dr.med., für Kinderheilkunde – liest nicht –, privat: M 2, Platenstraße 1/0 (77 35 24)

***Herrmann Alexander** (1.1.39), Dr.med., für Hals-,Nasen-, Ohrenkrankheiten – liest nicht –, privat: M-Solln, Voltzweg 5 (79 79 78)

***Bodechtel Gustav** (21.6.40), Dr.med., Dr.phil., für innere Medizin, Leiter des Instituts der Friedrich Baur-Stiftung (s.II.Med.Klinik) – liest nicht –, privat: M 19, Furtwänglerstr.14 (15 62 32)

***Kramer Kurt** (1.4.44), Dr.med., für Physiologie, Komm.Vorstand des Physiolog.Instituts, M 2, Pettenkoferstraße 12 (5 99 61), privat: M 2, Schubertstr.4 (53 11 99)

***Forst August Wilhelm** (1.6.46), Dr.med., Dr.phil., Dr.med.vet.h.c., für Pharmakologie, Toxikologie und Chemotherapie, privat: M 80, Schönbergstraße 12 (98 02 91)

***Eyer Hermann** (1.8.46), Dr.phil.nat., Dr.med., für Hygiene und med. Mikrobiologie, Komm.Vorstand des Max-v.Pettenkofer-Inst. für Hygiene und Med. Mikrobiologie, M 2, Pettenkoferstraße 9 a (53 93 21), privat: M 90, Gabriel-Max-Straße 14 (64 52 84)

***Laves Wolfgang** (1.3.47), Dr.med., Prof.h.c., für Gerichtliche Medizin und Versicherungsmedizin – liest nicht –, privat: M 40, Leopoldstraße 135 (37 92 36)

Kiese Manfred (1.8.50), Dr.med., für Pharmakologie, Toxikologie und Chemotherapie, Vorstand des Pharmakologischen Instituts, M 2, Nußbaumstraße 26 (5 38 41), privat: M 80, Cuvilliesstraße 21/III (98 64 35)

***Zenker Rudolf** (1.4.51), Dr.med., Dr.med.h.c., für Chirurgie – liest nicht –, privat: M 90, Hauensteinstraße 14 (64 61 00)

***Bachmann Rudolf** (24.7.52), Dr.med., für Anatomie – liest nicht –, privat: M 40, Osterwaldstraße 59/VI

***Schwiegk Herbert** (11.9.52), Dr.med., für Innere Medizin, Komm.Direktor der I. Med. Klinik, M 2, Ziemssenstr. 1 (53 99 11), privat: M 90, Hermine-Bland-Straße 4 (64 51 04)

***Kolle Kurt** (1.12.52), Dr.med., für Psychiatrie und Neurologie – liest nicht –, privat: 813 Starnberg/Obb., Oberholzstraße 10 (0 81 51 / 65 24)

Bücher Theodor (1953), Dr.rer.nat., Dr.med.h.c., für Physiologische Chemie, Vorstand des Instituts für Physiolog.Chemie, M 2, Goethestraße 33 (5 99 61), privat: M 90, Hermelinweg 7 (63 01 37)

***Lange Max** (1.10.54), Dr.med., für Orthopädie – liest nicht –, privat: M-Solln, Knotestraße 10

Witt Alfred Nikolaus (1.10.54), Dr.med., für Orthopädie, Direktor der Orthopäd.Klinik, M 90, Harlachinger Straße 51 (6 21 11) und Orthop.Poliklinik, Pettenkoferstraße 8 a (5 99 41), privat: M 90, Lengmoosstraße 5

The idea of the *Minerva* collection is to provide a reference work that allows researchers to identify peers working on similar or related topics to their own research. Therefore, the listing of professors and professorial denomination (alongside contact details) is at heart of the publication. Arguably, since this publication was aimed at researchers for identifying specialties of peers and coordination of their research (or at least correspondence and exchange of ideas with them), it should be seen as empirical validation for the theoretical remarks on why a professorial denomination conveys information on division of labor and specialization. Listings of all university professors are completely available in all issues and for all (established institutions used in this work) the professors are assigned a denomination demarcating topics they are researching on. As indicated in figure 1, professors are further divided in full, associate and assistant professors (and equivalents in the respective country's university system, e.g., *ordentlich-öffentlich, außerordentlich, außerplanmäßig* in the German system or senior lecturers in the UK system).

Since the *Minerva* publication was discontinued in 1970, a second source had to be used to extract the data on professorial denominations. For all institutions considered, professorial denominations were collected using digitalized academic calendars or bulletins, course catalogues (*Vorlesungsverzeichnisse*) or comparable documents individually requested at the universities' administrations, libraries or archives whenever available. In figure 2, the example of the MIT bulletin of 1975 to 1976 and the course catalogue of the LMU Munich of 1975 are given. As can easily be obtained, even though there are substantial differences between the publication types and the information they provide, the nucleus of (name of) professor and professorial denomination is given and further listed according to tenure status.

As pointed out in section 3.1.1, the documentation of professorial denominations is limited to natural sciences and engineering disciplines for theoretical reasons. In practice, distinguishing between denominations of natural sciences and engineering disciplines and denominations of formal and social sciences, as well as humanities was in the majority of cases intuitively possible (as opposed to the coding according to Web of Science subject categories, see section 3.2.3). In some cases, where denominations referred to technical terms unknown to us, a quick google search would enlighten whether there was a connection to the natural sciences and engineering disciplines or not. In cases, where the topics a denomination demarcates had a connection to the latter, but also to the other epistemic branches, it was documented as well. Given the scope of this work is to examine the coordination of research (also on the aggregate level of the discipline) this should come across as a natural choice.

3.2.2 Compilation

In this section, the choice of the time period considered as well as sampling of institutions will be motivated. First, the time frame considered is determined. Clearly, the degree of division of labor and specialization does not change quickly. Much rather, tenured professors supposedly keep their denomination for a lifetime. Changes will only occur very gradually, when the latter are given emeritus status or new positions are created and e.g., assistant professors specialize in new areas of research. Possibly, the latter rationale can be rediscovered by the movement from annual issues to issues in decade interval towards the end of the *Minerva* publication. The slow rate with which major changes among the professorial staff occurred might not have justified the workload linked to annual updates.

Also, since the manual documentation of denominations is very labor-intensive, limiting the documentations to fewer, spread apart periods will free resources to cover a larger span in time. In this work, the time period 1890 to 2020 will thus be covered in a decade-interval.

This time period is believed to represent the development of science as we understand it today or as Leydesdorff (2021: 3) summarized it:

> 'this […] period (1870-1910) […] the scientific-technical revolution … cannot be understood in terms of specific innovations – as is the case of the Industrial Revolution, which may be adequately characterized by a handful of key inventions – but must be understood rather in its totality as a mode of production into which science and exhaustive engineering investigations have been integrated as part of ordinary functioning. The key innovation is not to be found in chemistry, electronics, automatic machinery, aeronautics, atomic physics, or any of the products of these science-technologies, but rather in the transformation of science itself into capital.'

The decades 1930 to 1940 and 1940 to 1950 are not considered in the sample. For the latter decade, no Minerva publication was issued. The issue of the former period was published in the 'Third Reich' and the data collection on foreign universities is incomplete and cannot be proven to be reliable. Further, ethical concerns would arise with the documentation of natural sciences' denominations of the German institutions, which were both politically and due to their research activities involved in the atrocities committed by the Nazi regime. Sometimes professorial denominations like 'race hygiene' i.e., would give away the connection to the Nazi ideology, sometimes professors would hold (back then) commonly used professorial denominations like 'hygiene' and nonetheless conduct research on 'eugenics' and be heavily involved as NSDAP party member (see e.g., the role of Fritz Lenz in science landscape of the *third Reich* in Weber (1989), who is listed in issue 32 in the entry of the LMU Munich). Since a clear differentiation was not possible, we decided to exclude the data from issue 32 (1936) completely from the sample.

Secondly, the choice of institutions needs to be motivated. Given the theoretical remarks of chapter 2, it should be clear that while peer review and publication, as well as citation (and consequently recognition) bear information on the quality of a piece of research, this information is by no means perfect and probably biased by e.g., the *Matthew effect*. A second measure (, apart from limiting the data collection to natural sciences and engineering disciplines) put in place to ensure that the conditional efficiency analysis, conducted in the next chapter, actually contains information on the impact of DoL and Spec. on progressiveness (not mere quantitative productivity), is to limit the institutions, for which denominations are documented to the ones, for which we assume that scientific progress has already been achieved.

Regardless of all the flaws linked to university rankings as a fair measure of research or teaching quality (see 3.1.2.2 for details) they are a suitable tool for such an initial limitation of potential institutions. Since they measure performance based on the *bibliometric hypothesis,* they allow us to construct a sample of universities for which it is already known that they are good at producing large numbers of publication output. Also, the performance measurement in university rankings is partially built on quality criteria like e.g., survey data and the number of Nobel prizes received. And finally, third, they allow us to balance the bias towards European institutions, that would be introduced if we took data availability as primary criterion for selection. Since the issues of the Minerva collection employed were produced in the German empire, Weimar republic and the Federal Republic of Germany, data availability for institutions within the publication is biased towards German and European universities. Also, of course in the late 19th century, European universities were most certainly among the most renowned and longest established institutions in science, whereas universities in other parts of the world were not yet or just recently founded. Limiting the data acquisition process to top ranked universities will thus make sure that (European) universities that are considered have preserved their role as contributing towards scientific progress over time. (Daraio and Bonaccorsi 2016)

In summary, there are two criteria determining the sample of institutions. For one, when a university was founded before and established long enough to have a significant number of professors and denominations documented in the Minerva publication (forward-induced quality bias). And secondly, if it does count as an excellent institution today (backwards-induced quality bias). It should be critically noted that the intersections of the two criteria also introduce a bias towards institutions that are US American, Continental European or located in other English-native speaking countries (Australia and New Zealand). Then in turn, the development of division of labor and specialization needs to be assessed for institutions that operate at similar financial and infrastructural conditions over the whole period considered and this might indeed be best given for universities located in developed western societies, which were first in enforcing policies for general education of societies planting the seed for modern (higher) education institutions' success (Parsons and Platt 1990).

Therefore, the data collection process was initially limited to universities appearing simultaneously in the top 100 ranking of three most popular university rankings (Times Higher Education (THE), Shanghai Ranking, QS ranking). Since the choice of indicators in university rankings is rightfully criticized as arbitrary, we chose to consider the intersection of all three rankings, since they employ slightly different criteria (e.g., knowledge transfer in the THE, research focus in the Shanghai and teaching quality in the QS ranking) to sort out universities, which only rank highly because they perform well in particular categories (Van Raan 2019). The remaining institutions were limited to a sample for which an entry of the university was available before 1900[17] in the Minerva publication.[18] (Daraio et al. 2015a)

Effectively, the sample was initially limited to 62 universities. For the latter, academic calendars, course catalogues and archival material was requested at university libraries, archives and administrations. In total, 51 institutions took action upon the initial request. The two most frequent reasons declared for not being able to provide the requested data is that the university does not keep track of (historical) records of their academic staff (e.g., University of Edinburgh) or that it would be too labor-intensive to provide the requested resources (e.g., University of Basel). Some institutions were not willing to provide the data referring to their duty of disclosure by law being exceeded (e.g., University of Sheffield, referring to the UK Freedom of Information Act), whereas others would have only made the data available as a payable service (e.g., University of Freiburg). Another very common problem was that the organizational structure of a university proved to be incompatible with the other institutions (multi-institutional system of the university of Paris and London or the Humboldt university in Berlin, which was completely reorganized after the German reunification).

Indeed, the data collection process for later decades turned out to be extremely difficult and labor intensive for European universities (, for which good data availability in the *Minerva* publication was already secured), which seemingly put less emphasis on centralized units governing their institution. On the contrary, some US American institutions (, e.g. Harvard university archives) were able to redirect

[17] Note that the University of California system is the exception from this rule. Since they are branches of the UoC system initially concentrated at fewer locations (and since the data was made available to us), the UoC – Davis, UoC – Los Angeles and UoC – San Diego were also considered in the sample.

[18] In this first period, the data source is the Minerva publication of 1899 (issue 9) to allow the inclusion of institutions that were founded before the 1890s yet not available in issues 1-8. For all remaining periods, the issue that laid closest to the middle of each decade was chosen (1900-1910: issue 15 (1905), 1910-1920: 23 (1913), 1920-1930: 28 (1926), 1950-1960: 34 (1956) and 1960-1970: 35 (1966)). The latter scheme was equally applied to the academic calendars and course catalogues used to extract the data for the period 1970 to 2020.

and complete the request within a few workdays, providing digitalized academic calendars for the whole period with full availability of data.

Admittedly, the latter is of course only anecdotic evidence. Nonetheless, it points at a certain capacity of some institutions to monitor their task coordination and specialization in certain domains (since they collected data and could make it accessible), whereas other institutions apparently do not consider this to be an important part of university governance. Eventually, 24 institutions completed the request and provided data sources that enabled an extraction of professorial numbers and denominations. Four of the latter had to be excluded from the sample, because the documents did not allow an unambiguous assignment of professors to denominations, or because of the language barrier (e.g., University of Gent, which documented its denominations in Dutch).

For the remaining 20 universities, the distribution of the 10,167 manual entries documented is given according to institution and period in table 3. The final sample consists of 11 US American, three German, two Swiss, and single institutions from UK, Sweden, New Zealand and Australia respectively. Interestingly, even though initially we feared a European bias in the sample due to the good availability of data for European institutions in the Minerva publication, in the end US American institutions make up for more than half of the universities considered.

Nonetheless, given the comparatively small number of observations, the sample still offers a certain (even though unbalanced diversity) in characteristics.

Indeed, it can be differentiated in between technical and all-sciences universities, institutions located in English-native and non-English-native speaking countries, as well as according to region (North American, European and Oceanian). The European institutions stem from four different countries with different educational systems. Also, it can be differentiated among the US institutions in between private institutions (Caltech, Stanford, MIT), private Ivy League members (Harvard, Columbia) and state universities (, also sometimes referred to as public Ivy's) like the University of California and the University of Texas (with the University of Washington as aspiring candidate). While the effects of DoL and Spec. obtained for the latter sample certainly cannot be seen as representative for the whole scientific community, one could argue, given the diversity of the above-mentioned characteristics, that it is at least representative for most types of excellent (or highly ranked) universities.

Tab. 3: Denominations documented according to institution and period

institution	inc	1890	1900	1910	1920	1950	1960	1970	1980	1990	2000	2010
CALTECH	CAL	.	.	.	14	49	54	46	45	46	49	49
Columbia University	COL	31	57	65	87	96	70	.	.	.	.	6
ETH Zürich	ETH	31	34	36	35	64	102	.	.	.	154	55
Georg-August Universität Göttingen	GAU	.	33	36	38	64	96	104	159	196	180	.
Harvard University	HAR	47	75	77	98	119	86	70	67	69	89	.
LMU Munich	LMU	44	42	49	84	82	120	121	160	206	263	.
Massachusetts Inst. of Tech.	MIT	25	35	37	66	62	47	55	54	70	85	41
RFW Bonn	RFW	22	24	24	41	65	105	114	182	257	289	.
Stanford University	STA	.	19	36	30	42	64	26	24	27	28	29
Universität Zürich	UZH	40	23	33	32	48	67	78	112	131	150	.
University of Auckland	AUC	.	.	3	3	12	19	27	23	54	32	37
UoC – Berkeley	UCB	22	56	66	73	82	72	24	30	26	27	23
UoC – Davis	UCD	.	.	.	.	26	39	39	.	64	58	.
UoC – Los Angeles	UCL	.	.	.	.	44	31	.	35	33	46	48
UoC - San Diego	UCS	.	.	.	.	.	15	38	41	35	42	.
University of Leeds	LEE	.	28	28	28	55	55	53	.	.	177	.
University of Sydney	UOS	20	35	37	19	39	57	49	47	89	107	.
University of Texas - Austin	UTA	8	17	.	19	20	21	.	.	19	18	19
Univ. of Washington - Seattle	UOW	.	12	18	18	73	62	64	61	66	66	58
Uppsala Universitet	UPP	.	20	26	28	27	51	.	.	105	115	.

3.2.3 Denomination harmonization and matching according to WoS scheme

Before the actual variables for division of labor and specialization could be constructed, some further work on the raw denomination documented was necessary. First, the denominations had to be harmonized by translating all entries in (German) into the English language. With technical terms this was often very intuitive, yet occasionally denominations referred to very particular topics not trivial to translate. In the latter cases, the online bilingual concordance service of 'Linguee' (www.linguee.de) was used to achieve a proper translation of the denomination. As opposed to traditional translation services, 'Linguee' provides translations based on documents available on the internet, where the term of interest is used in context. In most unclear cases, this turned out to be very effective, because 'Linguee' referenced documents from the university to which the denomination belongs, where the institution itself would provide their own translation of the denomination. This harmonization process led to narrowing down the 10,167 entries initially documented to 2,549 denominations, which are semantically unique, either recombining existing topics or demarcating a specific topic differentiable from that of other denominations.

Second, in order to achieve a good compatibility of the data set with existing databases and studies based on the *bibliometric hypothesis*, the denominations were coded according to a *Web of Science* classification scheme (CWTS 2024), which allows to categorize publications and journals on three different levels of granularity (discipline, research field and subject area). By choosing this scheme, an analytical, internal perspective of science is taken over, thinking about research domains in terms of today's output as opposed to adopting a classification scheme based on sociological, historical or psychological perspectives on scientific inquiry (Bornmann and Mutz 2014). In total, the WoS classification system consists of 7 disciplines, 35 fields of research and 250 separated subject areas. After excluding all fields of research, which are not element of the definition of natural sciences and engineering disciplines provided in section 3.1.1, 3 disciplines, 18 fields of research and 166 subject areas remained, which are given in table 4.

Tab. 4: Web of Science categorization scheme limited to epistemic branches as specified in section 3.1.1

discipline	disc	research field	refi	subject area (nr.)
Multidisciplinary (misc.)	I	.	.	.
Engineering Sciences	E	Civil Engineering and Construction	E01	2
	E	Electrical Engineering and Telecomm.	E02	6
	E	Energy Science and Technology	E03	4
	E	General and Industrial Engineering	E04	4
	E	Instruments and Instrumentation	E05	2
	E	Mechanical Engineering and Aerospace	E06	5
Medical and Life Sciences	M	Health Sciences	M01	9
	M	Agriculture and Food Science	M02	8
	M	Basic Life Sciences	M03	10
	M	Biological Sciences	M04	12
	M	Basic Medical Sciences	M05	4
	M	Biomedical Sciences	M06	12
	M	Clinical Medicine	M07	34
Natural Sciences	N	Astronomy and Astrophysics	N01	1
	N	Chemistry and Chemical Engineering	N01	12
	N	Physics and Materials Science	N02	17
	N	Earth Sciences and Technology	N03	13
	N	Environmental Sciences and Technology	N04	10

See digital *Annex DS1* for full scheme including all sciences and subject areas

In practice, coding the individual denominations according to this scheme was less trivial than translating the denominations. A lot of the individual subject areas in the WoS classification do overlap (Rousseau et al. 2019) and for some denominations it took substantial effort and research on websites of institutes or departments to properly link a denomination to one or more of the 166 available options for subject areas. Assigning those denominations manually to subject areas guarantees for a high quality and consistency of the scheme.

While most denominations were attributable to one (e.g., 'Hemostaseology' → Hematology (M0713)) or two subject areas (e.g. 'Immunological Molecular Biology' → Biochemistry and Molecular Biology (M0302), Immunology (M0602)), others were attributable to multiple subject areas (e.g., 'Pharmacognosy and Analytic Pharmaceutical Chemistry and Anorganic Pharmaceutical Chemistry' → Medicinal Chemistry (M0501), Pharmacology and Pharmacy (M0608), Analytical Chemistry (N0102), Inorganic and Nuclear Chemistry (N0104)).

The different subject areas attributable could either belong to one research field and discipline ('Accident Medicine and Occupational Medicine' → Clinical Medicine (FIELD)), connect different research fields within a discipline ('Applied Genetics and Agronomy' → Agriculture and Food Sciences, Basic Life Sciences (FIELD) → Medical and Life Sciences (DISC)) or even span over different disciplines ('Engineering Science and Geophysics' → General and Industrial Engineering, Earth Sciences and Technology (FIELD) → Engineering Sciences, Natural Sciences (DISC)). The categorization of denominations according to fields and disciplines has been carried out entirely according to logic of the WoS scheme (, which of course is one of many possible ways to classify scientific activities). Consequently, the only subjective component involved in the categorization process was the assignment of subject areas, which may be retraced by looking at the full dataset containing the individual entries (sheet 'acc'), the semantically different denominations (sheet 'den') and the coding scheme (sheet 'cod') in the digital *Annex DS2*.

3.2.4 Variables

The categorization of the denominations according to the *WoS* scheme is also valuable for measuring the effects of DoL and Spec. on different levels of granularity. The latter falls in line with the idea promoted in section 2.2.3, to accommodate the sometimes very generally issued calls for inter- or transdisciplinarity by providing a more granular and concrete perspective on coordination of research activities.

One of the more concrete definitions of disciplinarity, multi-, pluri- and interdisciplinarity stems from the OECD (Apostel et al. 1972). In conjunction with the classification of denomination according to the *WoS* scheme, the latter can be applied to determine a denomination's inherent degree of disciplinarity. The latter measure may then be compared to the variables operationalized to determine division of labor and specialization to make a judgment on whether debates on disciplinarity would benefit from the more granular perspective.

Effectively, we considered a denomination to be disciplinary when it covers one subject area of a particular research field and discipline. A denomination is classified as interdisciplinary when it is attributable to subject areas belonging to at least two different disciplines in the *WoS* scheme (e.g., Engineering Sciences and Natural Sciences) The operationalization of the categories multi- and pluridisciplinary is slightly tailored to the available data yet oriented at the definition provided by Apostel et al. (1972). There, pluridisciplinarity is defined as 'Juxtaposition of disciplines more or less related, e.g. mathematics + physics (25)'. We can rediscover this information on the more granular level of subject areas and research fields, where e.g., the two different research fields 'Mathematics' and 'Physics and Materials Science' belong to the same discipline 'Natural Sciences'. Denominations are here considered 'pluridisciplinary' whenever they are attributable to different subject areas belonging to the same discipline.

Multidisciplinarity in turn was defined as 'Juxtaposition of various disciplines, sometimes with no apparent connection to one another (25)' Now of course, on level of individual denominations we may rarely ever encounter the latter. Thus, denominations were only considered multidisciplinary in case they were assigned to the subject area multidisciplinary (miscellaneous), which signals that they cover different topics, subjects and fields yet the denomination was too broad to assign individual subject categories. This only concerns a small portion within the sample (29 of the 2,459 denominations) and includes denomination like 'Applied Science', 'Natural Sciences' and 'Science' for example. The resulting four categorical variables of disciplinarity defined for individual denominations are by design more granular and informative than e.g., disciplinarity measures derived from affiliations in bibliometric databases and thus might indeed allow to enlighten the relationship of disciplinarity with the continuous variables of the DoL and Spec. components.

In table 5, the operationalization of the (theoretically derived) components of DoL and Spec. in continuous variables is given according to different levels of granularity. Regarding the operationalization of task division, two options were considered. As defined in section 2.2.2 task division means the institutionalization of a separable part of the production process into a differentiated task. So technically, task division could be operationalized by the sheer number of denominations, which reflects the division of the (teaching and research) production process of a university into different institutionalized tasks. On the contrary and in accordance with Adam Smith's idea of DoL being limited by the extent of the market, using the sheer number of denominations would neglect the size differences between institutions, which may only be able to institutionalize new tasks whenever sufficient numbers of professors are available.

Tab. 5: Operationalization of the components of DoL and Spec. in data set variables

	Division of Labor		**Specialization**	
	Task Division	Task Coordination	Concentration	Gravity
Discipline		Nr. of disciplines covered (weighted)	Nr. of professors	Individual specialization of denomination (weighted)
Research Field		Nr. of research fields covered (weighted)	Nr. of professors	
Subject Area		Nr. of subject areas covered (weighted)	Nr. of professors	
Denomination	Nr. of professors / denomination	Nr. of topics covered (weighted)	Nr. of professors	

Thus, task division here is defined by the number of professors divided by the number of denominations. A value of one would then return a size-normalized fully institutionalized task division, where on average a professor holds a denomination. Higher values in turn, point at a lower degree of institutionalized differentiation of tasks. Presumably, the latter signals that task division is not organized by institutionalizing it, but occurs e.g., rather team-based within larger departments, to which more professors are assigned sharing one common denomination. It should be noted though that while the measures employed here certainly to some extent convey information on organizational differences with low values for task division (=1) pointing at a university system operating with professorial chairs and higher values suggesting a departmental based organization, the primary concern here is the institutionalization of different tasks. Technically speaking, it is logically possible that multiple professors of an institution at a point of time share the same denomination, yet are organized in different departments, institutes, laboratories or even chairs. There is neither a logical necessity for them to cooperate, nor to (locally) cooperate at all, at least from all we know from the data. The main information conveyed and the primary interest within this work concerns the extent of the institutionalization of the differentiation of tasks.

As established in the theoretical line of thought, scientists choose to specialize in different research domains and universities allocate their professorial staff differently according to more labor-intensive research projects resulting in denominations with higher and lower numbers of professors assigned to them. The specialization concentration measure should thus allow to analyze (in-)equalities in distributions of professors according to the different institutionalized denominations. Further, the latter should be analyzed on the aggregated levels of the *WoS* categories to identify the domains an institution specializes in. For the descriptive analysis, the specialization concentration measure is analyzed by plotting the number of professors according to category of interest. Later, for the cluster and efficiency analysis, the qualitative information will be reduced to quantitative information only, by employing concentration measures (see 3.3.2).

In the theoretical remarks on specialization in section 2.2.2, a second component apart from the concentration on certain research domains alluding to the qualitative depth of specialization was defined. This gravity of specialization measure is operationalized by assigning each denomination a value of 1 if it demarcates a research topic, which is roughly equivalent to a subject area, a value of 2 if it demarcates a sub-topic that is fully contained within a subject area (and does not entirely cover the scope of the subject area) or a value of 1 if the denomination is as general as a research field or even discipline (= 0). Some denominations though cover more than one topic belonging either to one or more subject areas. 'Ophthalmology and Otology' for example covers two topics attributable to two different subject areas 'Ophthalmology' and 'Otorhinolaryngology'. It is evident that the degree of depth clearly differs for the two topics. The topic 'Ophthalmology' is identical to the subject category whereas 'Otology' demarcates a narrower scope than 'Otorhinolaryngology'. Whenever more than one topic is covered, each topic is assigned an individual value and the geometric mean is calculated for the denomination.

The latter could either be aggregated to the level of an institution at a point of time, a particular time period or an institution over the whole period. Of course, calculating geometric means for the latter needs to be normalized for the weight of an individual denomination within the considered aggregated level. So, in case the number of topics covered is calculated for an institution and a certain time period, the value of each denomination is weighted by the share of the sum of professors associated with the denomination relative to the sum of professors of an observation at the time period considered. Analogously, when the number of topics covered is calculated for the overall institution over the whole period, the value of topics covered for each denomination is weighted by the sum of professors associated with the denomination relative to the sum of professors of the institution over the whole period considered. The latter rationale was equally applied on level of subject areas, research fields and disciplines as well as applied to the specialization gravity variable.

Finally, the operationalization of the second component of division of labor, the task coordination variable, needs to be introduced. Analogously to the specialization concentration measure task coordination can be considered on different levels of granularity. Here, instead of the denomination the most granular information is assumed to be conveyed by the number of different topics a denomination coordinates. The latter could then belong to one or more different subject areas, research fields or disciplines a denomination coordinates. Here again, the aggregation to institutional level or time period is achieved for each variable by weighting the individual denominations by the share of associated professors relative to the category of interest.

3.3 Data

3.3.1 Descriptive statistics

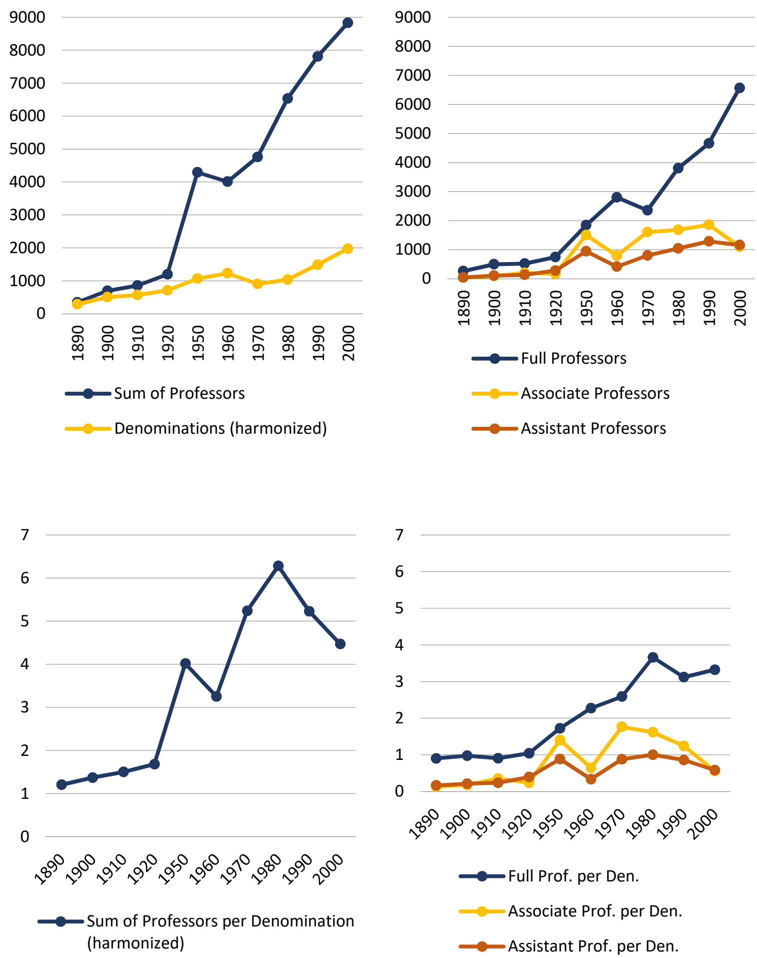
3.3.1.1 Full sample (zooming out)

In this section, a thorough descriptive analysis of the above introduced dataset is performed. For the 2,549 different denominations, documented in 10,167 individual entries, a total of 46,294 professors was counted. In figure 3 the development of sum of professors, professorial types and denominations over time is given. The year 2010 is not considered here because of the many missing observation in particular among European institutions (see table 3).

Clearly, an increasing trend in number of professors can be obtained over the considered period. For the period before the second world war (1890 to 1920) the absolute sum of professors and denominations moderately increased. When looking at the ratio of professors per denomination, we can obtain that they moderately increased for this pre-war period. When looking at the period 1950 to 2000 we can obtain a substantial shift in both patterns with enormous growth rates in overall number of professors, whereas the number of denominations follows a moderate growth path leading to significant increases in numbers of professors per denomination.

In the two graphs on the right-hand side, the numbers for the different professorial types are given. It should be noted that in the final Minerva publication used to document the data for the period 1960 to 1970 (Issue 35: 1966), for some institutions the differentiation between full, associate and assistant professors was not given. Therefore, all professors of those institutions had to be treated like full professors, which explains the gap in the trends for the different professorial types obtainable for the 1960s. Taking the latter into account, we may obtain a rather stagnating and saturating trend for increase in numbers of associate and assistant professors. The overall increase in professorial staff numbers is thus mainly caused by substantial increases in tenured positions. When relating the numbers of associate and assistant professors to the development of numbers of denominations we can even see a declining trend from 1970 on.

Fig. 3: Development of sum of professors, professorial types and denominations over time

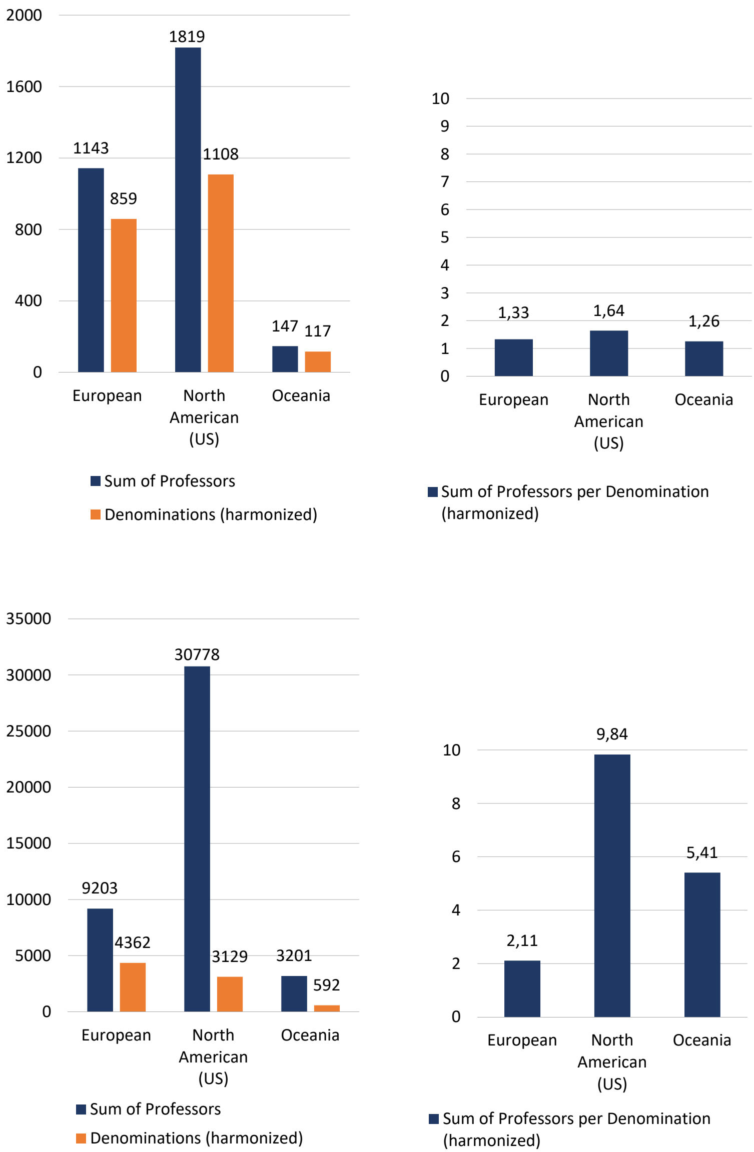


Overall, this suggests that there exists a structural break between scientific activity conducted before and after the second world war, reflected by the professorial staff numbers and the institutionalization of division of labor and specialization within the sample (Parsons and Platt 1990). Indeed, when analyzing cited references, Bornmann and Mutz (2014) find a relatively constant linear increase of 2 to 3 percent for the period between the two world wars and growth rates of about 8 to 10 percent afterwards. While the growth rates of cited references cannot be compared to the increase of professorial staff in this dataset, the structural change in size between pre- and post-war science is well-proxied by the development of professorial staff numbers in the sample.

In figure 4, sum of professors and denominations as well as the ratios are given for the different locations of the institutions according to pre-war and post-war period. While in the pre-war science the ratio of sum of professors per denomination is relatively equal for European, US American and Oceanian institutions, ranging from 1.26 to 1.64, the ratios reveal substantial differences in the post-war period. While the ratio only moderately increased from 1.33 to 2.11 professors per denomination for European institutions, it went up by a factor of 6 for US American universities (1.64 to 9.84) and over 4 for Oceanian institutions (1.26 to 5.41). It should be kept in mind though that the latter region is only represented by two institutions in the sample and its results should thus be interpreted carefully (University of Auckland, University of Sydney).

These findings coincide with the description of the differences between US American and German universities by Parsons and Platt (1990), which highlighted the substitution of the professorial chair system with the department system as the major innovation of the American university, building the organizational subunit of the faculties of arts and sciences or the professional schools where multiple full professors of the same subject area form a collegial body. The latter is reflected here by the upscaling in numbers of professors in US institutions, which seem to be decoupled from the number of denominations. For European institutions in turn, where the professorial chair system prevailed (, even though other forms of organizational and institutionalized bodies like projects, institutes, laboratories or affiliations with research organizations of course gained in importance), we can obtain that the growth in numbers of professors in the post-war period nearly

Fig. 4: Professorial staff, denominations and professors per denominations according to location for pre-war (top) and post-war science (bottom)



led to a proportional increase in numbers of denominations. When looking at the absolute numbers, this picture is confirmed. In absolute terms, the number of institutionalized tasks increased nearly by the factor 5 for European institutions yet only by factor 4 for US institutions regardless of the enormous growth in absolute numbers of university professors for the latter.

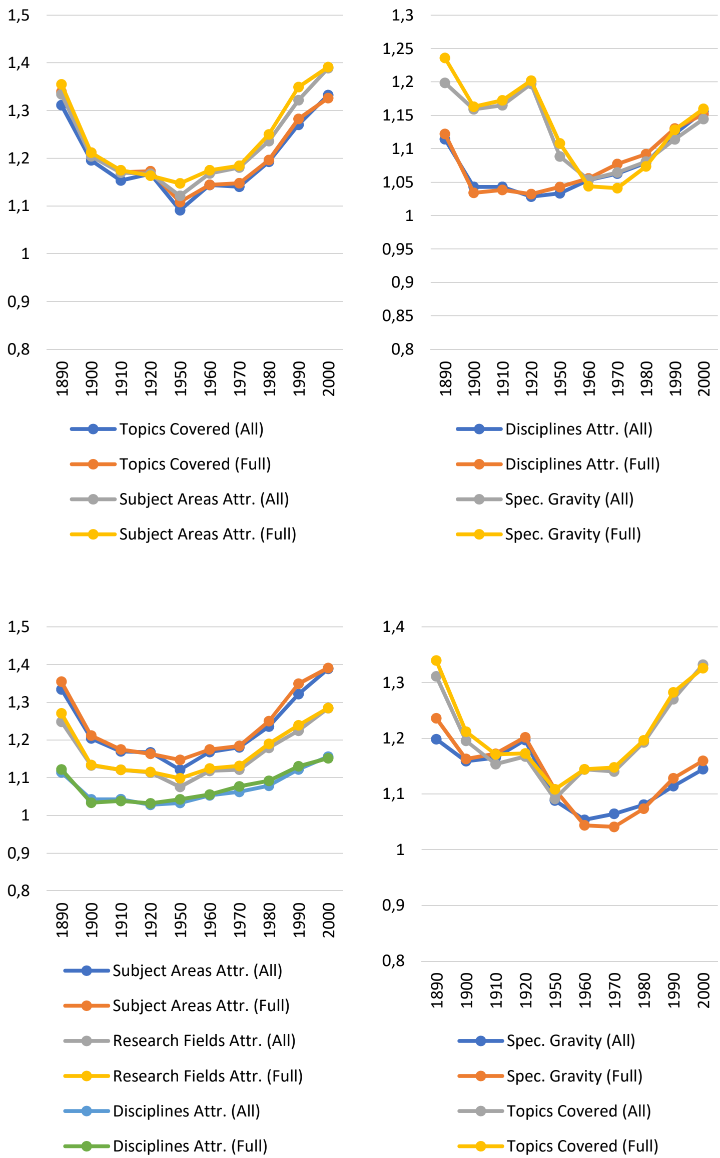
On the aggregated level of the location and for the rough division in pre- and post-war period, we can thus conclude for this sample that task division among professors is fully institutionalized in the pre-war period and still very differentiated in the post-war period for European institutions, whereas task division for the non-European institutions in the post-war period seems to be less institutionalized, supposedly more team- or department based.

In figure 5 the development of task coordination and specialization gravity is depicted for both full and sum of all professors. For all variables we find the trends to largely coincide for the latter, suggesting that the differences obtained in the development of numbers in between full, associate and assistant professors cannot be transferred to the division of labor and specialization variables.

In the upper left graph, we find that the number of topics and subject areas covered follows a U-shaped curve. Indeed, in the pre-war period the mean number of topics and subject areas covered by a denomination, drops from above 1.3 in 1890 to below 1.2 in 1920. Even though, the decades of 1930 and 1940 are missing because of ethical considerations as outlined in section 3.2.2, this decreasing trend nonetheless continues up until the 1950s before the variables start to increase substantially for the remaining post-war period, rising up above 1.3 again in 2000. Interestingly, in the post-war period we can obtain a moderate decoupling in the number of topics and subject areas covered, where in particular for the last decades individual topics are assignable to more subject areas on average. The latter can be explained by the increasing importance of departments, whose denomination might demarcate one topic which spans over multiple subject areas of a research field.

This idea is confirmed by the depiction of the different levels of granularity of the task coordination variable in the lower left graph. Indeed, by looking at the nearly continuously downwards shifted trend for subject areas covered when looking at the level of the research field

Fig. 5: Development of task coordination and specialization gravity variables for full and sum of professors over time



or discipline, two things can be noted. For one, if we were to project the number of topics covered into the lower left graph, we would find that while it develops parallelly with the subject areas in the pre-war period, it would in the post-war period lie in between subject areas and research fields attributable, confirming the idea of the department denomination's topics to be defined more general unifying multiple subject areas or larger parts of a research field. A second interesting finding here is that moving from the most granular levels of topics and subject areas to the aggregated level of the discipline, we can obtain that the information contained in the development of the trend seems to fade out or become less clear with increasing level of abstraction. The latter of course supports the idea that thinking about scientific coordination and cooperation in terms of interdisciplinarity might miss out on a lot of coordinative actions performed on more granular levels of interaction. Indeed, if we only considered the trend in development of disciplinarity of denominations for making a judgment on task coordination, we would probably conclude that (, after a steep move towards disciplinarity at the beginning of the 20th century) there is a moderately increasing trend towards more interdisciplinarity. Most certainly though, we would have missed the u-shaped trend obtainable on the more granular levels.

In the two graphs on the right-hand side, the trend of specialization gravity is plotted against the task coordination variables on level of the topics and on the level of the discipline. In general, the depth of an individual denomination's specialization is more volatile for the full sample when compared to the development of the task division or task coordination variables. In the pre-war period we can obtain high values for specialization depth, which seem to be relatively constant. In the post-war period, specialization depth seems to decrease until the 1960s before it stabilizes and increases again for the later decades. Interestingly, the trend for specialization depth falls in line with the trend of disciplines attributable in the post-war period and with the trend of topics covered in the pre-war period. The latter suggests that for this sample, in the post-war period we find that denominations demarcate slightly more 'interdisciplinary' yet less specialized research domains, whereas in the pre-war period they demarcate rather 'disciplinary' and highly specialized topics.

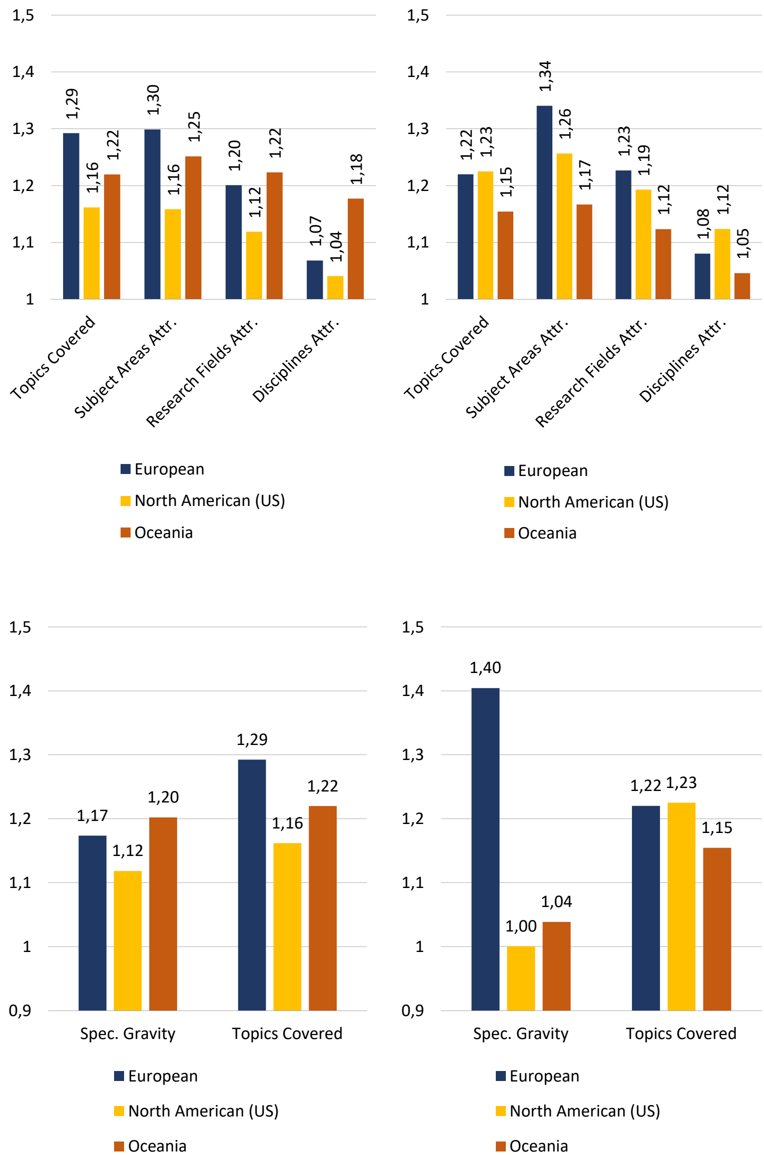
Clearly, this in opposition to the theoretically outlined idea of a lack of controlling coordination costs in scientific institutions. Indeed, task division and specialization gravity increase in the later sample periods, but task coordination equally increases on all different levels of granularity. This suggests that increased division of labor and specialization is kept in check by greater efforts of task coordination at least by the here considered universities. When looking at the trend over time for the full sample, a self-reinforcing effect of specialization gravity over time thus cannot be obtained.

In figure 6, the mean values for the task coordination and specialization gravity variables are given according to locations for pre-war and post-war science. In the pre-war science, the highest values for task coordination can be found among the European universities in the sample, which e.g., on average cover 1.3 topics and subject areas per denomination.

This is also reflected on the level of research field and discipline, where European institutions coordinate more when compared to US peers. While the denominations in the two Oceanian institutions cover less topics and subject areas, they coordinate slightly more research fields and substantially more in between disciplines than their European counterparts. When looking at the results for the post-war period, the structural differences found for pre-war and post-war period seem to be mainly caused by changes in US American institutions. While the European values for task coordination on the different levels of granularity remain relatively stable over the course of time, with a moderate decrease in number of topics covered and minor increases in subject areas and research fields attributable, we find substantial shifts in the patterns for non-European universities. Task coordination in US American institutions significantly increases on all aggregation levels, even surpassing the European institutions when it comes to connecting different disciplines in denominations. Oceanian universities on the other hand, operate at lower levels of task coordination in the post-war than in the pre-war period.

When looking at the lower graphs, where the specialization gravity variable is plotted against the topics covered, we see that this difference obtained in between European and US institutions in the post-war period does equally apply to specialization depth. On average in the post-war period the specialization depth of a denomination in non-Euro-

Fig. 6: Task coordination and specialization gravity according to location for pre-war (left) and post-war science (right)



pean institutions corresponds nearly exactly to the scope of a subject area. On the contrary, nearly every second denomination in a European institution demarcates a research topic that is narrower and fully nested in a subject area. When contrasted with the pre-war period, we can obtain that specialization depth of denominations in European institutions increased, whereas it decreased for US American and Oceanian universities.

The latter is particularly interesting, because task coordination mainly increased for US institutions yet nearly stayed the same for European institutions. Further, the task division variable suggests that European universities continued to differentiate tasks in the post-war period. Altogether, over time, European universities' denominations became more specialized in depth, continued to differentiate in tasks, while efforts for task coordination remained stable. Whereas we cannot find a lack of controlling coordination costs for the full sample, this might nonetheless be the case for the European institutions.

This picture is further confirmed by the opposite trends obtainable for the non-European institutions, in particular US universities. The latter are found with higher values for task division, which suggests less differentiation or institutionalization of tasks, moving towards organizing division of labor rather in departments or teams. In addition, the efforts for task coordination increased, whereas specialization in depth decreased, which suggests that coordination costs are kept in check over the here considered time period.

Finally, specialization concentration will be assessed by looking at the concentration of professors on different research domains. In table 6, a summary of the distribution of professors, DoL and Spec. variables is given according to disciplinarity and disciplines attributable (See Annex S1 for the full table, where information on numbers of individual professorial types, DoL and Spec. variables is provided according to time, location and institution). For the full sample, three-quarter of all professors is associated with one of the 1,086 disciplinary denominations. Overall, about 10 percent of sample professors work disciplinarily in the engineering sciences, 41 percent in the medical and life sciences and finally, about 23 percent in the natural sciences. About 10 percent of sample professors are associated with a denomination that is interdisciplinary, with the combinations of the disciplines originally considered being most pronounced making up about 5 percent of the pro-

Tab. 6: Nr. of denominations (DEN), share of sum of prof. (S_sum), sum of prof. per full prof. (Sum / Full), sum of prof. per denomination (sum / den), mean nr. of topics covered (TOPICS), ind. specialization (SPEC), subjects attributable (SUBJ), research fields attributable (FIELD) and instantiations (INS) in sample according to disciplinarity

	DEN	S_sum [%]	Sum / Full	Sum / Den	TOPICS	SPEC	SUBJ	FIELD	INST
Disciplinary	**1086**	**74.09**	**1.48**	**2.46**	**1.20**	**1.59**	**1.00**	**1.00**	**6.22**
Engineering Sciences	160	10.14	1.42	2.10	1.24	1.56	1.00	1.00	4.43
Medical and Life Sciences	629	41.12	1.60	2.50	1.21	1.57	1.00	1.00	6.37
Natural Sciences	297	22.82	1.27	2.58	1.16	1.63	1.00	1.00	6.87
Interdisciplinary	**451**	**9.57**	**1.31**	**2.71**	**1.77**	**1.60**	**2.30**	**2.18**	**2.24**
Engineering Sciences, Law, Arts and Humanities	2	0.01	1.00	1.50	1.50	1.00	2.00	2.00	1.50
Engineering Sciences, Medical and Life Sciences	19	0.09	1.51	1.49	1.84	1.52	2.21	2.21	1.37
Engineering Sciences, Medical and Life Sciences, Natural Sciences	6	0.03	1.00	1.08	2.50	1.75	3.33	3.33	1.83
Engineering Sciences, Multidisciplinary (miscellaneous)	3	1.45	1.51	17.66	2.00	1.17	2.00	2.00	10.00
Engineering Sciences, Multidisciplinary (miscellaneous), Natural Sciences	1	0.00	1.00	1.00	2.00	1.00	3.00	3.00	1.00
Engineering Sciences, Multidisciplinary (miscellaneous), Social and Behavioral Sciences	1	0.01	4.00	4.00	2.00	1.50	3.00	3.00	1.00
Engineering Sciences, Natural Sciences	97	3.73	1.18	3.61	1.79	1.41	2.16	2.10	2.18
Engineering Sciences, Natural Sciences, Social and Behavioral Sciences	1	0.00	1.00	1.00	2.00	2.00	3.00	3.00	1.00
Engineering Sciences, Social and Behavioral Sciences	6	0.17	1.26	6.58	1.83	1.42	2.17	2.00	1.50
Language, Information and Communication, Medical and Life Sciences	2	0.00	1.00	1.00	2.50	1.17	2.00	2.00	1.00
Law, Arts and Humanities, Medical and Life Sciences	25	0.15	1.14	1.06	1.96	1.63	2.36	2.24	2.16
Law, Arts and Humanities, Medical and Life Sciences, Natural Sciences	2	0.01	1.00	1.13	2.50	2.00	4.50	4.50	2.50

	DEN	S_sum [%]	Sum / Full	Sum / Den	TOPICS	SPEC	SUBJ	FIELD	INST
Law, Arts and Humanities, Multidisciplinary (miscellaneous), Medical and Life Sciences	1	0.00	2.00	1.00	2.00	1.00	3.00	3.00	2.00
Law, Arts and Humanities, Multidisciplinary (miscellaneous), Natural Sciences	1	0.00	1.00	1.00	3.00	2.00	3.00	3.00	1.00
Law, Arts and Humanities, Natural Sciences	5	0.07	1.15	1.25	1.60	1.80	2.40	2.00	3.60
Medical and Life Sciences, Multidisciplinary (miscellaneous)	23	0.20	1.27	2.26	1.91	1.40	2.13	2.09	1.52
Medical and Life Sciences, Multidisciplinary (miscellaneous), Social and Behav. Sci.	2	0.16	2.48	12.50	2.00	1.50	3.00	3.00	2.00
Medical and Life Sciences, Natural Sciences	160	2.33	1.34	2.15	1.68	1.77	2.33	2.18	2.38
Medical and Life Sciences, Natural Sciences, Social and Behavioral Sciences	1	0.05	1.14	12.50	2.00	2.00	3.00	3.00	2.00
Medical and Life Sciences, Social and Behavioral Sciences	55	0.72	1.51	3.54	1.71	1.61	2.27	2.11	2.02
Multidisciplinary (miscellaneous), Natural Sciences	13	0.10	1.51	2.12	1.92	1.67	2.15	2.08	1.46
Multidisciplinary (miscellaneous), Natural Sciences, Social and Behavioral Sciences	1	0.00	1.00	1.00	3.00	2.00	5.00	5.00	1.00
Multidisciplinary (miscellaneous), Social and Behavioral Sciences	5	0.04	1.36	1.23	1.60	1.10	2.40	2.00	2.60
Natural Sciences, Social and Behavioral Sciences	8	0.04	1.20	1.41	1.75	1.63	2.25	2.00	1.63
Multidisciplinary (miscellaneous)	**29**	**1.93**	**1.40**	**3.81**	**1.00**	**1.31**	**1.00**	**1.00**	**4.21**
Multidisciplinary (miscellaneous)	29	1.93	1.40	3.81	1.00	1.31	1.00	1.00	4.21
Pluridisciplinary	**983**	**14.41**	**1.34**	**2.47**	**1.68**	**1.65**	**2.15**	**1.69**	**2.32**
Engineering Sciences	50	0.93	1.43	3.93	2.06	1.39	2.08	1.74	1.50
Medical and Life Sciences	752	10.67	1.36	2.27	1.65	1.69	2.17	1.72	2.39
Natural Sciences	181	2.81	1.24	2.90	1.73	1.52	2.07	1.53	2.27
Sample	**2549**	**100.00**	**1.40**	**2.53**	**1.49**	**1.61**	**1.67**	**1.48**	**3.99**

fessors in the sample. Finally, about 14 percent of sample professors are linked to denominations categorized as pluridisciplinary. The latter category is mainly constituted by denominations, which connect different research fields in the medical and life sciences.

When looking at the ratio of sum of professors to full professors, we find that among the disciplinary denominations, medical and life sciences have higher shares of associate and assistant professors (1.60) followed by the engineering sciences (1.42). Among the pluridisciplinary denominations this is the other way around. In both cases, the natural sciences have the least associate and assistant professors relative to full professors in the sample. Regarding task division (sum of professors per denomination) we can obtain that disciplinary denominations seem to be more institutionalized (2.46) whereas inter- (2.71), multi- (3.81) and pluridisciplinary (2.47) denominations come along greater numbers of professors. The different variables employed to measure task coordination also vary across different disciplines and degree of disciplinarity. While the number of topics covered is relatively moderate and comparable for disciplinary denominations, it is absolutely higher for pluridisciplinary denominations and varies, with pluridisciplinary engineering sciences denominations covering more than two topics per denomination on average. Interestingly though, the latter cannot be transferred to the level of the subject areas covered, which are highest for the medical and life sciences (2.17). The number of subject areas coordinated lies close to two for inter- and pluridisciplinary combinations, which suggests that multi-, inter- or pluridisciplinarity above all means the coordination of two manageable subject areas (as opposed to e.g., integrating a broad range of different research domains). For the integration of different research fields differences between pluri- and interdisciplinary denominations are obtainable. When compared to interdisciplinary denominations, we find that pluridisciplinary denominations are more likely to integrate subject areas belonging to the same research field (as opposed to subject areas of different research fields within the same discipline). This suggests a tendency for collaboration due to proximity, which is reasonable when we assume that coordination costs in between subject areas within a research field should be lower than coordination costs of subject areas lying in different research fields even when they belong to the same discipline.

Interestingly, the specialization depth is relatively equal for disciplinary, inter- and pluridisciplinary denominations. Among the latter though moderate differences exist with pluridisciplinary denominations of the medical and life sciences demarcating the narrowest scope of research. Finally, the mean number of appearances of denominations of particular categories is given in the last column. As expected, strictly disciplinary denominations have the highest probability of appearing on multiple different occasions with about 6 repetitions, whereas on average the same inter- or pluridisciplinary denomination appears about twice in the whole sample. The latter could of course belong to the same institution at a different point in time. The highest uniqueness of denominations is to be found in engineering sciences, where a disciplinary or pluridisciplinary denomination is instantiated no more than 4 or 1.5 times in the full sample respectively.

In table 7, disciplinarity, disciplinary combination and research field is given for the largest denominations which account for half of all sample professors. In this subsample, the most concentrated denominations belong to the field 'Clinical Medicine (11.95%)'. Other densely concentrated denominations can be found in the research fields 'Physics and Materials Science (5.61%)', 'Chemistry and Chemical Engineering (4.94%)', 'Biomedical Sciences (4.49%)' as well as 'Biological Sciences (4.42%)'. We can further obtain that 'big' denominations are above all disciplinary with few interdisciplinary denominations, for which use cases of the connections between the research fields can easily be imagined. In the whole sample no pluridisciplinary denomination exists, which accounts for a share of professors large enough to belong to this subsample. Given that the share of professors assigned to pluridisciplinary denominations in the whole sample is higher than its interdisciplinary counterpart for example and considering that the pluridisciplinary denominations are mainly concentrated in the 'Medical and Life Sciences' discipline, this is a quite surprising finding. It certainly points at a concentration of resources in disciplinary denominations and a greater differentiation or diversity of resources when it comes to deviating (re-)combinations of neighboring subject areas. The slightly lower shares of full professors, as well as the higher ratio of sum of professors per denomination when compared to the ratio of full professor per denomination support this claim, confirming that associate and assistant positions are more likely to be allocated to 'big' denominations.

Tab. 7: DEN, S_all, S_full, R_sum / full, R_sum /den, R_full / den, TOPICS, SPEC and INST. for 50th percentile of sample professors according to disciplinarity, discipline and research field

	DEN	S_sum [%]	S_full [%]	R_sum / full	R_sum / den	R_full / den	TOPICS	SPEC	INST
Disciplinary	**27**	**44.76**	**41.64**	**1.87**	**9.49**	**5.09**	**1.04**	**0.96**	**83.96**
Engineering Sciences	**3**	**7.53**	**7.17**	**1.75**	**14.00**	**8.12**	**1.00**	**1.00**	**81.33**
Civil Engineering and Construction	1	1.68	1.42	1.93	10.43	5.39	1.00	1.00	74.00
Electrical Engineering and Telecommunication	1	2.85	2.93	1.59	14.83	9.32	1.00	1.00	88.00
Mechanical Engineering and Aerospace	1	3.00	2.82	1.74	16.74	9.65	1.00	1.00	82.00
Medical and Life Sciences	**17**	**22.72**	**18.65**	**2.04**	**9.25**	**4.43**	**1.06**	**1.00**	**75.71**
Basic Life Sciences	1	0.95	1.00	1.55	5.29	3.41	1.00	1.00	82.00
Biological Sciences	3	4.42	4.69	1.54	8.36	5.45	1.00	1.00	86.33
Biomedical Sciences	5	4.49	4.12	1.81	6.41	3.41	1.00	1.00	75.80
Clinical Medicine	7	11.95	8.37	2.31	11.38	4.83	1.14	1.00	77.00
Health Sciences	1	0.92	0.46	3.25	15.11	4.64	1.00	1.00	28.00
Natural Sciences	**7**	**14.51**	**15.82**	**1.51**	**8.14**	**5.40**	**1.00**	**0.86**	**105.14**
Astronomy and Astrophysics	1	1.14	1.41	1.33	5.28	3.98	1.00	1.00	99.00
Chemistry and Chemical Engineering	2	4.94	5.23	1.55	10.48	6.75	1.00	1.00	107.00
Earth Sciences and Technology	2	1.72	1.84	1.52	4.63	3.04	1.00	1.00	85.50
Environmental Sciences and Technology	1	1.11	1.13	1.60	5.77	3.60	1.00	1.00	88.00
Physics and Materials Science	1	5.61	6.21	1.48	15.67	10.62	1.00	0.00	164.00

	DEN	S_sum [%]	S_full [%]	R_sum / full	R_sum / den	R_full / den	TOPICS	SPEC	INST
Interdisciplinary	**6**	**4.39**	**4.66**	**1.55**	**29.33**	**18.95**	**2.00**	**0.83**	**13.50**
Engineering Sciences, Multidisciplinary (miscellaneous)	**2**	**1.44**	**1.44**	**1.64**	**25.24**	**15.31**	**2.00**	**0.75**	**14.00**
Electrical Engineering and Telecommunication, Multiple Fields	1	0.68	0.66	1.70	31.20	18.40	2.00	1.00	10.00
Mechanical Engineering and Aerospace, Multiple Fields	1	0.76	0.78	1.58	19.28	12.22	2.00	0.50	18.00
Engineering Sciences, Natural Sciences	**3**	**2.21**	**2.31**	**1.56**	**27.59**	**17.10**	**2.00**	**0.83**	**15.00**
Civil Engineering and Construction, Environmental Sciences and Technology	1	0.77	0.81	1.56	23.53	15.07	2.00	1.00	15.00
Electrical Engineering and Telecommunication, Computer Sciences	1	0.80	0.77	1.71	46.00	26.88	2.00	1.00	8.00
General and Industrial Engineering, Physics and Materials Science	1	0.63	0.73	1.41	13.23	9.36	2.00	0.50	22.00
Medical and Life Sciences, Natural Sciences	**1**	**0.75**	**0.91**	**1.35**	**42.75**	**31.75**	**2.00**	**1.00**	**8.00**
Basic Life Sciences, Chemistry and Chemical Engineering	1	0.75	0.91	1.35	42.75	31.75	2.00	1.00	8.00
Multidisciplinary (miscellaneous)	**1**	**0.92**	**0.80**	**1.87**	**28.00**	**15.00**	**1.00**	**0.00**	**15.00**
Multidisciplinary (miscellaneous)	**1**	**0.92**	**0.80**	**1.87**	**28.00**	**15.00**	**1.00**	**0.00**	**15.00**
Multiple Fields	1	0.92	0.80	1.87	28.00	15.00	1.00	0.00	15.00
Sample	**34**	**50.07**	**47.10**	**1.81**	**13.53**	**7.83**	**1.21**	**0.91**	**69.50**

Tab. 8: DEN, S_all, S_full, R_sum / full, R_sum /den, R_full / den, TOPICS, SPEC and INST. for 50th percentile of sample professors according to disciplinarity and (harmonized) denomination

	S_all [%]	**S_full [%]**	**R_all / full**	**R_all / den**	**R_full / den**	**INST**
Disciplinary	**44.76**	**41.64**	**1.87**	**9.49**	**5.09**	**83.96**
Anatomy	0.82	0.81	1.65	3.63	2.19	104.00
Anesthesiology	0.75	0.52	2.35	6.92	2.94	50.00
Astronomy	1.14	1.41	1.33	5.28	3.98	99.00
Biochemistry	0.95	1.00	1.55	5.29	3.41	82.00
Biology	2.22	2.40	1.51	14.56	9.61	70.00
Botany	0.85	0.94	1.48	4.37	2.96	89.00
Chemical Engineering	1.34	1.40	1.57	10.27	6.53	60.00
Chemistry	3.59	3.83	1.54	10.70	6.97	154.00
Civil Engineering	1.68	1.42	1.93	10.43	5.39	74.00
Electrical Engineering	2.85	2.93	1.59	14.83	9.32	88.00
Geography	1.11	1.13	1.60	5.77	3.60	88.00
Geology	1.10	1.18	1.53	4.55	2.97	111.00
Geophysics	0.62	0.67	1.51	4.72	3.12	60.00
Internal Medicine	1.88	1.26	2.45	16.94	6.90	51.00
Mechanical Engineering	3.00	2.82	1.74	16.74	9.65	82.00
Medicine	3.20	2.12	2.47	26.16	10.59	56.00
Nursing	0.92	0.46	3.25	15.11	4.64	28.00
Obstetrics and Gynecology	1.01	0.70	2.35	4.36	1.86	106.00
Pathology	1.26	1.02	2.02	7.40	3.65	78.00
Pediatrics	1.77	1.34	2.16	9.34	4.33	87.00
Pharmacology	0.78	0.78	1.63	4.77	2.93	75.00
Pharmacy	0.78	0.72	1.77	7.64	4.32	47.00
Physics	5.61	6.21	1.48	15.67	10.62	164.00
Physiology	1.06	1.08	1.60	4.35	2.71	112.00
Radiology	1.04	0.72	2.38	11.68	4.90	41.00
Surgery	2.07	1.41	2.40	8.55	3.57	111.00
Zoology	1.34	1.35	1.63	6.16	3.79	100.00
Interdisciplinary	**4.39**	**4.66**	**1.55**	**29.33**	**18.95**	**13.50**
Aeronautics & Astronautics	0.76	0.78	1.58	19.28	12.22	18.00
Chemistry & Biochemistry	0.75	0.91	1.35	42.75	31.75	8.00
Civil and Environmental Engineering	0.77	0.81	1.56	23.53	15.07	15.00
Computer Science & Engineering	0.68	0.66	1.70	31.20	18.40	10.00
Electrical and Computer Engineering	0.80	0.77	1.71	46.00	26.88	8.00
Materials Science & Materials Engin.	0.63	0.73	1.41	13.23	9.36	22.00
Multidisciplinary (miscellaneous)	**0.92**	**0.80**	**1.87**	**28.00**	**15.00**	**15.00**
Biological Sciences	0.92	0.80	1.87	28.00	15.00	15.00
Sample	**50.07**	**47.10**	**1.81**	**13.53**	**7.83**	**69.50**

When taking a look at table 8, where the actual denominations are listed that account for 50 percent of the share of professors, it becomes clear that those ‘big’ denominations largely correspond to the layman’s understanding of natural sciences’ disciplines or what could be regarded as a typical way to define school subjects (e.g., ‘Biology’, ‘Chemistry’ and ‘Physics’). The latter is not surprising, given that in earlier sample periods, task division meant exactly the differentiation in those subjects. In addition, the latter today still sometimes serve as denominations for larger departments, demarcating for example the scope of PhD student programs.

Overall, half of the sample professors are concentrated on 34 denominations, corresponding to only 1.33% of all semantically different denominations documented. The latter are instantiated 2,363 times for different institutions and points in time, making up for about a quarter of all entries (10,167). The biggest denominations in the sample are ‘Physics (5.61%)’, ‘Chemistry (3.59%)’ and ‘Medicine (3.20%)’. When taken together ‘Medicine’ and ‘Internal Medicine’ would even rank second, accounting for 5.08 percent of sample professors.[19]

This table also illustrates quite well, why the dataset might on the aggregated level bear information on the organizational structure of universities (e.g., department or professorial chair organization) yet the individual denomination exclusively contains information on the institutionalized pursuit of a particular research domain. During the documentation process outlined in section 3.2.1, all of the here listed disciplinary denominations appeared in earlier and later periods as denomination of professors assigned a professorial chair or affiliated to a department. This can be retraced by looking at the high absolute number of instantiations in conjunction with the mixed ratios of sum of professors per denomination. When turning to the denominations categorized as interdisciplinary[20], big enough to appear in this sub-

[19] ‘Internal Medicine’ and ‘Medicine’ were not reduced to a single denomination due to simultaneous appearances in institutions at one point in time, signaling a different scope of research.

[20] When coding the denominations according to the WoS scheme, ‘Electrical Engineering and Computer Engineering’ for example was attributable to the research field ‘Electrical Engineering and Telecommunications’ as well as ‘Computer Sciences’, with the former belonging to the discipline ‘Engineering Sciences’ and the latter to ‘Natural Sciences’. Consequently, the denomination is catego-

sample, we might suspect that those are mainly organized in large (engineering) departments since the number of instantiations is low and the ratio of sum of professors per denomination is (, except for 'Materials Science and Materials Engineering') comparatively high.

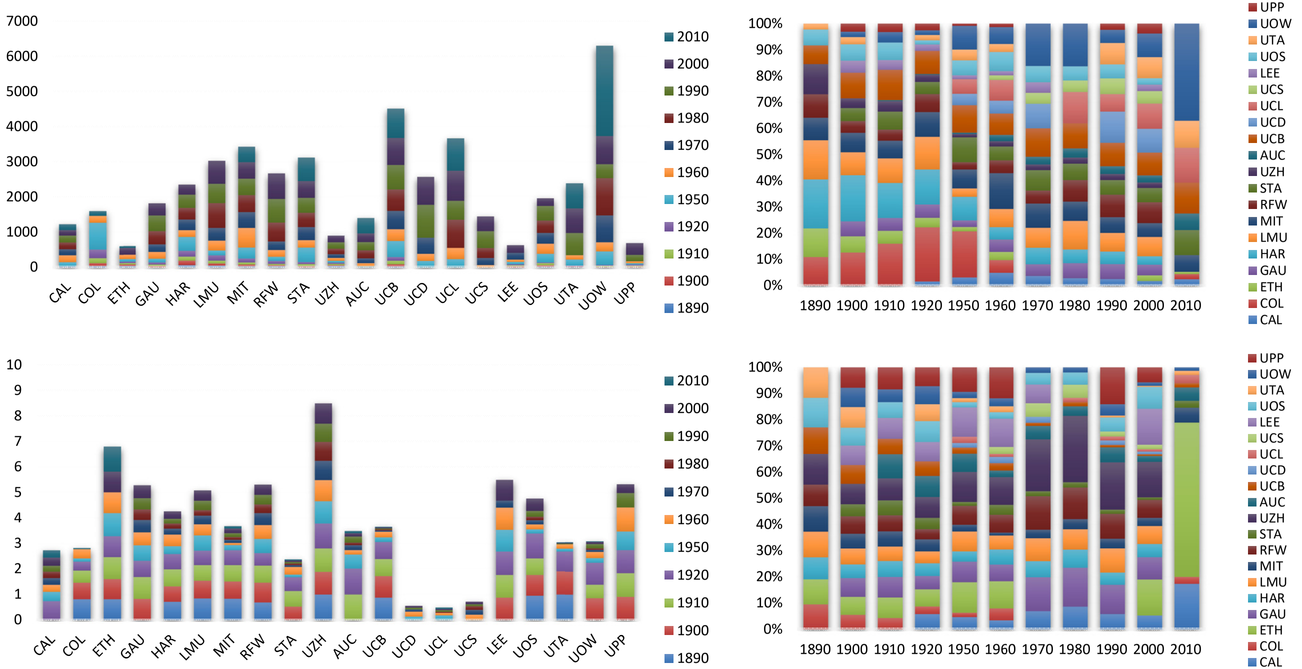
Since, given a denomination, it is not possible to differentiate a priori or necessarily between department or chair system (especially when shares of professorial types are not available), the denominations reveal a sort-of continuous quality, making them a suitable tool for comparisons of universities with different organizational designs.

3.3.1.2 Institution-specific (zooming in)

In this section, the general trends according to locations will be complemented by looking at the institution-specific variables. In figure 7, the number of denominations and professors per denomination are given as absolute numbers as well as relative shares for each institution according to time period. At first glance, we are able to attribute a large portion of the size differences obtained between US institutions and the rest of the sample institutions to the state university systems of the University of California and the University of Washington at Seattle. The accumulated sum of professors is comparatively high for the latter, driven above all by the observation for 2010. The latter also does not fit the institution's relative development over the course of time, potentially qualifying as an outlier observation (This will be addressed in context of the data employed in the quantitative empirical analysis in section 4.2.1.3). Interestingly, the size difference for US and European institutions does mainly concern the latter state universities and the non-German European institutions. The German universities and the other US institutions seem to be quite comparable in absolute numbers of professors accumulated over the here considered period.

rized as interdisciplinary regardless of a certain intuitively presumed proximity of the two topics covered by the denomination.

Fig. 7: Institution-specific absolute number (left) and relative shares (right) of sum of professors (upper) and professors per denom. (lower)

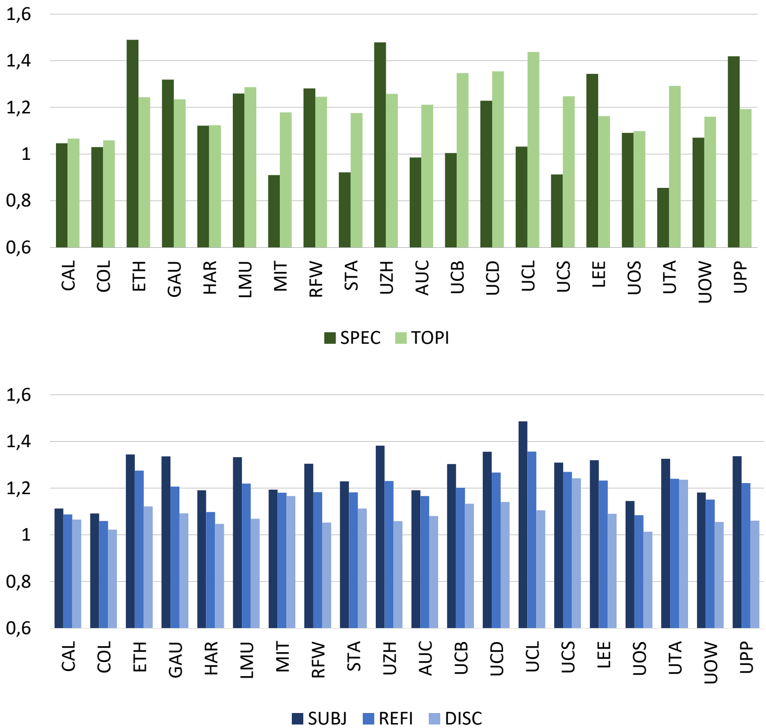


When looking at the relative shares according to decade, we find that those institutions maintain fairly equal shares in each period over the whole time considered, whereas the growth of the large state universities at first sets in in the post-war period. Overall, even though this could not have been influenced anyway, the panel of distributions of professors according to institutions is quite balanced with the notable exception of the decade 2010 to 2020, for which as previously noted no source for European institutions were available. Here, in particular the share of the Columbia University is odd given its huge proportion in the pre-war period. Certainly, the observation of the Columbia University in 2010 also qualifies as potential outlier observation, which will later be addressed.

In the lower panels, institution-specific development of denominations per professor (, or inverted task division) is provided. Here we can clearly differentiate in between pre-war and post-war period, as well as rediscover the substantial differences in between European and US American institutions. Apart from the California Institute of Technology and the Harvard University the values for inverted task division approximate 0 for US American institutions in the post-war period, contrasting sharply with the accumulation of values analogous to their European peers in the pre-war period. Interestingly, when looking at the institution-specific data we can obtain that the German institutions are subject to a scaled-down version of reduction in denominations per professors, whereas the other European institutions (e.g., the UZH) reveal a constant steadily high differentiation in all periods considered.

In figure 8, institutional means for task coordination and specialization gravity variable are given. Here, the mean values were calculated employing the variables for each denomination weighted by its professorial share relative to the number of professors of the overall institution. As opposed to the discussion of the institution-specific (inverted) task division variable, the differences between European and US American institutions in degree of specialization depth seem to hold for all institutions. Over the whole period considered, the European universities dominate their US peers, with the University of California – Davis being the only one that comes close to the least specialized European with the LMU Munich.

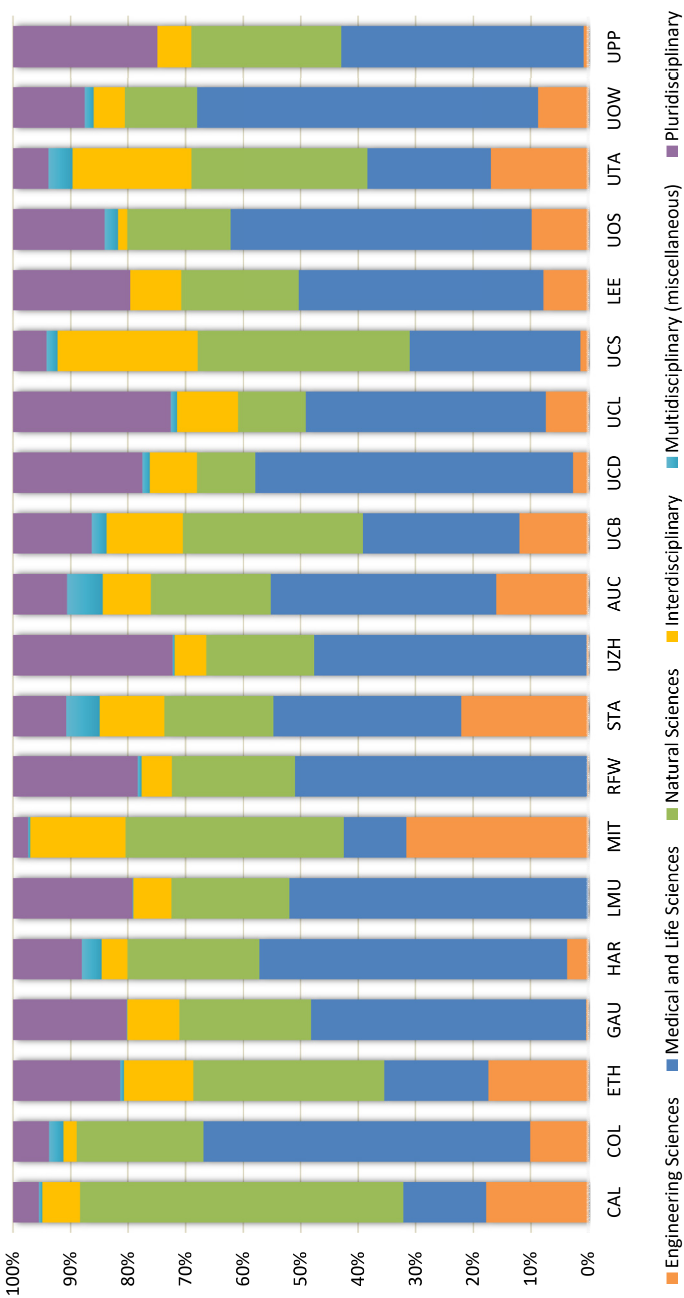
Fig. 8: Institution-specific mean values for task coordination and specialization gravity



The results for the task coordination variable in turn seem to suggest that the gap in between subject areas, research fields and disciplines coordinated is not strictly divided in European and US institutions. The Harvard University and the institutions of the University of California (, except for the UCS) for example, reveal higher values for number of topics and seem to coordinate research rather on the level of the subject area, whereas other US peers like e.g., the Caltech or the Columbia university seem to coordinate research equally on the different levels of granularity.

In figure 9, the institution-specific shares of disciplines, inter-, multi- and pluridisciplinary denominations are given. First, it needs to be emphasized that the disciplinary profiles are varying substantially on

Fig. 9: Institution-specific shares of disciplines (disciplinary) and shares of inter-, multi- and pluridisciplinary denominations

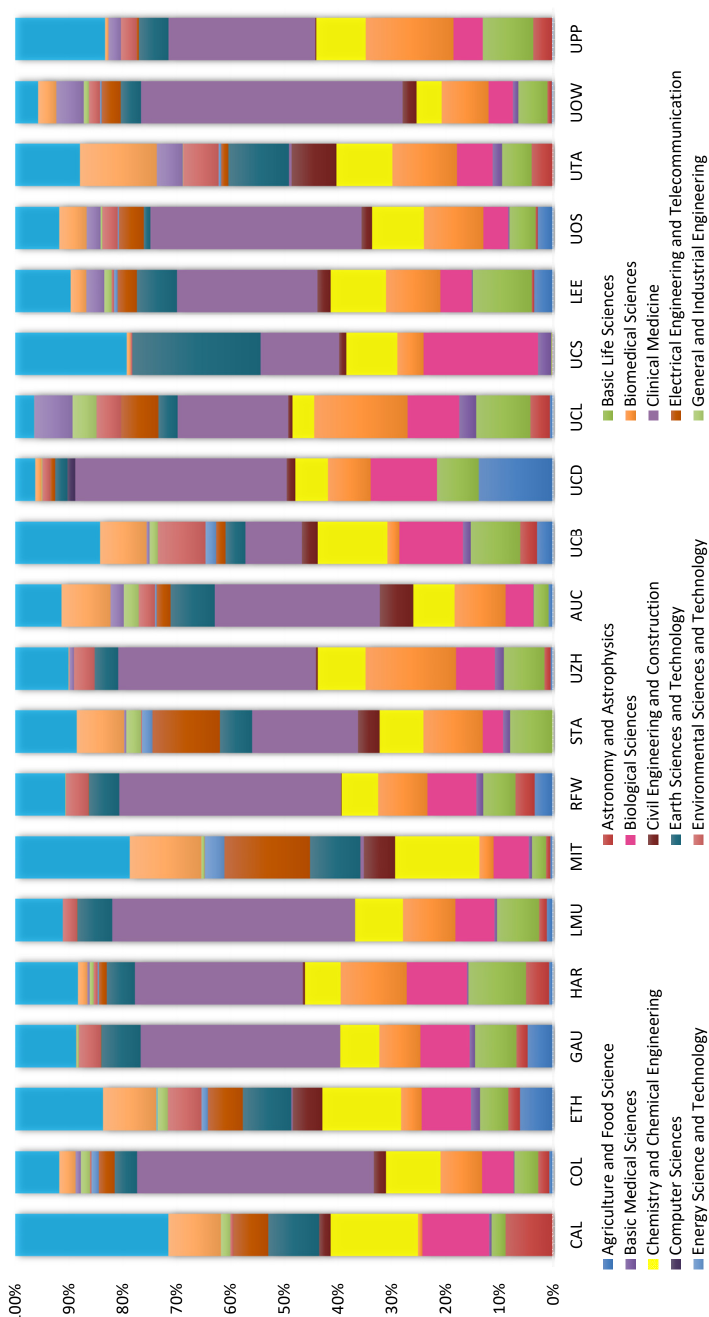


the level of the individual institution. If indeed, performance measurement of universities, e.g., university rankings, were to incorporate institutional data in the future, this needs to be considered. One interesting finding here is that the German institutions and the Swiss university of Zurich have no significant share of disciplinary denominations attributable to the engineering sciences. For the LMU Munich and the University of Zurich, this could be explained by the very renowned Technical University of Munich and ETH Zurich being located nearby. Indeed, the latter of course being a part of the sample here as well, reveals a comparatively high concentration in engineering sciences. Supposedly, a scaled-down version of this case could be made for the Harvard University and the MIT as well, with the former having a relatively low share of engineering sciences when compared to US peers and the latter having the strongest concentration on engineering in the sample.

Most universities in the sample have their strongest concentration in the medical and life sciences. The remaining institutions, among other the polytechnic universities are dominated by denominations attributable to the natural sciences. Regardless of the composition of the disciplinary denominations we find a certain corridor of disciplinarity ranging from 65 to 85 percent with only few institutions deviating with higher (e.g., Caltech) or lower (e.g., UCL) shares of disciplinary denominations. The shares of multi-, pluri- and interdisciplinarity are unevenly distributed though, pointing at a substitution effect of coordinative denominations on institutional level.

In figure 10, the institution-specific shares of research fields attributable to disciplinary denominations is given. Here we can obtain that the universities with a large share of denominations belonging to the Medical and Life Sciences nearly all have a large share of denominations in the field ‘Clinical Medicine’. Interestingly, smaller to medium shares of denominations in the Medical and Life Sciences as found for all polytechnic universities and the University of Texas at Austin come along no denomination in the field of ‘Clinical Medicine’. Obviously, this points at a clinic affiliation being an important determinant of the share of denominations in the Medical and Life Sciences and also predicts the size of the shares of denominations dedicated to the fields ‘Biomedical Sciences’ and ‘Basic Medical Sciences’ for example.

Fig. 10: Institution-specific shares of research fields attributable to disciplines (disciplinary)



When looking at the disciplinary profiles, polytechnic universities seemed to be quite comparable to one another and differentiable from their all-sciences counterparts. Here, on the level of the research field though, a more nuanced picture is provided. Indeed, polytechnic universities seem to have pronounced concentrations of professorial resources on varying fields. The Caltech for example has a large share of professors in the field of 'Astronomy and Astrophysics', for which the shares of the ETH and MIT rather fall in line with their all-sciences counterparts. The ETH has a large concentration in 'Agriculture and Food Science', while the MIT specializes more in 'Mechanical Engineering' and 'Electrical Engineering' when compared to its technical peers. On the contrary, common denominators of the technical institutions are relatively large shares in the fields 'Physics and Materials Science' as well as 'Chemistry and Chemical Engineering'.

Further, while the disciplinary profiles of the institutions of the University of California (with the exception of the UCS) seemed roughly comparable to other all-sciences peers, their research field profile suggests a significantly larger heterogeneity. The UCB for example shows a quite equal distribution of professors according to fields, whereas the UCD exhibits an emphasis on 'Agriculture and Food Science' and the UCL on 'Biomedical Sciences'. The UCS in turn, seems to deviate significantly from the other institutions of the University of California with particularly pronounced concentrations in 'Biological Sciences' and 'Earth Sciences and Technology'. When looking at the individual denominations of the UCS we find that the denominations contributing to the concentration in 'Biological Sciences' are e.g. 'Marine Biology and Geology', 'Marine Geophysics' and 'Biological Oceanography' or 'Physical Oceanography'. A quick look at the history of the University of California – San Diego reveals that the institution was founded near the preexisting renowned Scripps institution of Oceanography. Broadly speaking, the enduring concentration of professors on those research fields over the course of the whole period is explained by institutional prerequisites. This points at a certain profanity of specialization concentration processes, confirming the idea that concentration on a research domain relies on factors linked to interests, talent and opportunities, given certain institutional prerequisites, already highlighted as major argument against the idea of an efficient cognitive division of labor.[21]

3.3.2 *Identifying university types (cluster analysis)*

The descriptive analysis of the previous sections will be complemented by a correlation analysis to provide statistical evidence for whether DoL and Spec. indeed create path dependencies over the course of time. Next to the institution-specific differences highlighted, the descriptive analysis also suggests structural differences according to size and location, but also in disciplinary profiles, pointing at an existence of certain institutional types, which are not sufficiently accounted for by the broad categories of e.g., 'all sciences vs. polytechnic university'.

Thus, before the correlation analysis is performed, a hierarchical cluster analysis is employed to segregate university types. This may as well be considered as a form of cross validating the remarks made in the descriptive analysis. Further, in case the identified clusters reflect to some extent our theoretical expectations, it also affirms the idea of the *denomination hypothesis* since the DoL and Spec. variables at least convey information we deem to be meaningful given our theoretical understanding of institutional differences.

Before the institutions can be clustered according to DoL and Spec., task coordination and specialization concentration need to be reduced to single indicators. In addition, the information of specialization con-

[21] Indeed, during the documentation process a lot of such allegedly profane institutional constraints appeared to be the reason for concentration in certain research domains. The LMU Munich for example had a considerable number of denominations linked to the subject area forestry in the sample periods 1890 to 1980, which suddenly disappeared in the last two periods. The latter provoked brief research on the issue, which revealed that the LMU's professors initially chose to specialize in forestry, because Duke Maximilian Joseph I gifted the Bavarian state forest to the university. Because of the institutional requirement to manage the forest, a concentration in this subject area was established. When the Technical University Munich took over those management tasks and the forest was transferred to the latter in 1999, the LMU's concentration on the subject area forestry disappears. (See https://www.lmu.de/de/die-lmu/foerdern-und-unterstuetzen/stiftungen-lmu/leistet/der-universitaetswald/historie/index.html)

centration, which was mainly explored graphically in the descriptive analysis, now needs to be operationalized quantitatively. To achieve the latter, we employ the Herfindahl-index (, alternatively known as Hirschman-Herfindahl or the reciprocal of the Simpson-diversity index) to quantify the absolute concentration of professorial shares according to the different levels of granularity (Tabner 2007). The concentration measure is defined as (Tabner 2007):

$$C_H = \sum_{i=1}^{n} c_i^2 , \quad (1)$$

with $c_i = P_i / \Sigma_i^n P_i$, where P_i stands for the number of professors sharing a common denomination, subject area, research field or discipline and $\Sigma_i^n P_i$ is the sum of professors for an institution in a certain period. In case of maximum concentration, the Herfindahl-index takes a value of 1. (Tabner 2007)

Here, a single-dimensional, absolute concentration measure is employed instead of using e.g., a Rao-Stirling diversity or Gini index,[22] because it suits the theoretical motivation of the specialization concentration variable best. Given two institutions, which have equally large absolute shares of professors devoted to e.g., two research fields with one of the institutions also having a marginal share of professors associated with a third research field, the Gini index would return substantially different values, whereas the Herfindahl index would basically remain the same for the two universities. The latter is certainly the more desirable property for quantifying the specialization concentration component, since what we are trying to understand here is whether an institution concentrates significant absolute resources on selected e.g., denominations, not how equally the professors at a university are

[22] In bibliometric analysis different indices have been employed to measure (publication) specialization. Lopez-Illescas et al. (2011) and Daraio et al. (2015) used a Gini index to categorize whole institutions as generalist vs. specialist by looking at the disciplinary specialization based on the equality of distribution of publications across a finite number of disciplines. Rafols and Meyer (2010) and Leydesdorff and Rafols (2011) used a multi-component indicator with the Rao-Stirling-index to measure 'Interdisciplinarity' of publications. Moschini et al. (2020) applied the Herfindahl-index to measure 'Multidisciplinarity' of publications based on categorization according to subject areas.

distributed across all research domains. Also, by design employing the Gini index on the most granular level of the denomination would require accounting for institutional shares of all 2,549 denominations. In case denominations with no share were excluded, the number of denominations of an institution would influence the results and consequently an undesirable logical connection to the task division variable would be introduced.

Prior to performing the cluster analysis, the dimension of the set of variables will be reduced by performing a factor analysis. By design and judging from the trends obtained in the descriptive analysis, the task coordination and specialization concentration variables should be strongly correlated on the different levels of granularity. To maintain the differing information conveyed on the different levels of granularity, without assigning disproportionately high weights to the task coordination and specialization concentration components in cluster, correlation and efficiency analysis, factor loading seems to be preferable over preselecting individual variables e.g., on the most granular or most aggregated level or employing all variables.

The segregation of the university types in clusters should be based on time-independent information to identify characteristics that are institution-inherent and not part of general trends within science or the scientific community. Consequently, the values for all variables are aggregated to the institutional level for the whole period, by weighting all values of denominations assigned to an institution i with the number of assigned professors relative to the sum of professors associated with the institution over the whole sample period.

Ideally, the results of the factor analysis confirm that the variables of the task coordination and specialization concentration component are suited for factor loading, whereas the individual components in which division of labor and specialization were divided based on the theoretical remarks in section 2.2.2 should be graded as not suitable for factor loading (, or else one could argue in favor of one indicator for e.g., division of labor incorporating both the information for task division and coordination). Indeed, for division of labor the latter can be confirmed. The Kaiser-Meyer-Olkin criterion suggests a value of 0.64 for overall sampling adequacy of all division of labor variables with a value of 0.52 for task division. When task coordination is considered exclusively, the overall value slightly increases (0.65), even though the individual values suggest excluding disciplinary task coordination (0.46). Thus, we conclude that a vertical integration of the different levels of granularity of task coordination into a factor variable is sound, yet we refrain from constructing a joint division of labor factor, given the low KMO criterion when simultaneously accounting for task division.

When limited to task coordination on the level of the denomination, subject area and research field, the KMO increases to a reasonable value of 0.76 with highly significant positive correlations of the three task coordination variables (0.86*** for topics and subject areas, 0.87*** for topics and fields and 0.93*** for subject areas and fields). In table 9 and 10, the results of the factor loading procedure for task coordination are given.

Tab. 9: (Unrotated) principal component analysis for the task coordination variables

Obs. = 20	Eigenvalue	Difference	Proportion	Cumulative
Factor 1	2.621	2.656	1.034	1.034
Factor 2	-0.035	0.016	-0.014	1.020
Factor 3	-0.050	.	-0.020	1

Tab. 10: KMO-criterion, factor loadings, unique variances and predicted scores for the task coordination factor

Obs. = 20	KMO (0.761)	Rotated Factor Loadings (F1)	Unique Variances	Scoring coeff.
Task Coord. - Topics	0.837	0.894	0.201	0.172
Task Coord. – Subject Areas	0.956	0.956	0.086	0.434
Task Coord. – Research Field	0.953	0.953	0.092	0.400

For the two specialization variables a different picture emerges. In general, the values for the KMO-criterion are lower than for the task coordination components. The configuration with all components yields a KMO value of only 0.57 with the individual value for research fields lying only at 0.33, which suggests excluding the concentration measure on field level. If focusing only on specialization concentration variables, the KMO lies even lower at 0.55. In the latter case though, the value for the research field slightly increases to 0.42. Also, while the concentration measures on level of topics and subject areas correlate positively (0.97***), the latter correlate negatively with the two concentration measures on the aggregate level (-0.04 for subject areas with fields, -0.51** with disciplines) and the specialization gravity variable (-0.77***). In conjunction with the low value of the individual KMO criterion of the concentration measure for research fields, the negative correlation of denominations and subject areas with research fields and disciplines could indicate that there is a shift (or potentially even a trade-off) in concentration on granular versus aggregated levels.

Also, here on the aggregated institutional level, the KMO criterion would not advise against constructing a single factor variable for specialization. On the contrary, the construction of the specialization gravity variable is theoretically independent from the numbers of professors concentrated on disciplines and its operationalization with the Herfindahl-index. In addition, the low individual KMO value for research fields indicates that here on the aggregated institutional level, the information conveyed on the aggregated levels of fields and disciplines should be interpreted carefully.[23] In accordance with the procedure employed for the factor loading of the two DoL components, we thus decided against loading the two specialization components into a joint factor. In table 11 and 12, the results of the factor loading procedure for specialization concentration are given.

[23] Indeed, when the factor analysis is performed for time-specific observations according to institutions (n = 169) for the efficiency analysis in section 4.2.2, the specialization concentration factor can be based on the three most granular levels of topics, subject areas and fields just like the task coordination factor. That both factors can be based on the three most granular levels provides a more coherent picture and points at a lower reliability of the results on the aggregated institutional level where only one observation for each institution (n = 20) is available.

Tab. 11: (Unrotated) principal component analysis for specialization concentration variables

Obs. = 20	Eigenvalue	Difference	Proportion	Cumulative
Factor 1	2.229	2.152	0.985	0.985
Factor 2	0.077	0.121	0.034	1.019
Factor 3	-0.044	.	-0.019	1

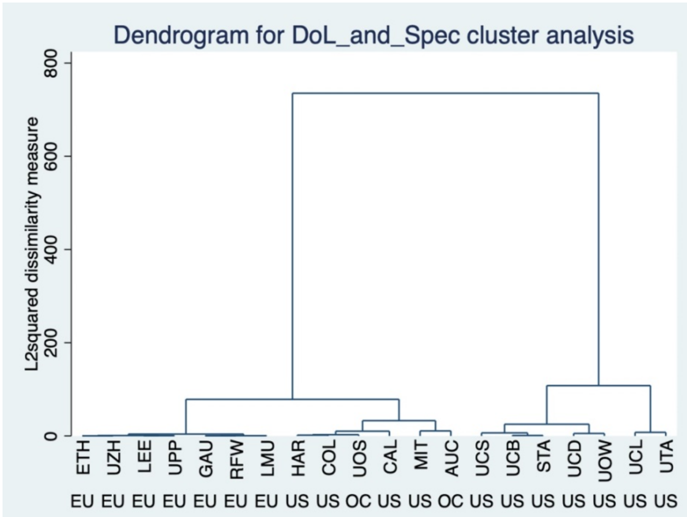
Tab. 12: KMO-criterion, factor loadings, unique variances and predicted scores the specialization concentration factor

		Rotated Factor Loadings			Scoring coeff.	
Obs. = 20	KMO (0.5725)	(F1)	(F2)	Unique Var.	(F1)	(F2)
Topics	0.541	0.941	0.292	0.029	0.123	2.331
Subject Areas	0.549	0.967	0.135	0.046	0.863	-2.202
Disciplines	0.728	-0.480	-0.389	0.619	0.029	-0.161

In figure 11, the dendrogram of the wards-linkage hierarchical cluster analysis[24] employing the task division and specialization gravity variable as well as the factors and remaining individual variables for task coordination and specialization concentration is given. Judging based on the dissimilarity measure the configurations of DoL and Spec. of the sample institutions may reasonably be segregated in two to four different clusters.

[24] Employing a hierarchical cluster analysis instead of e.g., a k-means based approach is a natural choice here, given the low number of observations enabling a meaningful and easy to interpret graphical inspection of proximities with a dendrogram. Since there are no special requirements for employing a particular distance measure, wards-linkage was chosen in accordance with the majority of applications in the literature.

Fig. 11: Dendrogram of hierarchical cluster analysis according to DoL and Spec. variables



Indeed, some theoretically expected differences as well as the observations of the descriptive analysis are affirmed by the cluster analysis results. Assuming two clusters, we can segregate the US state university systems from the rest of the sample. Supposedly, the differences in DoL and Spec. are here also linked to differences in size. Interestingly, the Stanford university belongs to this cluster as well, nested in its Californian peers of the University of California system. On the contrary, potential similarities among Californian institutions should not be overestimated in importance since the Caltech belongs to a different cluster. Generally, though, a clear distance of US public institutions to US private and non-US institutions can be obtained.

If we were to define four clusters, the two resulting clusters on the left-hand side of the dendrogram affirm the idea of structural differences between European and non-European institutions. Indeed, the dissimilarity measure reveals the lowest dissimilarity for all seven European universities when compared to the dissimilarities obtainable within the other three clusters constituted by non-European institutions. The expected differences within the European institutions are reflected by the order within the cluster, where e.g., the Swiss and German institutions are grouped. Yet in comparison with the non-European institutions, the dissimilarity of the latter seems to be insignificant suggesting a surprisingly high homogeneity of European institutions. This is an interesting finding, because following the literature one would intuitively suspect the dissimilarity in DoL and Spec. to be higher in between the comparatively smaller ETH as a polytechnic university and the LMU as a larger all-sciences university than the dissimilarity between the former with other US technical institutions. Further, the universities in the European cluster belong to four different countries, where in particular the University of Leeds as UK institution was expected to reveal a greater proximity to Oceanian and US American institutions than its Continental European peers.

Even though the dissimilarities in between the institutions of the second cluster are higher, their order and configuration fall in line with theoretical expectations. First, we find that the two Oceanian institutions are grouped together with the private Ivy's, suggesting a certain proximity towards the US system (, which could be explained by the historical connection to the British empire or a certain proximity and greater exchange of professorial staff due to the English language), which could not be obtained for the University of Leeds. Secondly, the two technical institutions of the US, the Caltech and the MIT show a certain expected proximity. And third, the longest established US institutions with the Harvard University and the Columbia university are positioned on the left end of the cluster revealing the least distance to the European cluster. The latter coincides with the theoretical information retrieved from the Minerva publication, which highlighted a close orientation of the latter two institutions towards the German university system in the late 19^{th} and early 20^{th} century, with the Columbia university having strong ties to the Humboldt university in Berlin for example.

In figure 12, dendrograms for wards-linkage based hierarchical cluster analysis for DoL and Spec. separately, as well as box-whisker plots for variables and factors are given. The clear segregation of European, private Ivy's and public state universities seems to be mainly driven by differences in task division and task coordination. Some of the patterns, like a close proximity of the European institutions can be rediscovered for specialization as well. Yet here there are also some deviations observ-

Fig. 12a: Dendrogram for DoL cluster analysis and box-whisker plots for DoL components and factor variables

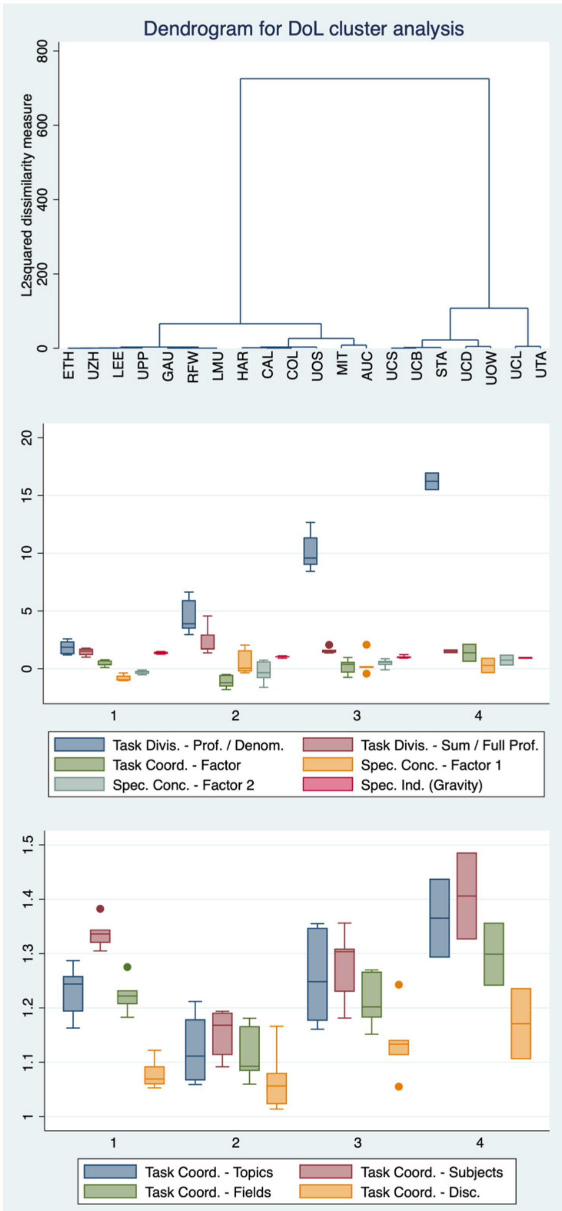
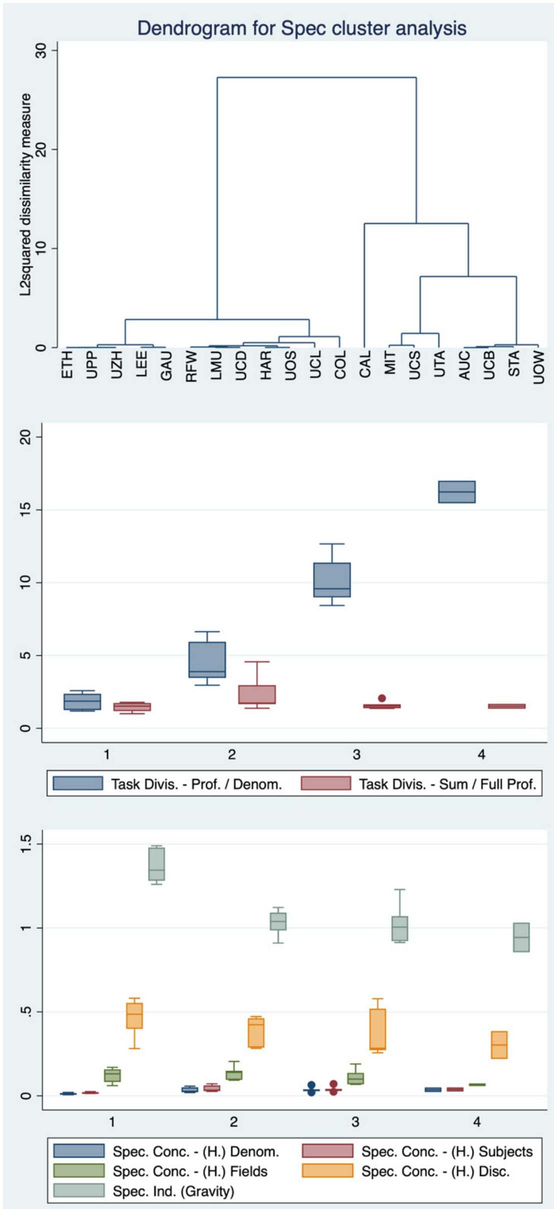


Fig. 12b: Dendrogram for Spec. cluster analysis and box-whisker plots for Spec. components and factor variables



able. Indeed, we find a greater proximity of two German institutions (RFW and LMU) with the longer established private US universities. Further, the results for the cluster analysis focused exclusively on the specialization components reveals that apart from the ETH, the technical institutions show a certain individuality with the MIT being part of a small cluster with only two other institutions and the Caltech even constituting its own cluster.

The box whisker plots allow to identify some of the major differences in patterns according to clusters. First, there is a dominating difference in task division, which is ranging from approximating full institutionalization (=1) in cluster 1 to large supposedly departmentally organized task division in cluster 4. More nuanced differences can be obtained for the composition of the task coordination component, where we see very homogenous narrow, yet substantially spread apart corridors for task coordination on the different levels of granularity for cluster 1. For the remaining clusters in turn the corridors are more heterogenous. Also, cluster 2 reveals significantly lower overall values for task coordination. When turning to the configuration of the specialization components, cluster 1 shows a high specialization depth as well as the highest disciplinary concentration. Clusters 2 and 3 in turn have lower specialization depth yet higher concentration on the level of topics, subject area and research fields.

Overall, there seems to be a certain continuity in task division and specialization gravity values shifting from cluster 1 (high task differentiation, high specialization depth) to cluster 4 (low task differentiation, low specialization depth). Apart from these continuous differences, structural differences exist for task coordination and specialization concentration. Cluster 2 reveals low overall task coordination, yet higher specialization concentration in topics, subject areas and research fields. While the average task coordination seems comparable for cluster 1, 3 and 4, the former shows very narrow homogenous corridors on the different levels of granularity. The latter homogeneity found in the European cluster (1) could be explained by the organizational difference in between European and US American universities with the introduction of the department system as highlighted by Parsons and Platt (1990). In any case, differences between European and other sample institutions are not limited to specific components, but much rather exist for all aspects of DoL and Spec. analyzed.

3.3.3 Identifying path dependencies of DoL and Spec. (correlation analysis)

Given the segregated institutional types, it can be examined if DoL and Spec. create path dependencies within institutions over the course of time. From the theoretical remarks on institutional DoL and Spec. we would expect such path dependencies to arise, since the gravitational force of specialization should promote further task division in dominant areas of research, leading to more concentration on the latter, which then should again enable more specialization in depth. If this mechanism is applicable to the scientific production process, we should observe that initial configurations of DoL and Spec. (e.g., a high (low) share of concentration in engineering denominations) predict today's configurations of DoL and Spec. (e.g., an even higher (lower) share of concentration in engineering). From a methodological point of view this may be operationalized by performing a simple correlation analysis. Pairwise Pearson correlation coefficients are calculated for all variables DoL and Spec. has been operationalized in, for the first and last observation in the sample (FL), as well as pre-war and post-war mean values (PP) of an institution. This procedure will also be applied to the relative shares of professors according to disciplinarity, which is particularly suited to analyze the specialization concentration component in detail, enabling us to identify disciplines (and forms of disciplinarity), which are particularly affected by path dependencies.

If significant, positive correlations are found, it is concluded that institutional DoL and Spec. are applicable to the scientific production process. This is an important requirement for establishing institutional DoL and Spec. as determinants of epistemic outcomes. Given that for some institutions first and last observation lie 120 years apart and considering that only 20 observations at maximum exist for all pairs, stating only significant correlations as proof can be considered as a conservative way to approach this. This particularly concerns the relative shares of professors according to disciplinarity, since here a positive correlation can only be obtained if the relative proportional concentration of professorial staff according to discipline has increased.

Tab. 13: Correlation analysis results for DoL and Spec variables for full sample and according to clusters

	Sample		Cluster **(1)**		Cluster **(2)**		Cluster **(3)** und **(4)**	
	FL	PP	FL	PP	FL	PP	FL	PP
Task Div. Prof / Denom.	0.090 (0.707)	0.066 (0.801)	0.324 (0.479)	0.786** (0.036)	-0.434 (0.390)	0.090 (0.865)	-0.865 (0.135)	-0.834 (0.166)
Task Coord. – Topics	-0.234 (0.320)	0.315 (0.218)	-0.225 (0.627)	0.040 (0.932)	0.234 (0.655)	0.513 (0.298)	0.522 (0.478)	0.436 (0.564)
Task Coord. – Subjects	0.010 (0.966)	0.495** (0.043)	0.458 (0.301)	0.458 (0.301)	0.212 (0.687)	0.177 (0.737)	-0.077 (0.923)	0.515 (0.485)
Task Coord. – Fields	-0.167 (0.481)	0.332 (0.193)	0.038 (0.935)	0.039 (0.934)	0.230 (0.661)	0.459 (0.360)	-0.698 (0.302)	0.388 (0.612)
Task Coord. – Disc.	0.072 (0.763)	-0.024 (0.926)	-0.112 (0.812)	-0.051 (0.913)	0.388 (0.447)	0.174 (0.741)	-0.114 (0.886)	-0.344 (0.656)
Spec. Conc. – (H) Denom.	0.429* (0.059)	0.167 (0.521)	0.461 (0.298)	-0.079 (0.867)	0.247 (0.637)	-0.073 (0.891)	0.191 (0.809)	0.513 (0.487)
Spec. Conc. – (H) Subjects	0.429* (0.060)	0.210 (0.418)	0.703* (0.078)	-0.013 (0.978)	0.232 (0.658)	-0.067 (0.899)	0.188 (0.812)	0.445 (0.555)
Spec. Conc. - (H) Fields	0.013 (0.957)	0.232 (0.370)	0.278 (0.546)	0.663 (0.104)	-0.389 (0.446)	-0.216 (0.681)	-0.232 (0.768)	-0.720 (0.281)
Spec. Conc. - (H) Disc.	0.074 (0.758)	0.355 (0.162)	0.089 (0.850)	0.836** (0.019)	0.342 (0.507)	-0.050 (0.926)	-0.585 (0.416)	-0.777 (0.223)
Spec. Ind. (Gravity)	0.384* (0.095)	0.219 (0.399)	0.284 (0.537)	0.530 (0.221)	-0.424 (0.402)	-0.531 (0.279)	-0.287 (0.713)	0.299 (0.701)

Tab. 14: Correlation analysis results for disciplinary shares for full sample and according to clusters

	Sample		Cluster **(1)**		Cluster **(2)**		Cluster **(3)** und **(4)**	
	FL	PP	FL	PP	FL	PP	FL	PP
Disciplinary	0.438* (0.079)	0.001 (0.997)	0.569 (0.183)	0.682 (0.092)	0.303 (0.560)	0.218 (0.724)	0.300 (0.700)	-0.083 (0.917)
Engineering Sciences	0.741*** (0.004)	0.623** (0.023)	0.998*** (0.002)	0.992* (0.008)	0.916*** (0.029)	0.528 (0.472)	-0.361 (0.639)	0.2670 (0.826)
Medical and Life Sciences	0.574** (0.025)	0.451* (0.060)	0.964*** (0.005)	0.771** (0.043)	0.959** (0.042)	0.562 (0.438)	-0.583 (0.604)	-0.452 (0.548)
Natural Sciences	0.385 (0.127)	0.462** (0.046)	0.730* (0.063)	0.294 (0.522)	0.376 (0.462)	0.670 (0.216)	-0.563 (0.437)	0.560 (0.441)
Interdisciplinary	-0.007 (0.980)	0.049 (0.853)	-0.027 (0.955)	-0.259 (0.621)	0.178 (0.736)	0.760 (0.136)	-0.362 (0.638)	-0.039 (0.976)
Multidisciplinary (miscellaneous)	0.380 (0.528)	.	.	.	0.426 (0.720)	.	.	.
Pluridisciplinary	0.623* (0.099)	0.194 (0.426)	0.392 (0.385)	0.207 (0.656)	0.361 (0.550)	0.812 (0.095)	-0.554 (0.446)	-0.821 (0.179)
Engineering Sciences	0.232 (0.387)	.	0.956*** (0.001)	.	0.330 (0.588)	.	.	.
Medical and Life Sciences	0.745*** (0.001)	0.268 (0.354)	0.637 (0.124)	0.168 (0.719)	0.987*** (0.002)	0.934 (0.234)	-0.582 (0.418)	-1 (1)
Natural Sciences	0.037 (0.893)	-0.236 (0.417)	-0.094 (0.842)	-0.661 (0.153)	0.121 (0.847)	0.778 (0.222)	-0.760 (0.240)	-1 (1)

In table 13, results of the correlation analysis for DoL and Spec. variables for full sample and according to clusters is given.[25] In the full sample case, significant positive correlations can be found for first and last sample observation of specialization concentration on the level of the denomination and the subject area, as well as for specialization gravity. For pre- and post-war mean, a positive significant effect exists for task coordination of subject areas. When looking at the results for the individual clusters, we find significant correlations only in cluster 1 for (PP) task division and concentration on disciplinary level as well as on the level of subject areas (FL). It should be kept in mind though that for the individually analyzed clusters, the pairwise correlations further drop in numbers of observations to 7 institutions for cluster 1, 6 for cluster 2 and 7 for the clusters 3 and 4.

In table 14, results of the correlation analysis for disciplinary shares for full sample and according to clusters is given. In the full sample case, significant positive correlations can be obtained for first and last sample observations, shares of disciplinary and pluridisciplinary professors. Within the disciplinary shares, correlations are significant for the engineering sciences and medical and life sciences. Within the pluridisciplinary shares, this seems to particularly concern the medical and life sciences. Turning to the pre- and post-war perspective, significant correlations can be found for the disciplinary shares of the engineering sciences, medical and life sciences and natural sciences.

For individual clusters, significant correlations of the (FL) disciplinary shares of engineering and medical and life sciences are found for cluster 1 and 2. Analogous to the analysis of individual DoL and Spec. variables, no significant correlations are detected for the joint sample of cluster 3 and 4. The latter could signal an independence from path dependencies, or that the combination of the two clusters might mask relevant information. Significant correlations of pluridisciplinary shares in first and last sample observations are obtained for cluster 1 in the engineering sciences and cluster 2 in the medical and life sciences. When considering pre- and post-war means instead of first and last

[25] Clusters 3 and 4 are considered as one cluster since the latter is only constituted by two institutions and the differences obtained between the two clusters based on the box-whisker plots are mainly proportional in size (, not structural as it is the case with clusters 1 and 2).

period observations, the findings for disciplinary shares of cluster 1 are robust.

This can be interpreted as such that a university in cluster 1, starting with a larger share in disciplinary engineering in 1890 and for the period 1890 to 1920 is more likely to show an even higher proportional concentration in disciplinary engineering in 2020 and for the period 1950 to 2020 on average. On the contrary, a university in cluster 1, with a larger share in disciplinary medical and life sciences in 1890 and from 1890 to 1920 is likely to have an even higher relative share of professors working in the medical and life sciences in 2020 and over the whole post-war period considered. In conjunction with the detected growth in professorial numbers in section 3.3.1.1, this must not mean a reallocation from professorial staff from one discipline to another, but much rather suggests a disproportional growth in research domains where the relative concentration was higher in earlier periods.

To summarize, specialization creates path dependencies within institutions over the course of time. Empirical evidence for the latter is clearest for the European institutions (cluster 1) and sufficient for the universities in cluster 2 as well as for the full sample. The latter are also the institutions with the best data availability in pre- and post-war period making it more likely to obtain significant results for the correlation analysis when compared to the merger of cluster 3 and 4 (, which are significantly younger and for which no path dependencies could be securely established).

The moderate correlations obtained for the DoL components point at a potential existence of path dependencies created by initial configurations of division of labor. The latter seems to be above all applicable to cluster 1. Again, the correlation analysis here is implemented in a very strict form, considering pairs of observations of only 6 to 20 institutions lying 120 years apart. Nonetheless, the mixed results for the DoL component, which cannot be supported by a more granular analysis as it is the case for Spec. with the correlations of disciplinary profiles, are not sufficient to establish that initial configurations of DoL necessarily significantly induce path dependencies. On the contrary, given the conservative approach and low number of observations chosen, it is most certainly not evidence against it either. The latter should be subject to future research when micro-data for a larger number of institutions is available.

4. Examining the Effect of Institutional DoL and Spec. on Universities' Publication Productivity

(Quantitative Empirical Analysis)

4.1 Methodology: Conditional efficiency framework

4.1.1 A brief introduction into non-parametric analysis of production efficiency

The second main goal of this work is to (challenge the idea of an efficient cognitive division of labor and) establish that the way labor is divided within the scientific community is not necessarily efficient. To provide empirical evidence for the latter claim, the introduced measures of division of labor and specialization need to be related to a measure of efficiency of the scientific production process. As outlined in section 2.1, the scientific production process is complex, making an operationalization of an efficiency measurement here much more complicated than in standard production contexts with a set of limited, homogenous and comparable production factors, for which data availability is not restricted. In the following sections, the choice of a state-of-the-art conditional nonparametric methodology will be motivated, which allows to account for the complexity of the scientific production process and the specific requirements linked to the available data.

Modern economists' interest in evaluating production efficiency dates back as far as to the seminal works of Debreu (1951) and Koopmans (1951). In production theory, efficiency is understood as the maximum attainable productivity within a given production process (Daraio and Simar 2007). Productivity in turn is defined by the ratio of (a set of) outputs (produced) to (a set of) inputs employed. The maximum attainable productivity can be described by a production function modeling the input-output relationship. Ideally, the efficient production function can be defined theoretically or is known to the researcher, whenever production efficiency is evaluated. (Mastromarco et al. 2019; Seiford and Thrall 1990)

Alternatively, if the efficient production function is unknown, it may be defined nonparametrically or estimated parametrically using e.g., a Stochastic Frontier Analysis approach. The latter though requires for the specification of the functional form of the relationships of inputs and outputs. For the scientific production process, making any *a priori* assumptions at all is clearly undesirable though, since we neither have any information on the true production function nor on the underlying functional form (production, cost, profit or distance function). It thus seems reasonable to turn to nonparametric efficiency analysis methods.[26,27] (Bornmann et al. 2023)

The idea of measuring technical efficiency nonparametrically was introduced in the pioneering work of Farrell (1957), which proposed to define observations as inefficient if they operate below a production function constructed by the empirically observed observations with the highest productivity ratios. This deterministic production function envelops all observations, for which efficiency is evaluated, by spanning an efficiency frontier across the most productive units within a sample, given a specific input or output level. A firm or a decision-making unit (DMU) is thus efficient if it operates at the highest empirically observed productivity in converting inputs into outputs, which is assumed to be the maximum attainable productivity of a particular scale section. (Farrell 1957)

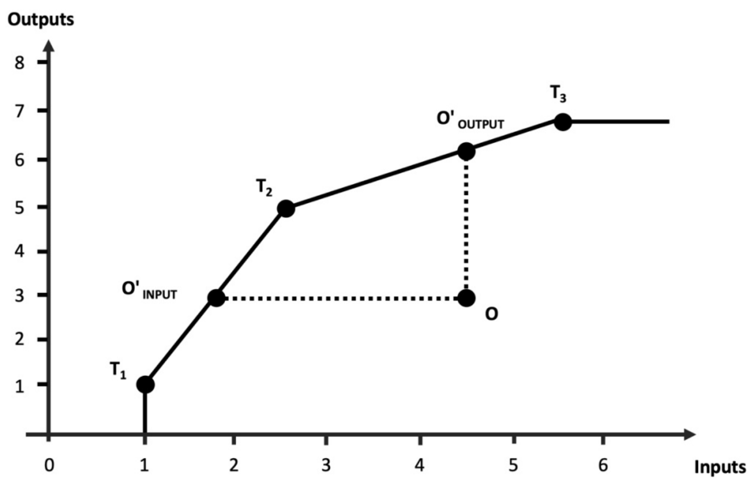
The degree of inefficiency of the units below the efficient production function can be assessed by their distance to the frontier. Different methods have been proposed to calculate efficiency estimates using different distance functions in a multidimensional space, with the Data Envelopment Analysis (DEA) as first introduced by Charnes, Cooper

[26] Daraio (2019) argued that the lack of specification of a functional form is a clear advantage of nonparametric methods in context of scientific production. For one, the relationship of inputs and outputs in research activities on individual level is known to be uncertain, lagged, nonlinear, with scientists' productivity being extremely skewed. Further, it is unclear how these factors combine on an institutional level, making a nonparametric approach the best fit.

[27] Further, there are recent examples for nonparametric methods being used to assess efficiency of universities. (See e.g., Bornmann et al. 2023, Daraio et al. 2015, Daraio et al. 2015a)

and Rhodes (1978) being the most frequently used and adapted approach (Liu et al. 2013). A simple illustration of a one input, one output model under variable returns to scale assumption is provided in figure 13.

Fig. 13: Toy Example – Banker, Charnes Cooper (1984) DEA Model



Own illustration based on Cooper et al. (2007: 90)

Some of the properties of the DEA approach can be visually inspected in the toy example above. In the case of variable returns to scale, the production possibility set is defined by the observations T_1, T_2 and T_3 which envelop the inefficient observation O and constitute the efficiency frontier. The distinct sections of the efficiency frontier describe different economies of scale, with the section spanned by T_1 and T_2 being substantially steeper than the section spanned by T_2 and T_3. Overall, the empirically defined boundary seems to suggest a production process subject to decreasing returns to scale. (Cooper et al. 2007)

The figure also reveals one of the assumptions embedded in the DEA approach, namely the convexity of the attainable set. The radial projection of the distance function to the frontier for O results in O1', for which no empirical observation is available. Assuming convexity though, all projections to the frontier denote attainable input-output combinations in the production possibility set. The convexity assumption thus implicates divisibility, meaning that it is feasible to proportionally reduce the input-output combination for any observed production unit without disproportionate productivity losses. (Johnson 2007)

Finally, the graphical inspection of the toy example allows us to obtain two distinct perspectives on the quality of the inefficiency of point D. Employing radial distance functions, we can either project O along the input dimension to the frontier in $O1'_{INPUT}$ (input-orientation) or alternatively, along the output dimension to the frontier in $O1'_{OUTPUT}$ (output-orientation). In the former case, O could become technically efficient by saving on approximately 2.5 input units, while producing the same number of outputs. In the latter case, O is found to be technically inefficient, because given the number of inputs employed, it should be feasible to produce about three more output units. (Cooper et al. 2007)

In summary, the basic idea of evaluating production efficiency nonparametrically is to derive measures of (in-)efficiency by evaluating the distance of observations with average or low productivity ratios towards an efficient production function. As opposed to parametric approaches, the latter is constituted by the most productive empirically observed observations, which span an efficiency frontier that defines the production possibility set deterministically. While the DEA approach is the most frequently employed approach and is also well-suited for a comprehensible introduction, there are advanced nonparametric methods available, which will be introduced in the upcoming sections.

4.1.2 Efficiency measurement

4.1.2.1 FDH model

In reference to Koopmans (1951) and Debreu (1951), a production process in which q outputs are produced employing p inputs can formally be defined by the production set of technically feasible input-output combinations in the Euclidean space $\mathbb{R}_{+}^{p+q}$ as (Simar and Wilson 2015):

$$\Psi = \left\{(x, y) \in \mathbb{R}^{p+q} \mid x \text{ can produce } y\right\}, \tag{1}$$

with the boundary (efficient production frontier) denoted by:

$$\Psi^{\vartheta} = \left\{(x, y) \in \Psi \mid (\gamma^{-1}x, \gamma y) \notin \Psi \text{ for all } \gamma > 1\right\}. \tag{2}$$

The basic assumptions usually made regarding the attainable production set are free disposability, meaning the (logically true) requirement that more (input) resources could be employed than necessary:

$$\begin{aligned} &\forall (x, y) \in \Psi \text{ and all } (x', y') \text{ such that } x' \geq x \\ &\text{and } y' \leq y, (x', y') \in \Psi, \end{aligned} \tag{3}$$

'no free lunch', hence the requirement for some of the elements of p to be positive:

$$(x, y) \notin \Psi \text{ if } x = 0 \text{ and } y \geq 0, y \neq 0 \tag{4}$$

and finally, the assumption that the boundary Ψ is convex for all $\tau \in [0,1]$:

$$(x, y) = \tau(x_1, y_1) + (1-\tau)(x_2, y_2) \in \Psi, \tag{5}$$

if (x_1, y_1) and (x_2, y_2) are elements of the attainable set. (Simar and Wilson 2015)

Technical efficiency, as introduced in the section above may now be evaluated oriented in the direction of the input or output space. The Debreu-Farrell measure of efficiency $\theta(x, y)$ is defined by the minimal contraction in inputs necessary to project an observation (x, y) onto Ψ^{ϑ} (Simar and Wilson 2015):

$$\theta(x, y) = \inf\left\{\theta \middle| (\theta x, y) \in \Psi\right\}. \tag{6}$$

Accordingly, the (output-oriented) Farrell measure of efficiency, where the projection of a point onto the frontier is made alongside the maximum feasible expansion in the output space, is given by (Simar and Wilson 2015):

$$\lambda(x,y) = \sup\left\{\lambda \middle| (x,\lambda y) \in \Psi\right\}. \tag{7}$$

Given the remarks on the characteristics of the scientific production process in section 2.1, the assumptions of free disposability and 'no free lunch' seem reasonable. It should be unproblematic to assume that a fixed number of e.g., publication output could be produced with excess (i.e. academic staff or funding) resources available. And even though indeed, one could argue that large proportions of citational return do not require for extra input (*Matthew effect*), at least some positive input, in the sense of resources for the initial production of the cited paper, is necessary.

The convexity assumption on the other hand is problematic for a variety of reasons.[28] Regardless of the choice of outputs, we already know from the analysis in chapter 4 that the professorial staff (, which will be employed as input) differs structurally for what was characterized as pre- and post-war science and further differs according to different university types or clusters.[29] While the former will be addressed by splitting the data in two subsamples and estimating efficiency using two distinct attainable sets, a 'pre- and a post-war frontier', the issue of the different university types remains. Also, the observations within the pre-war sample and the post-war sample refer to the same institution at different points of time. While it is common practice to measure a joint frontier for panel data, e.g. to provide more robust measures of efficiency, the time frame is usually more moderate.[30] This issue par-

[28] Note that a more thorough discussion of the input and output data employed will occur in section 4.2.1, yet a brief discussion of the dataset structure is necessary at this point.

[29] These differences include but are not limited to the access to funding resources, recruiting and number of academic staff, as well as teaching responsibilities, resulting in systematic heterogeneity of professorial staff input (quality) according to the university types identified.

[30] See Bornmann et al. (2023) for an example of measuring a joint frontier for efficiency measurement of universities over the period 2003 to 2018. In their analysis, no interaction of efficiency and time could be obtained.

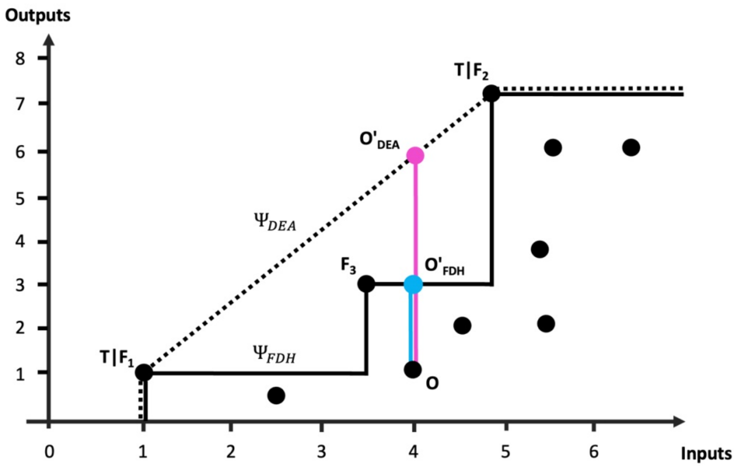
ticularly concerns the post-war sample, which covers the extensive period of 1950 to 2020.

Naturally, over the course of half a decade, the attainable set should be affected by major shifts in technology (i.e. the introduction of modern communication technologies) supposedly favoring production. On the contrary, the accumulation of knowledge ('increasing knowledge burden' argument), could make production of original knowledge more resource intensive over the course of time. Assuming convexity when benchmarking the observations against one joint frontier is thus inadequate, because the empirically unobserved fragments of the constructed boundary might not be feasible or at best it remains unclear whether those input-output combinations are feasible. To account for the uncertainty linked to the specific structure of the dataset at hand, the first measure put in place to guarantee for a sound evaluation of efficiency is to turn to robust advanced nonparametric methods.

As opposed to the traditional DEA estimator, the Free Disposal Hull (FDH) approach does rely exclusively on the free disposability assumption, thus not imposing convexity. Instead of enveloping the data with piece-wise linear segments connecting the efficient observations that constitute the boundary, the FDH estimator defines a facet $\hat{\Psi}$, constituted by all input-output combinations, which are not dominated by other observations. For the remaining observations, efficiency is determined relative to the dominating facet of $\hat{\Psi}$. (Daraio and Simar 2007)

In the simple one input, one output case, this facet may be imagined as the staircase function depicted in figure 14. The dotted line represents the DEA frontier as described in the section above, the solid line demarcates the dominating facet of the FDH estimator. As we can see, imposing the convexity assumption, observation F_3 is classified as inefficient. In case of the FDH approach though, F_3 is an efficient observation belonging to the attainable set, since there is no other observation available that dominates F_3 in the input and output dimension simultaneously. Provided that we accept the free disposability assumption as reasonable, this guarantees that for all inefficient observations it is empirically proven that it is technically feasible to produce more output using less input.

Fig. 14: Toy Example – DEA versus FDH Model



Own illustration based on Daraio and Simar (2007: 36)

It is obvious that this advancement comes at the cost of a slower convergence rate. In cases where convexity is a reasonable assumption, the DEA estimator will converge faster than the FDH estimator (and with a lower number of observations) to the true underlying functional form. This is also of practical importance, since the FDH estimator is therefore also more prone to the so-called *curse of dimensionality*, which means that in an increased input-output space (multiple inputs and multiple outputs), the number of efficient observations tends to increase disproportionately. (Simar and Wilson 2015)

Technically, for an observed sample $\chi_n = \{(X_i, Y_i)\}_{i=1}^n$, the FDH estimator of Ψ is constituted by the union of all orthants (positive in x and negative in y) with their vertex at the empirical observations and can be denoted by (Simar and Wilson 2015):

$$\hat{\Psi}_{FDH}(\chi_n) = \left\{(x, y) \in \mathbb{R}_+^{p+q} \mid y \le Y_i, x \ge X_i, (X_i, Y_i) \in \chi_n\right\}. \quad (8)$$

To obtain the Farrell-Debreu and the (output-oriented) Farrell measure of efficiency the FDH estimator can simply be plugged in equations

(6) and (7) respectively. The output oriented FDH efficiency is thus given by (Simar and Wilson 2015):

$$\hat{\lambda}_{FDH}(x,y) = \sup\left\{\lambda \middle| (x,\lambda y) \in \hat{\Psi}_{FDH}(\chi_n)\right\} \tag{9}$$

4.1.2.2 Directional distance functions

The actual projection of inefficient observations to the respective attainable set is operationalized using distance functions. In the standard DEA model this is usually done employing radial distance functions, which allow for the calculation of either the input-oriented or output-oriented efficiency estimates. In some use cases it might be desirable though to employ a more flexible way of measuring efficiency, e.g., projecting the inefficient points in the dimension of selected inputs or outputs or to account for non-discretionary inputs and outputs, outside of the sphere of influence of the decision-maker.

Chambers et al. (1996) proposed to use a directional vector $d = (d_x, d_y) \geq 0$ instead of radial distance functions to project inefficient observations onto the attainable set, which in the FDH can be denoted as (Daraio et al. 2020; Simar and Wilson 2015):

$$\begin{aligned} &\beta_{FDH}(x,y \mid d_x,d_y) \\ &= \sup\left\{\beta > 0 \middle| (x - \beta_{FDH} d_x,\, y + \beta_{FDH} d_y) \in \hat{\Psi}_{FDH}(\chi_n)\right\}. \end{aligned} \tag{10}$$

Note that a distance of $\beta_{FDH}(x,y \mid d_x,d_y) = 0$ means that an observation is efficient and constitutes the efficient hull. Given its additive nature, directional distances allow for an efficiency measurement with negative or zero values in inputs and outputs (, as opposed to the traditional DEA, which can only handle positive integers). (Daraio et al. 2020)

While this might not seem useful at first sight, given that it is hard to imagine zero values in a production process (, let alone negative values), this property can be beneficial to tackle the above addressed issue of the *curse of dimensionality*. In case of a moderate information loss, variables employed in the efficiency model may be implemented as factor variables, which of course contain negative values by definition. This will be particularly useful when external factors are introduced into the model (in section 4.2.2). Analogously to the procedure in the cluster analysis in the previous chapter, the different variants for task coordination and concentration may be reduced to one factor variable to restrict the analysis to a reasonable number of different models with a maximum of four variables (one input, one output, one or two external factors) producing considerate numbers of efficient observations. (Daraio et al. 2015)

Another interesting property of directional distances is that they are a generalized version of the radial distance functions. By setting vector d_y to zero, input-oriented efficiency is measured. By setting vector d_x to zero, output-oriented efficiency is measured. And in case a comparison with radial DEA efficiency estimates is of interest, the original Farrell measure of efficiency can be rediscovered by $1+\beta(x,y|0,y)$ (Simar and Vanhems 2012)

In this work, an output-oriented specification of directional distances will be used $d_x = 0$ and $d_y = \overline{Y}$. It seems appropriate to assume that decision-makers (here: universities) are more interested in maximizing their scientific output given their input resources at hand than saving on academic staff input while keeping similar levels of output. Arguably, the latter could be interesting in cases where efficiency is not directed at research activity, but rather at an analysis of teaching efficiency. Regardless of this rationale, the output-orientation is also the common choice for efficiency analysis of universities in the literature (e.g. Daraio et al. 2015 and Daraio et al. 2015a).

4.1.3 Accounting for extreme observations and external factors

4.1.3.1 Partial frontiers

The FDH approach allows us to benchmark observations only against empirically observed units, thus eliminating the risk of defining areas in the attainable set as feasible, which in reality are not feasible, or for which there is simply no information on feasibility. Nonetheless, this does not prevent that observations of universities from later periods (of the post-war sample) could be benchmarked against a frontier constituted by observations of earlier periods (of the post-war sample), or the other way around. If indeed, there would exist substantial shifts in technology[31], enabling a much higher productivity today and e.g. an observation of 2010 serves as dominating unit of all observations from 1960, the upwards shifted attainable set would make the universities appear a lot less efficient in earlier periods than they would be if they were benchmarked against a boundary constituted only by observations of 1960 (, defining their 'own' technology so to speak). Supposedly one would intuitively argue that such a comparison is unfair and subsequently this might distort the obtained effects of the influence of division of labor and specialization on productivity, which is of course our primary interest. Even though it can be shown that this issue is subordinate because of the specific relationship of size with institutional types and period (see 4.3), it might in any case be useful to methodologically take precautions against (different kinds of) outlier observations potentially distorting the shape of the efficient boundary.

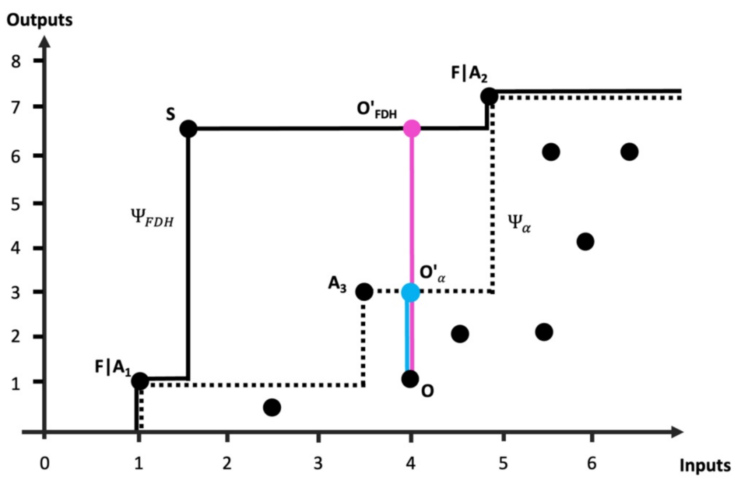
Robust nonparametric estimators have first been proposed by Cazals et al. (2002), Aragon et al. (2005), Daouia and Simar (2007) for radial measures and have been extended to directional distance functions by Simar and Vanhems (2012). The idea of robust measures is to disclose some particularly productive sample observations as superefficient. A small quantity of observations, which constitute the original full frontier are now permitted to lie beyond the efficient boundary. The resulting partial frontier, against which the inefficient observations are now benchmarked, is more robust in the sense of being less sensitive to outlier observations influencing the attainable set. (Daraio et al. 2020)

In the simple one input, one output case portrayed in figure 15, we can see that the productivity ratio of S deviates substantially from the productivity ratios of the remaining sample observations. Supposedly, the high productivity of S is caused by a measurement error or is due to reasons not disclosed to the researcher, making it a bad fit for the

[31] Shifts in technology need not be restricted to technological progress. Shifts in the attainable set could also be due to changing publication habits (, prioritizing for example the publication of multiple short peer-reviewed journal articles over one enlarged thesis), different levels of non-professorial academic staff, or disproportional funding due to political emphasis on scientific activity in certain limited periods and research fields (Bonaccorsi et al. 2017).

comparison group of decision-makers under evaluation. In the toy example, if S is permitted to lie beyond the frontier, a partial frontier (dotted line) results that seems to provide a more reasonable benchmark for evaluating efficiency. The new partial frontier seems to be a better fit for enveloping the inefficient observations, accounting better for what seems to be a production context subject to increasing returns to scale by revealing the varying feasibility of output quantities given different input levels.

Fig. 15: Toy Example - FDH versus Orderalpha Model



Own illustration based on Daraio and Simar (2007: 38)

In order to implement the partial frontier approach, Cazals et al. (2002) proposed to adopt a probabilistic formulation of the nonparametric estimators. The probability of finding a unit (X,Y) that dominates the observation (x,y) can be denoted by the joint probability $H_{XY}(x,y) = \Pr(X \leq x, Y \geq y)$, where the joint distribution of (X,Y) has support over Ψ [32], which is given by (Simar and Vanhems 2012):

[32] Note that all equations refer to the FDH estimator Ψ_{FDH} introduced in section 4.1.2.1. For a better overview, subscripts are left out of the equations.

$$\Psi = \left\{ (x, y) \in \mathbb{R}^{p+q} \middle| H_{XY}(x, y) > 0 \right\}. \tag{11}$$

Simar and Vanhems (2012) define this for directional distance functions as:

$$\beta(x, y | d_x, d_y) = \sup \left\{ \beta > 0 \middle| H_{XY}(x - \beta d_x, y + \beta d_y) > 0 \right\} \tag{12}$$

Two variants for robust frontiers have been proposed, the order-m and the order-α quantile frontier. In the former case, the percentage of points classified as superefficient is (iteratively) trimmed by m units randomly drawn from the sample to calculate an average maximal value of outputs (or minimum value of inputs), which define the boundary. In the order-α quantile frontier approach the probability of obtaining observations below the efficient boundary is fixed at the $(1-\alpha)$ percentile. For the sake of a good comparison of the different models assessed in this work (pre-war and post-war frontier, with and without DoL and Spec. external variables) avoiding an iterative approximation of the frontier seems desirable. (Daraio and Simar 2007; Daraio and Simar 2014)

For any $\alpha \in [0,1]$ in the above specified output-oriented one input, one output model, the directional distance of order-α corresponds to (Daraio et al. 2020):

$$\beta_a(x, y | 0, \bar{Y}) = \sup \left\{ \beta \middle| S_{Y|X}(y + \beta | x) > 1 - \alpha \right\}, \tag{13}$$

with the conditional survival function of Y given $X \leq x$ defined by $S_{Y|X}(y \mid x) = \Pr(Y \geq y \mid X \leq x)$. For any $\alpha < 1$, the α-quantile of the conditional distribution of the output of observations with lower input employment than x serves as the efficient boundary for observation (x, y). Subsequently, the order-α quantile frontier can be defined as (Daraio et al. 2020):

$$\varphi_a(x) = y + \beta_\alpha(x, y \mid 0, \bar{Y}), \tag{14}$$

where for large values of y (in the support of $S_{Y|X}$) it could result that $\beta_\alpha(x, y \mid 0, \bar{Y}) < 0$, meaning that the observation lies beyond the efficient partial boundary and is classified as superefficient. In case of $\alpha = 1$, we rediscover the original FDH estimator. (Daraio et al. 2020)

4.1.3.2 Conditional frontiers

Finally, in this section the introduction of external factors into the efficiency model will be explained. First, it should be pointed out that neither division of labor nor specialization or the four components by which the two phenomena are operationalized (task division, task coordination, concentration, specialization gravity) can be considered external to the production process in the sense of a shock or non-discretionary effects of stochastic nature.

Clearly though, they can neither be considered an input or output of the scientific production process. In addition, given the lack of institutions controlling coordination costs in science, and the here established self-reinforcing character of DoL and Spec., in this work the hypothesis is supported that they are not controlled and disposed like a production factor by a decision-maker.[33] Consequently, the effects of DoL and Spec. will be introduced into the efficiency model as external factors.

Given the probabilistic notion of the production process introduced in the previous section, we may accordingly define the attainable set $\Psi^{\mathcal{Z}}$ of universities facing heterogeneity factors $Z \in \mathcal{Z} \subset \mathbb{R}^r$ that may influence the production process (here: DoL and Spec.), with the joint support of (X,Y) given $Z = z$ by (Daraio et al. 2020):

$$\Psi^{\mathcal{Z}} = \left\{(x, y) \in \mathbb{R}^{p+q} \mid x \text{ can produce } y \text{ if } Z = z\right\}$$
$$= \left\{(x, y) \in \mathbb{R}^{p+q} \mid H_{XY|Z}(x, y \mid z) > 0\right\} \tag{15}$$

The external factors may affect the efficiency of an observation either by affecting the probability of reaching the attainable set, by affecting the attainable set itself or by affecting both (Daraio et al. 2020). The possibility that the external conditions may affect $\Psi^{\mathcal{Z}}$ itself, is the primary reason why the so-called two-stage approach, where in a first step

33 Be it the self-governed scientific community or a centralized university or faculty administration for that matter.

a nonparametric efficiency analysis is conducted and in a second step covariates are regressed on the yielded efficiency estimates (using a censored or truncated regression model), is seen as invalid (see Simar and Wilson (2011) for a discussion of this issue). The latter approach is only a valid option in cases, where it can be proven that the so-called separability condition holds, that is that $\Psi^{\mathcal{Z}} = \Psi$ for all $z \in \mathcal{Z}$.

Indeed, it might be hard to come up with empirical examples of external factors that, in case they have an impact on the production process, strictly influence only inefficient observations. The theoretical remarks on division of labor and specialization suggest that in this work the separability condition is violated. There is no evidence supporting the idea that the effect of division of labor or specialization is restricted to inefficient universities' productivity. Conversely, in accordance with the Smithian idea of DoL driving technological progress, we may much rather expect the productivity of peers to be determined to some extent by their degree of division of labor. This is particularly true for the pre-war sample, which may also be seen as a growth phase of modern science (and some relatively young US institutions) where task division and gravity of specialization should be subject to the still limited extent of the scientific 'market'.

The directional distance function conditioned on Z is given by (Daraio et al. 2020):

$$\beta(x, y \mid 0, \bar{Y}, z) = \sup\left\{\beta > 0 \mid H_{XY|Z}(x, y + \beta\bar{Y} \mid z) > 0\right\}, \tag{16}$$

where $H_{XY|Z}(x, y \mid z)$ is a nonparametric estimator, applying Kernel smoothing techniques over the neighborhood of z for Z_i, since observed values $Z_i = z$ may not be continuously available. The localized analog of $H_{XY|Z}(x, y \mid z)$ can be defined as (Daraio et al. 2020):

$$\hat{H}_{n,\, XY|Z}(x, y \mid Z = z) = \frac{\sum_{i=1}^{n} (X_i \leq x, Y_i \geq y) K_h(Z_i, z)}{\sum_{i=1}^{n} K_h(Z_i, z)}, \tag{17}$$

where $K_h(Z_i, z)$ is a Kernel function with compact support[34] and h is a bandwidth vector, to be individually determined for each factor of z.

[34] In this work Epanechnikov Kernel functions are employed.

It is evident that bandwidths selection for the different components of z is crucial since the smoothing of Z can hardly be evaluated *a posteriori* (, as opposed to the fitted functions of the effect of Z on the efficiency estimates, which can be supported by visual inspection, scatter plots etc.). (Daraio et al. 2020)

Initially, a least squares cross validation (LSCV) approach traditionally employed in nonparametric regression literature was proposed by Bădin et al. (2010) to determine appropriate bandwidth. Later the approach was amended to incorporate the conditional distribution function into the bandwidth selection process, adapting the selection process proposed by Li et al. (2013). Here, the implementation of the latter approach by Bădin et al. (2019) is used to generate cross validation plots and choose appropriate bandwidths for all considered external factors based on two criteria, (1) to minimize the LSCV criterion by choosing bandwidths at local minima or locations with sharp increases while (2) simultaneously minimizing the number of efficient observations in the conditional models. (Daraio et al. 2020)

Finally, we can extend the conditional model to the order-α case by plugging in the nonparametric estimator $H_{XY|Z}(x,y\,|\,z)$ in equation (15) and define the order-α conditional directional distance model as (Daraio et al. 2015):

$$\beta_\alpha(x,y\,|\,0,\bar{Y},z)=\sup\left\{\beta>0\,|\,H_{XY|Z}(x,y+\beta\bar{Y}\,|\,z)>1-\alpha\right\} \quad (18)$$

Calculations in this work have been performed by adapting the codes provided in Bădin et al. (2019) and Daraio et al. (2020) using the software MATLAB (version R2023b 23.2.0) provided by *The MathWorks Inc*. A general introduction into nonparametric efficiency analysis based on DEA and FDH estimators can be found in Cooper et al. (2007) and Daraio and Simar (2007) respectively. A more detailed introduction into the different models, variants and their properties can be found in Cazals et al. (2002) for order-α quantile frontiers, Daraio and Simar (2014) for directional distance functions and Daraio et al. (2020) for conditional frontiers.

4.1.4 *Interpreting the effect of DoL and Spec. on the efficiency measure*

The interpretation of the effects of the external factors incorporated in the conditional efficiency models is based on the methodology proposed in Bădin et al. (2012) and Bădin et al. (2014). In order to distill the effect of *Z* on efficiency the relationship of conditional to unconditional efficiency measure is of interest (Bădin et al. 2012):

$$R_\alpha(x, y \mid 0, \bar{Y}, z) = \frac{\beta_\alpha(x, y \mid 0, \bar{Y}, z)}{\beta_\alpha(x, y \mid 0, \bar{Y})}, \tag{19}$$

where for an observation $(X, Y) \in \Psi^Z$ producing under $Z = z$, the random variable is given by $R_\alpha(X, Y \mid Z = z)$ and the conditional expectation is defined by (Bădin et al. 2012):

$$\tau_\alpha^z(P) = \mathbb{E}(R_\alpha(X, Y \mid Z = z)), \tag{20}$$

with $\tau_\alpha^z(P) \to \tau^z(P)$ for $\alpha \to 1$. (Bădin et al. 2012)

For all $R_\alpha(x, y \mid 0, \bar{Y}, z) > 1$ we obtain an unfavorable and for all $R_\alpha(x, y \mid 0, \bar{Y}, z) < 1$ a favorable effect of *z* on the production process. Ratios above one result when the distance of an observation to the frontier under the condition of *z* is larger than its distance in the unconditional case. Accordingly, a value below one means that in the conditional case an observation moved closer to its boundary relative to the distance to the frontier where *z* is not accounted for (, with a ratio of 0 being the extreme case where accounting for *z* makes an observation move onto the frontier). In case of $R_\alpha(x, y \mid 0, \bar{Y}, z) = 1$ we find no effect of *z* on the efficiency estimate of the evaluated observation, hence the distance to the attainable set in the conditional and unconditional case are identical. (Bădin et al. 2012; Bădin et al. 2014)

Technically, where local ratios are above one, *z* acts as an additional undesirable output to be produced. A favorable effect in turn corresponds to the external factor acting as an additional input that is freely available. When describing the ratios as a function of *z* in an output-oriented framework, the interaction of *z* with the input thus needs to be accounted for. Ideally, there is no interaction between size and the external factor. Such ‘partial separability’ may not always be the case, so the effect of z on the ratios should be compared with the effect of size on the ratios to eliminate the possibility that the effect of the external factor can partially or fully be reduced to a size effect. In case of the variables employed to measure DoL and Spec. this seems especially important. By construction, the considered components are to a certain extent size dependent. This particularly concerns task division (, defined as number of professors per denomination). (Bădin et al. 2012; Bădin et al. 2014)

Arguably though, the perspective on the interaction of z with the input (size) must be a different one in this work since DoL and Spec. are part of the mechanism constitutive of size effects. From the theoretical remarks we know that they are explanatory factors of economies of scale. So instead of looking for partial separability, we may rather want to obtain an interaction with size and this interaction to be consistent with the theoretical understanding developed in the prior chapters. If e.g., a strictly positive effect of size and task division for the pre-war period is found, this would in principle be a consistent finding. Ideally, the effect of task division should reveal a clear, more robust pattern of the effect than the ratios as a function of the input. If the latter is not the case, it needs to be concluded that other factors relating to size explain the (favorable effect of size) better than task division. The interpretation of the effect of task division then needs to be limited to observations operating at input levels independent of the ratios. (Bădin et al. 2012; Bădin et al. 2014)

Bădin et al. (2012) suggest using two-dimensional scatter plots for inspecting the marginal effect of z on the efficiency ratios. Further, we may analyze the latter according to different levels of α in order to differentiate between the effect of z on the attainable set and its effect on the inefficiency distribution. This is useful to validate or reject the above addressed (full) separability condition and learn more about the role of z in the production process. In case the separability condition holds, we should find an effect of z on the inefficiency distribution yet no effect on the boundary.[35] To implement the latter the authors propose to look at the marginal effects on the full frontier and a robust frontier (, where α lies close to one, permitting a few outlier observa-

[35] In principle, the above introduced methodology could then be complemented by a parametric truncated regression analysis.

tions beyond the attainable set) as well as specifying a percentile by α (e.g., 50th percentile) that constructs a boundary on the middle of the distribution of the inefficiencies. The latter allows to unhinge the effect z exclusively has on the inefficiency distribution. The interaction with the input can be implemented using three- and two-dimensional scatter plots for full frontier, robust frontier and the boundary of the inefficiency distribution. (Bădin et al. 2012; Bădin et al. 2014)

Note that in the here employed nonparametric set-up, there is no testing for significance required (, or useful). In case we find local effects of Z on the efficiency ratios (for full / robust frontier and / or partial frontier given they are independent of the input), this is sufficient proof that DoL and Spec. influence efficiency and thus epistemic outcomes.[36] (In case we find a random distribution of the ratios close or equal to 1, an influence on efficiency and thus epistemic outcomes cannot be claimed) In case the local analysis of ratios reveals consistent patterns for the analyzed components, it is possible to go one step further and make propositions regarding the nature of the influence of DoL and Spec. on epistemic outcomes.

To assess the patterns of the analyzed components, the marginal analysis of the ratios will be complemented by fitting local linear kernel smoothing regression lines for all components of DoL and Spec. For the full frontier case, the local linear regression function is given by (Li and Racine 2023):

$$\hat{\tau}_n^z = \frac{\sum_{i=1}^{n}(s_2 - s_1(z - Z_i) * K(u) * \hat{R}_\alpha(X_i, Y_i \mid Z_i))}{s_2 * \sum_i^n K(u) - s_1^2}, \tag{21}$$

with $s_1 = \Sigma_{i=1}^n (z - Z_i) * K(u)$ and $s_2 = \Sigma_{i=1}^n (z - Z_i)^2 * K(u)$ capturing the local behavior of the observations, h being the bandwidth parameter of Gaussian Kernel function[37] (Li and Racine 2023):

[36] Thus far it has only been established that division of labor and specialization create path dependencies within universities. Proof that DoL and Spec. influence epistemic outcomes will thus be provided in case we find an effect on the efficiency ratios.

[37] There is no requirement for a specific Kernel function here.

$$K(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}}, \tag{22}$$

where $K(u) = K(z - Z_i / h)$. Since the choice of the bandwidth parameter is uncritical here (as opposed to the bandwidth selection process in the conditional frontier model), h will be specified according to Scott's and Silverman's rule of thumb respectively. Deviations of the regression lines fitted according to the two plug-in rules should give us an intuitive understanding of the nonlinear relationship of the variables, reveal heteroskedasticity within the data structure and allow for a good comparability of the fitted effects in the different model set-ups. (Li and Racine 2023)

Finally, to allow for a joint interpretation of the individual components of DoL (task division and task coordination) and Spec. (concentration and gravity), the plots for the marginal effects will be backed by 3D surfaces fitted using Gaussian Processes Regression (GPR). The latter method is well-suited for modelling non-linear relationships in cases where the number of observations is restricted. GPR nonparametrically models the interactions of the variables as probability distributions over all possible functional forms. All interactions were modelled in identical set-ups, using squared exponentials as covariance function. As dependent variable, z-scores of the efficiency ratios were employed and all observations lying more than two standard deviations outside of the two considered independent components were eliminated to facilitate generating a smooth interaction based on a dense core of observations. For a more detailed introduction into GPR, see Schulz et al. (2018) or Wang (2023).[38]

[38] Mean Squared Errors (MSE), Rooted Mean Squared Errors (RMSE) and R-squared estimates are provided along each plot created using GPR. Please keep in mind though that there is a trade-off in providing a model that is suited for visual inspection and one that is suited for inference. Here GPR is exclusively employed to facilitate the visual interpretation of the interacting effects. Error values approximating zero and an R-squared approximating one could easily be obtained by tuning the parameters of the model accordingly (, resulting in comparable overall patterns, yet at the cost of a crinkly surface with local minima and maxima).

4.2 Data

4.2.1 Efficiency model

4.2.1.1 Input data

In the scientometrics field, performance measurement of HEI (e.g., university rankings) often relies on the number of publications and citations according to institutions. Bornmann et al. (2023) have just recently pointed out that the information conveyed by the latter is limited whenever the amount of available resources is not accounted for. Simply put, instead of judging the performance of a university by its publication and citation output, the authors advocate to evaluate performance based on publication or citation productivity instead. (Bornmann et al. 2023)

Needless to say, that there are of course examples in the literature, where performance measurement in the higher education sector has been conducted considering resource input. But just like pointed out in section 3.1.3, where the need for a new institutional data set to measure DoL and Spec. was motivated, the input data employed in these studies is either not directly observed (e.g., Bornmann et al. 2023; Pastor and Serrano 2016), not differentiated according to academic function (e.g., Lepori et al. 2019, Daraio et al. 2015), or derived from the publication output (e.g., López-Illescas et al. 2011; Moed et al. 2011). In addition, even if micro-level data of universities was used (e.g., Herberholz and Wigger 2021; Daraio et al. 2015a), the analysis was limited to recent years or comparability of independently collected sources (Daraio 2019). Another benefit of the new data set is thus that it allows to analyze the efficiency of universities using empirically observed institutional input data for an enlarged time period.

In previous chapters, the idea has been discussed that this dataset allows to fill the gap between division of labor in scientific teams on the micro-level and specialization on the aggregated level of broadly defined epistemic branches or the overall institution (e.g., polytechnic universities) on the macro-level. The latter is achieved by taking into account the professorial denominations as informative intermediate medium that enables a sufficiently granular analysis of division of labor within institutions, yet beyond the scope of e.g., the individual laboratory. Analogously, instead of the denominations, the number of professorial staff can be used as an input for an efficiency model measuring universities' publication and citational productivity.

The rationale behind this idea is twofold. Given the data at hand, there are characteristics of the scientific production process we can account for (DoL, Spec. and size) and certain aspects which are known to be determinants of epistemic outcomes (e.g., funding, administrative staff, other academic staff, teaching requirements etc.)[39] yet for which data over the here considered period is not available nor comparable. While the methodology introduced in the previous section allows the isolation of the impact of external factors on efficiency, it does not enable an all-things considered judgment on why an individual university is inefficient at a particular point in time. Since the purpose of this work though is neither to provide a complete model of the scientific production process, nor to construct an alternative ranking method that allows for a fair judgment of efficiency,[40] we can naturally limit the interpretation of the results towards the impact of DoL and Spec. on the efficiency of the (sub-)samples.

Consequently, instead of adopting a total factor productivity perspective in the efficiency model (, which would be the requirement for a performance evaluation taking all inputs affiliated with the production process into account), a partial productivity model can be estimated using the number of professorial staff as sole input. Arguably, it may even be desirable to reduce the inputs of the efficiency model to just one factor.

For one, additional input factors considered in the literature, e.g., total academic staff or capital endowment are subject to severe heterogeneity in quality, which is not limited to, but particularly affects the different university types and the time period considered in this work. Heterogeneity in quality could result from the work of e.g., a nuclear physicist in 2010 requiring for disproportionally higher funding to achieve the same amount of epistemic outcomes when compared to a predecessor in 1960. Even in case a variable for capital endowment

[39] Lepori et al. (2019) for example showed for a sample of European and US universities that there exists a super-linear scaling relationship between universities' revenues and volume of publications and (field-normalized) citations.

[40] If the latter is of interest, see Daraio et al. (2015) for an approach that allows to 'whiten' the efficiency score from the influence of external effects.

was available and adjusted for purchasing power parity, incorporating it as an input into the model would disregard that scientific production might get increasingly more resource-intensive with ongoing knowledge accumulation. This idea previously introduced as the 'knowledge burden' is i.e. expressed by the enormous amount of capital necessary to build up and operate facilities like particle accelerators to push the knowledge boundary further or the previously addressed shift towards team-science.

Secondly, taking the idea of this work seriously, a large portion of the heterogeneity in additional input factors can be explained by division of labor and specialization. In a model, where a university specializes in particle physics (more than its competitors), incorporating capital endowment as an input variable would distort its efficiency estimate. While cost efficiency might be of interest in some cases, it is not the primary concern in this work and any effect of the external factors considered on efficiency would then relate to an undifferentiable measure of efficiency, which is partially constructed based on information on costs and partially on allocation efficiency. In other words, considering additional input variables might not be very informative when evaluating efficiency on institutional level since it neglects that the micro-production processes of e.g. 'nuclear physics' and 'occupational medicine' have totally different requirements for capital endowment. This equally concerns the number of administrative and other academic staff necessary to operate e.g., a molecular biology laboratory as opposed to a chemical laboratory. Acknowledging the latter, it will provide a more informative picture to permit unobserved heterogeneity in an efficiency model for which we have a clear concept of what it measures.[41] In a second step then, the above introduced methodology can be employed to distill the effect of the explanatory factors of interest, explaining a limited part of the heterogeneity (, but not all of it).

[41] Certainly, one may object that the latter quality differences are not always primarily caused by DoL and Spec, which would be the prerequisite for ruling out that the later observed effects are too to some extent determined by other factors, not directly accounted for in this framework. When limits of this work are discussed in chapter 5, this issue will be addressed.

In accordance with the remarks on the scientific production process, the professorial staff is a suitable choice for the sole input. University professors are in charge of the micro-level production process, managing the scientific production within teams (Lee et al. 2015; Häussler and Sauermann 2020), or organizing the laboratory life (Latour and Woolgar 1986; Knorr-Cetina 1984). Further, they are the entities interacting on the meso-level of the institution, allocating the cognitive labor and therefore determining the produced output. Even though their role changed and adapted over time, the latter is true for the whole period of modern science, making the professor the nucleus of institutionalized science, constitutive of universities as an organization. (Parsons and Platt 1990)

Parsons and Platt (1990) highlighted the role of the university professor as most important depositor of value commitments and influence in the academic system. Their privileged position allows them to dispose over their own research and teaching activities, while at the same time exerting influence on reputation and governing of the institution they are affiliated with by e.g., participating in new appointments or the design of curricula. This confirms the idea of the professor as a decision-maker of an isolated production process on the micro-level, while functioning as an input on the meso-level, where he is one of many actors within the institution. The second role resonates with the idea of the self-governing scientific community, in particular the appointment of new professors according to priorly demarcated denominations, which directly influences task differentiation. (Parsons and Platt 1990)

The authors further suggested that there are different perceptions of responsibility and pressure and different levels of influence between tenured full or associate professors and the temporarily employed assistant professors. Since the role of the latter is expected to differ according to university system and research field though, here the total number of professors (sum of full, associate and assistant professors and their equivalents in different university systems respectively) will be used as input in all efficiency models. This builds on the assumption that granted that there are significant differences in influence between different professorial types, all university professors are assumed to have sufficient impact on the production process by taking part in hiring and fundraising activities, determine task division and specialization, as well as choice of research activities. (Parsons and Platt 1990)

4.2.1.2 (Publication and citation) output data

The choice of the outputs is built on the literature review on the science studies in section 2.1. In this work, a publication productivity model (PPM) will be calculated using the number of natural sciences publications attributable to an affiliation and a citation productivity model (CPM) using the number of natural sciences citations attributable to an affiliation as outputs.

As already established in previous chapters, publications and citations are acknowledged indicators of scientific performance. Sociologists of science define publications as main product of scientific activity in all disciplines. The authors of the laboratory studies, putting themselves in the position of ‘anthropologists’ of the scientific production process emphasized the strange importance of producing a piece of paper, which is seemingly unrelated to complex interactions and operations performed using costly substances, materials and instruments (Knorr-Cetina 1984; Latour and Woolgar 1986). Given their insights in the laboratory, Latour and Woolgar (1986) even proposed to ‘consider papers as objects in much the same way as manufactured goods (71)’. And indeed, it is certainly not controversial to argue that in science the publication is the closest equivalent to produce yielded in other production contexts and therefore qualifies as output in the first (PP) efficiency model.

In science studies and scientometrics, the importance of the publication has in recent years been pushed back by the analysis of citations. Indicators based on citations regularly serve as a criterion in university rankings, determine funding or hiring opportunities of scientists and have eventually superseded peer-review as primary quality criterion in science. While the paramount importance of citations is critically discussed, there is a broad consensus that the attention a publication receives and the impact it has on future research (referencing the latter) to some extent reflects the quality or impact of research. (Bornmann and Haunschild 2019; Wang et al. 2013)

Clearly though, an ‘anthropologist’ studying the activities in a laboratory could not directly observe the production of citations. While he might observe a researcher utilize resources and engage in activities to promote his work e.g., by attending conferences, bringing it to the attention of peers in his field, this by no means guarantees that a publication is highly cited. On the contrary, a publication could receive outstanding attention, because of prior merits of the author (*Matthew effect*), only loosely corresponding to the resources employed to produce the publication. The correspondence of scientific production activity with citations is thus a lot less direct and more abstract than in case of publications. This raises the question whether citations can be considered as an output of scientific production and whether an efficiency model with citations as output meets all production theoretic demands.

Given the advanced nonparametric methodology applied, the latter should be uncritical, because the free disposability assumption equally holds for both publications and citations. It is possible for an institution to employ more professorial staff input yet produce the same number of (publications or) citations as before. The 'no free lunch' requirement is an interesting case, because at extreme occasions the input necessary to produce citational output could be fully reduced to the input used to produce the initial publication. The issue of cumulative advantage, also known as the *Matthew effect*, has been introduced in previous sections and while it concerns both publications and citations, the distribution of the latter might be more vulnerable to it, resulting in skewed distributions. A Nobel prize winner for example could publish a paper and employ zero resources linked to promoting the publication or convincing others in the field of its importance and nonetheless receive a substantial number of citations. Of course, technically, the 'no free lunch' requirement is nonetheless always fulfilled since receiving citations at least requires the initial production of the publication and consequently also requires a (non-zero) input.

Acknowledging that the productional dependency of professorial staff input and citations is more abstract than with publications, citations can nonetheless be considered an output of the scientific production process. Employing citations as an output in an efficiency model can be compared to other cases of production efficiency analysis, where the outputs employed contain information that goes beyond the mere amount of goods produced. Consider for example the incorporation of farm gross results in agricultural production efficiency analysis. Instead of employing physical produce as an output, agricultural economists traditionally use farm gross results when assessing technical efficiency of farms or farming sectors (Liu et al. 2013). Just like farm gross results correspond to the value of the physical produce sold, citation output can be understood as the publication output valuated by the (demand of the) scientific community. The number of citations thus serve as output in the second citation productivity (CP) efficiency model.

While both outputs might not be perfect proxies for what we intuitively understand as scientific progress, they are carriers of epistemic outcomes, contribute towards the accumulation of knowledge and puzzle-solving activities. Or as van Raan (2019) put it:

> 'But work of at least some importance provokes reactions from colleagues. They are the invisible college by which research results are discussed, and they play their role as members of this invisible college by referring in their own work to earlier work of other scientists. Scientific performance relates to the quality of the contribution in terms of increasing our knowledge (scientific progress) as perceived by other knowledgeable researchers (peer review), quantified and archived by citations. (243)'

As established in section 3.1.1, the latter understanding of progress in science might be best applicable to the branches of natural sciences and engineering, where the connection of quantity of research with quality of research is (by its empirical design) more pronounced than in the social sciences or humanities.

Before sources and sampling of the outputs are addressed, a few words on potential alternatives are in order. Daraio et al. (2015) differentiate between printed and non-printed outputs of research, which have impacts on either scientific-scholarly, economic and technological or social level. While citations and publications count to the printed outputs, the authors mention datasets, products or artwork as examples for non-printed outputs of research. A third output belonging to the former category are patent applications, which recently received significant attention. Patent applications deserve special attention here, because in the literature a potential substitution effect of patents and publications is discussed for the engineering sciences.

In this work patents are not considered though, because estimating a third model using only patents as sole output would require focusing exclusively on engineering denominations, for which not enough observations exist to generate an own subsample. Measuring the performance of the latter together with all-sciences universities with a small share of engineering departments or chairs respectively would create upward biases for polytechnic universities. In simpler terms, patent numbers depend on (specialization) concentration in a particular field.[42]

While it is justified to neglect patent applications as an additional output, it would have been desirable to include the number of students enrolled as second output into additional models accounting for both research activity and teaching. Previous studies that considered numbers of students in their model (e.g., Catalano et al. 2019; Daraio et al. 2015a; Daraio et al. 2015c) have claimed a trade-off between focusing on research activity and fulfilling teaching requirements. Certainly, it would have been informative to compare the efficiency models using one research output with models considering enrolled students as additional teaching activity output.[43] Even when left out of the efficiency models, student numbers could have been examined as additional external factor to illuminate possible interactions e.g., of teaching responsibilities and task division. Undeniably, not accounting for the number of students is a weakness of this analysis and needs to be critically addressed when discussing the results and potential tracks for future research.

Although detailed information on the bibliometric perspective on performance measurement is provided in section 3.1.2.2, some information on the properties of publications and citations should be examined prior to a careful implementation. In general, an issue related to publication databases is incomplete coverage. Here, the documentation of publications and citations in the pre-war period is expected to be less complete than the documentation in the post-war period. Also,

[42] Supposing the substitution effects of patents and publications are comparable for the engineering departments of the universities, the effect of patents could partially be captured by the specialization concentration variable.

[43] Unfortunately, the sources used to construct the data set did not contain reliable data on enrolled students. While a few entries of the *Minerva* publication indeed provided numbers of students, surprisingly this was not the case for the academic calendars and course catalogues used to cover the period 1970 to 2020.

it is well-known that publication databases only just recently started to catch up on the integration of monographs for example (Bonaccorsi et al. 2017). This problem is further enhanced by the shift of publication habits over the course of time, supposedly substituting enlarged monographs with more frequent publication of shorter notes and papers. The latter is further aggravated by the circumstance that publication habits are potentially unevenly distributed across the university systems and regions considered in the analysis. Finally, completeness of coverage also differs according to database and disciplines. (Bornmann and Mutz 2014)

Some of these issues are alleviated by the scope of the analysis being limited to the branch of natural sciences and the estimation of efficiency models according to separated pre-war and post-war samples. Here in the post-war period, we expect that regardless of university type (region, or language) the vast majority of relevant findings with sufficient impact were published in article form (, not monographs), fully available in publication databases. Even if some monographs attracted substantial attention (, let's say as standard works or introductory textbooks), groundbreaking findings were most certainly published in research articles before, as soon as the empirical results were considered final.[44] Securing priority and striving for the related credit is an important and well-understood mechanism within the scientific community (see 2.1.1). In conjunction with the requirement for a good comparability of the output numbers across the institutions, the publication numbers employed as outputs are not based on all documents affiliated with a university (e.g., books, book chapters, editorials or conference papers) but limited exclusively to the number of research articles. (Van Raan 2019)

Another issue relating to using publications and citations for performance measurement are lagged impact of publications. Van Raan (2019) calls it the delay problem, that even though peers might be

[44] Since monographs might be crucial elements of theory building and the 'progress' in the humanities and social sciences, this argument is only valid for the natural sciences disciplines. One may for example consider the work of Michel Foucault, consisting of both monographs as well as journal articles, amounting to a total of 1,349,822 citations according to google scholar. (Of course, though, monographs of Foucault are supposedly completely covered in all relevant publication databases.)

aware of new publications containing relevant findings, it takes them some time to incorporate them into their own work which creates a delay between awareness and bibliometric notification (the actual referencing of a work). Chu and Evans (2021) obtained a median time of 9 years for top-tier publications to catch up to the most cited of a field. Further, dissemination time is substantially shorter in large fields, where promising articles tend to shoot to the top fast (Chu and Evans 2021). Most authors in turn propose that the average peer review time cycle is no more than three to four years and choose i.e. 4-year windows for citation analysis (Bongioanni et al. 2014; Van Raan 2019).

Effectively, it is unclear whether time-lags are a fundamental problem in citation analysis or not. Nonetheless, rare cases of the so-called sleeping beauties are well-documented. Sleeping beauties suffer from an absence of recognition for several years before they get rediscovered and suddenly generate substantial recognition within the community. While the latter might indeed cause serious issues in bibliometric studies (, especially when focusing on individual scientists' performance), in this work the problem is restricted to the last decade of the sample (2010 to 2020). For articles published e.g. in 2020, it is possible that they gain disproportionately more citations in the upcoming years than publications of earlier periods. For the remaining periods all considered publications should approximate their maximum potential in recognition (and only gain proportionately if they are for example seminal works, which are relevant up until today and still get regularly cited). (Van Raan 2019)

We follow the approach of Lin et al. (2022), which used the number of papers accumulated over the course of a decade. This fits the input data used in this work and should further also account for lagged publication and outlier years (in particular of the pre-war period, where the absolute number of publications or citations of a young university could be close to zero). Finally, all problems linked to the performance measurement (arising at a statistically low aggregation level) when individual researchers' performance is examined (using citation-based measures e.g., the h-index), are of no concern in our setup since efficiency is assessed on aggregated institutional level.

The output data is built on 2,681,761 publications with a total of 148,576,471 received citations and stems from Elsevier's Scopus database (https://www.scopus.com). Initially launched in 2004, Scopus ended the monopoly of Clarivate Analytics' Web of Science (WoS) and developed to an equally high-quality database now interchangeably used with WoS in Scientometrics' studies (Van Raan 2019). For each university, the sum of natural sciences' paper publications and linked citations affiliated with a university is extracted according to each decade. The Scopus database offers two options to choose which documents belong to an institution. The option 'affiliation only' contains publications that are directly associated with the institutional profile. The alternative 'documents, whole institution' additionally considers publications from affiliations contained within its hierarchy. For the example of the LMU Munich the number of documents of the whole institution amounts to 214,449, whereas 186,808 documents are listed under the affiliation only option. In its affiliation hierarchy, we find a number of specialized research centers (e.g., Munich Center for Integrated Protein Science), clusters (e.g., Munich Cluster for Systems Neurology), affiliated schools of the Max-Planck society (e.g., Max Planck Research School for Molecular Life Sciences) and the LMU clinic., the clinic alone is involved in publication of 70,088 documents.[45]

While in the latter example, institutionally separated research organizations only contribute a small share to the overall number of publications (13,301 in total) we advocate to integrate them in this work in the performance analysis of universities. It has previously been pointed out that university rankings lack a consideration of the circumstance that in the continental European research landscape, a lot of resource-intensive science is conducted in research organizations (e.g., belonging to the Max-Planck society) that even though affiliated with universities are treated as separate organizations (, with own academic and administrative staff and funding resources not belonging to the university). In the here employed model though, professorial staff input is considered. Since professors working in clinics or research societies are necessarily linked to the universities (, because of the uni-

[45] Note that the gap of 'affiliation only' and 'whole institution' documents is smaller than the number of publications affiliated with the LMU clinic. Seemingly, a lot of publications are co-authored by authors associated with the clinic and authors associated with the university. Analogously, some of the authors could be affiliated with clinic and university (e.g., full professors with teaching responsibilities), while there is only a small number of research staff affiliated exclusively with the clinic.

versities' monopoly for issuing the title professor), numbers and denominations were equally documented for the input data.[46] The heterogeneity in other academic staff, administrative staff and funding resources is simultaneously treated like the heterogeneity between professors working in different research fields. The inclusion of affiliated institutions is thus uncritical and might help mitigate the supposed spread in performance between European and other sample institutions.

The number of documents was limited to the following Scopus subject areas belonging to domains of engineering and natural sciences: *Physics* and *Astronomy*, *Engineering*, *Biochemistry*, *Genetics* and *Molecular Biology*, *Materials Science*, *Chemistry*, *Medicine*, *Earth* and *Planetary Sciences*, *Chemical Engineering*, *Neuroscience*, *Energy*, *Environmental Science*, *Immunology* and *Microbiology*, *Agricultural* and *Biological Sciences*, *Pharmacology*, *Toxicology* and *Pharmaceutics*, *Health Professions*, *Nursing*, *Veterinary Sciences*, *Dentistry* and *Multidisciplinary*.[47] In accordance with the definition of the term multidisciplinary in chapter 2, the subject area refers to journals publishing articles on different topics within a discipline or field like i.e. *Nature* and *Science* (Rousseau et al. 2019). Since papers published in these journals belong to the natural sciences, they should principally be considered. In case of the LMU Munich *Multidisciplinary* articles make up for about two percent of the publication output, making it desirable to include the category into the analysis.

Technically though, there might be articles belonging to the category *Multidisciplinary*, which span over multiple research fields in the social sciences for example, having no connection to the natural sciences at all. In addition, the content of the articles belonging to the category need not be inherently multidisciplinary (Moschini et al. 2020). An

[46] The data sources used to document denominations and numbers of professors explicitly contained additional information if a professor was associated with a particular department or institute of a clinic or chaired the clinical department.

[47] Consequently, the categories *Social Sciences*, *Arts* and *Humanities*, *Economics*, *Econometrics* and *Finance*, *Business*, *Management* and *Accounting*, *Decision Sciences*, *Psychology*, *Mathematics* and *Computer Science* are not included. Since the operator 'limit to' was used (instead of 'exclude') articles i.e. belonging to the latter two categories might still be accounted for when they are simultaneously listed in the category *Engineering* for example, signaling that the content of the article is not restricted to the formal sciences branch.

'eye-test' for a variety of universities confirmed though that most publications affiliated with this category are primarily connected to the natural sciences. Further, the categories affiliated with an article are not exclusive, but multiple categories can be linked to a paper in case the content matches a particular subject area. So, limiting the publications of e.g., the LMU Munich to the *Multidisciplinary* category returns other subject categories the papers are affiliated with as well. Of the 3,945 Multidisciplinary articles over the whole period recorded, 1,041 where also affiliated with another category. Of the latter, only 4 publications were affiliated with categories not included in the natural sciences branch (e.g., *Arts* and *Humanities*). Given the focus of this work on coordination of subject areas and research fields (, potentially resulting in more *Multidisciplinary* produce), papers belonging to the *Multidisciplinary* subject area are thus included in the output data set.

4.2.1.3 Outliers

As detailed in the methodology section, the partial frontier approach allows to construct frontiers that are robust towards the inclusion of outliers. In this work, outliers could stem from two different sources. There could be outliers due to data collection errors in the input data set or the Scopus database employed for publication and citation output. As a result, there could be measurement errors resulting from incompatibility of the two different data sources, when for example the output values are correct, and the institutional data is flawed or the other way around. This in turn could result in distorted productivity ratios dominating the attainable sets. Secondly, as pointed out earlier, using the partial frontier approach is also employed to account for potential conceptual outliers, meaning to prohibit the comparison of observations with observations operating under different conditions or a different technology.

As opposed to the latter issue, data collection and measurement errors can in the one input, one output case also be addressed using conservative methods for outlier detection. Since this is not the case for the conceptual outliers, it might be reasonable to get rid of any measurement and data collection errors prior to conducting the actual efficiency analysis and thus kind of 'reserving' the capacity of the partial frontier approach for the detection of conceptual outliers. In figure 16, the ratios of publications and citations to professorial staff input are given for each institution at a point in time t (from earliest (left) to latest (right) period) according to clusters.

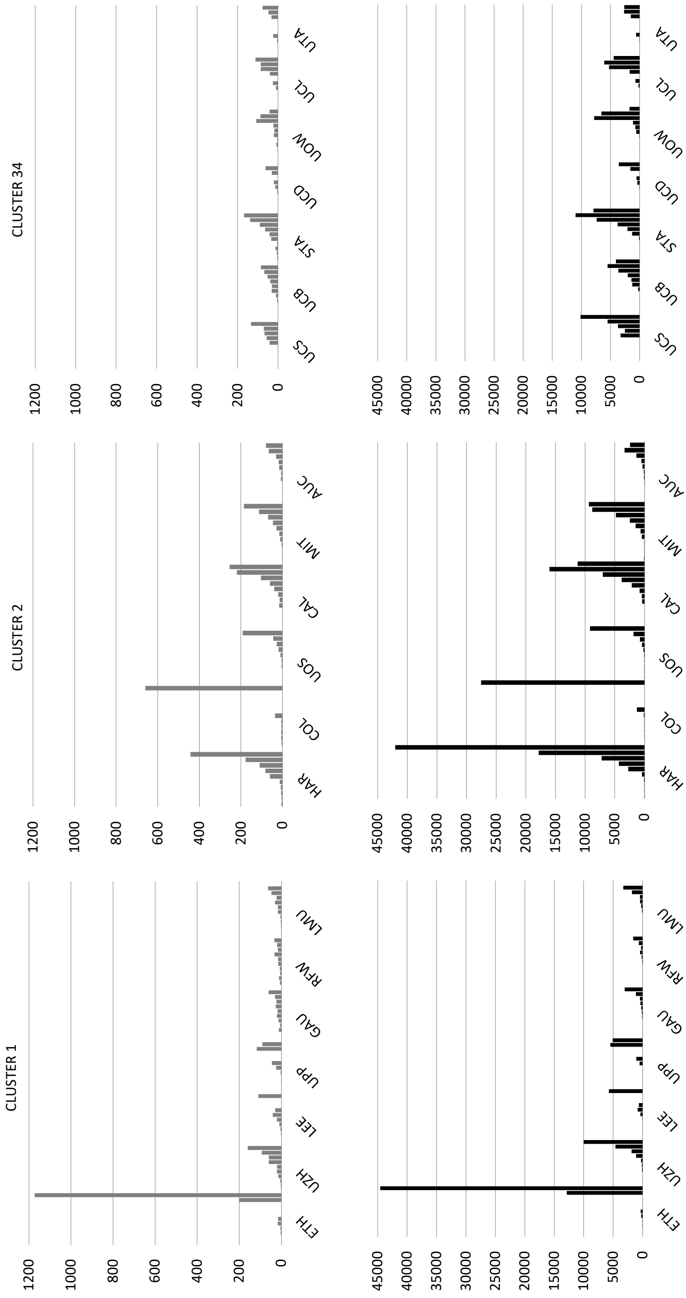
It is easy to obtain that professorial productivity in generating publications and citations increased for institutions over the course of time. This might be related to the increasing size over the considered period, as well as the heterogeneity in endowment with non-professorial staff, which we cannot account for. Both partial productivity of publications and citations follow similar trends for each institution, with the latter pattern revealing more steep inclines and some abrupt changes, capturing the expected cumulative advantages (*Matthew effect*) linked to receiving citations. Interestingly, we can obtain that citation productivity peeks in the decade 2000 to 2010 and slightly declines for the last sample period in the institutions, for which the period 2010 to 2020 is available. This supports the idea of the importance of late recognition of publications (*sleeping beauties*) and is opposed to the idea that most citations are received by a paper within three years of publication.

On average, for the full sample, a professor is involved in the production of 51 publications generating 2,569 citations per decade. These numbers are in line with findings of previous studies and seem realistic for professors working in the natural sciences branch and for the sample of excellent universities (Bornmann and Mutz 2014). (See table 15 for descriptive statistics of inputs, outputs and productivity ratios of the full sample)

Tab. 15: Descriptive statistics of inputs, outputs and productivity ratios of the full sample

Obs. = 169	Mean	Std. Dev.	Min	Max
Sum of Professors (Nr.)	273.67	300.22	3	2,575
Publications (Nr.)	15,868	24,524	3	128,406
Citations (Nr.)	879,150	1,617,768	2	12,188,913
Publication Productivity	50.65	114.21	0.14	1,173
Citation Productivity	2,569	5,820	0.07	44,569

Fig. 16: Nr. of Publications (Grey) and Citations (Black) per Professor According to Institution and Period



The accumulated productivity ratios give us an intuitive understanding of the productivity development of an institution over the course of time, with institutions like Harvard seemingly continuously increasing their productivity levels, while e.g., the German institutions reveal constant productivity rates in the post-war period.

Regarding publication productivity, ratios of the observations of the Columbia university (660) and ETH Zurich in 2010 (1173) and the Harvard university in 2000 (443) (last bars in the chart) do not match with productivity ratios of peers in 2010 and 2000 respectively. In addition, the values of the former two (, in particular the ETH Zurich) do not fit their intra-institutional productivity development pattern, suggesting that here there is indeed an incompatibility of input and output data. Looking at the citation productivity ratios this picture is confirmed (COL 2010: 27,557, ETH 2010: 44,569, HAR 2000: 42,031). For all three observations the publication and citation productivity value lie beyond the range of three standard deviations outside of the sample mean with 393 publications and 20,032 citations per professor and decade respectively. Consequently, they are classified as outliers and excluded from the efficiency analysis. (While the latter numbers for the threshold are certainly high, given adequate endowment with staff, resources and international collaborations, they should still be attainable.[48])

As an alternative, two standard deviations outside of the mean could be chosen as a commonly used threshold (279 publications and 14,211 citations per professor and decade respectively). Applying the latter, the citation productivity values of Caltech in 2000 and Harvard in 1990 would be classified as outliers as well. Looking at both charts (bars second to the last), an outlier role could indeed be supposed for citation productivity yet not for publication productivity. This suggests that here we might obtain one of the issues related to using citation numbers for performance measurement. For the Harvard university, for which an outstanding reputation (even when compared to other excellent universities) can be assumed, we might observe the institutional equivalent to the *Matthew effect*, while the last sample observation of

[48] Keep in mind that while we are relating the number of publications to the professorial staff input this does not mean that all publications are necessarily co-authored by them.

the Caltech might be subject to the delay-problem distorting its pattern of citation productivity development. In any case, since the publication productivity pattern seems normal for both observations, they cannot be classified as measurement or data collection errors and are thus included in the efficiency analysis.

Although all other observations lie within these thresholds, there are some remaining observations which could also be considered as odd either regarding conspicuously low productivity ratios or not fitting the intra-institutional pattern. This concerns for example the ratio of the University of Sydney in 2000 (192 and 9,200) or the University of Washington in general, which abruptly increases at one point and afterwards, as opposed to the overall pattern of all institutions, decreases again in productivity. While those values might be odd, there could be structural issues we have no information on, causing those pattern shifts. Given they are within the commonly employed standard deviation range for outlier classification, these observations are not excluded as outliers and regularly enter the efficiency analysis.

4.2.2 External factors

4.2.2.1 DoL – task division and task coord. (factor)

Choice and calculation of the external factors corresponds largely to the remarks in section 4.2.3. The two variables used to portray the effect of division of labor on the scientific production process are task division and task coordination. Task division is given by the sum of professors divided by the number of denominations. The highest task division on institutional level is thus returned by a value of one. Given a fixed number of professors and presuming that a denomination demarcates a professorial chair or department dedicated to an institutionalized task (topic), the tasks linked to a production process of a university can only be divided to the extent, where each professor is assigned a distinct denomination (task division = 1).

On the other hand, high values of the task division measure reveal a strong prevalence of team science within an institution. Given a fixed number of professors, dividing the production process in less denominations than professors available (low task division on the institutional level), results in greater numbers of professors working on each institutionalized task (higher division of labor in teams). To summarize, values of 1 show maximum feasible task division on institutional level, while higher values indicate low task division on institutional level (, where task division is shifted to the team or department level). Of course, the measure does not account for the absolute number of denominations an institution covers and the concentration of the professors according to the latter denominations, which will be accounted for by the specialization concentration measure in the upcoming section.

The second external factor that will be employed to analyze division of labor, is task coordination. As opposed to task division, task coordination tries to measure the coordinative quality of a denomination. In plain language, it measures how big the scope of an institutionalized task is. A denomination dedicated to the task 'organic and inorganic chemistry' needs to consider a broader area of research than 'organic chemistry' for example. On the level of denominations, the measure returns how many topics on average one denomination bridges. On the level of the WoS categories it measures how many distinct subject areas, research fields and disciplines a denomination within an institution on average connects with one another.

The two measures of division of labor thus measure in how many distinct tasks the production process of a university is institutionalized in and how well those individual tasks are connected by the domains they cover. As outlined in the theoretical part of this work, those effects are believed to interact with one another and influence efficiency. For one, we would expect higher values of task division (more team science) to be favorable for efficiency, whereas we would expect low values of task division (high task division on institutional level) to be favorable only when tasks are also highly coordinated (to keep coordination costs low).

Further, we might expect differences according to the time periods considered. For pre-war science, the role of task coordination could be unclear or less pronounced than in the post-war period. In this (growing) phase of science we would expect that task division on institutional level is in principle favorable and does not require for a high coordination of the newly institutionalized tasks. Given the lower absolute numbers of professors a lot of the coordinative requirements linked to the latter could have been naturally handled by local proximity or discourse.

In table 16, the variables available to measure division of labor (and spec.) according to different levels of granularity are reproduced. Both task division and task coordination could be measured on the level of the individual denomination or the levels of the WoS subject area, research field and discipline it was assigned to. In the case of task coordination, the indicators of these levels indeed provide qualitatively different information. On average, coordination of subject areas could be very high, yet the different disciplines the institutionalized tasks are attributable to, could still be isolated from one another. On the contrary, some institutionalized tasks could coordinate different disciplines very well yet within those disciplines the coordination of subject areas could still be low.

Tab. 16: Replication of Tab 5 in section 3.2.4 with variables implemented as external factors printed in bold

	Division of Labor		**Specialization**	
	Task Division	Task Coordination	Concentration	Gravity
Discipline		[*Nr. of disciplines covered (weighted)*]	[H_{DISC}]	Individual specialization of denomination (weighted)
Research Field		Nr. of research fields covered (weighted)	$\mathbf{H}_{\boldsymbol{REFI}}$	
Subject Area		Nr. of subject areas covered (weighted)	$\mathbf{H}_{\boldsymbol{SUBJ}}$	
Denomination	Nr. of professors / denomination	Nr. of topics covered (weighted)	$\mathbf{H}_{\boldsymbol{DENO}}$	

For the task division measure on the other hand, it was established that the insights on the aggregated level of the subject area or the discipline are rather limited. In accordance with the procedure for conducting the cluster analysis, for task coordination a factor variable is constructed loading the information from the different aggregate levels in one joint measure of task coordination.

As opposed to the approach of the cluster analysis, here the values are not aggregated to the institutional level for the whole period (, by weighting all values of denominations assigned to an institution *i* with the number of assigned professors relative to the sum of professors associated with the institution over the whole sample period). Instead, they are aggregated to the level of the individual observations (institution *i* at point in time *t*) entering the efficiency analysis. Hence the variables are calculated based on the denominations assigned to a university *i* at a certain point in time *t*, weighted by the sum of professors assigned to each denomination relative to the sum of professors assigned to *i* in *t*.

On the level of individual observations, the overall Kaiser-Meyer-Olkin criterion (0.73) suggests that the measures for task coordination of the individual denomination, subject area and research field can reasonably be loaded into one joint factor. (See table 17 and 18 for the results of the factor loading procedure.)

Tab. 17: (Unrotated) principal component analysis for the task coordination variables

Obs. = 166	Eigenvalue	Difference	Proportion	Cumulative
Factor 1	2.637	2.644	1.023	1.023
Factor 2	-0.007	0.046	-0.003	1.021
Factor 3	-0.053	.	-0.021	1.00

Tab. 18: KMO-criterion, factor loadings, unique variances and predicted scores for the task coordination factor

Obs. = 166	KMO (0.731)	Rotated Factor Loadings (F1)	Unique Variances	Scoring coeff.
Task Coord. – Topics	0.837	0.895	0.198	0.141
Task Coord. – Subject Areas	0.656	0.972	0.054	0.605
Task Coord. – Research Field	0.730	0.943	0.111	0.258

Even though task coordination on the disciplinary level correlates positively and significantly with all other variables employed (0.56*** with topics, 0.59*** with subject areas and 0.73*** with research fields covered), it cannot be loaded into the same factor and thus would have to be considered as individual external factor. Presumably, since the correlation with the other task coordination variables is fading with increasing granularity, the information conveyed by disciplinary coordination is supposedly not accurate enough to fit the factor, contributing evidence towards the idea that thinking about scientific collaboration in terms of the very broad scope of disciplines might not be the best fit.

To keep the input-output space as small as possible (and limit the effect of the curse of dimensionality (Daraio et al. 2015a)) the task coordination factor is built on the information conveyed on the level of the individual denomination, subject area and research field. In table 19, descriptive statistics for task division and task coordination (factor) are provided for the full sample.

Tab. 19: Descriptive statistics for external factors representing DoL in the full sample.

Obs. = 166	Mean	Std. Dev.	Min	Max
Task Division	5.95	7.98	1	44.40
f. Task Coordination	0.00	0.98	-1.63	5.09

For the task coordination (factor) we can obtain that the negative pole is deviating less from the mean than its positive counterpart indicating that very high task coordination at all levels is more likely than no task coordination at all levels. Regarding task division, we can see that the extreme poles range from observations which operate at a maximum task division on institutional level (task division of 1) to universities where on average 44 professors share the same denomination working on one joint institutionalized task (, supposedly as teams in large departments). On average 6 professors share the same denomination, who could either be organized in a large professorial chair or a small department.

4.2.2.2 Spec. Concentration (factor) and gravity

In order to measure the influence of specialization on efficiency, two variables were defined in section For one, specialization can be thought of as the concentration of resources (here: professorial staff) on tasks, sort of reflecting topical focuses of an institution. This was operationalized by using the Hirschman-Herfindahl index as a measure of absolute concentration of numbers of professors on denominations, subject areas, research fields or disciplines according to institution. In accordance with economic theory, we would in principle expect higher concentration on tasks to be favorable. The authors of the science studies in turn highlight the importance of diversity and combinatorial novelty, indicating that there might exist a trade-off here between efficiency gains through concentration versus diversification.

Secondly, in context of scientific production, another aspect linked to specialization is the increasing depth in demarcating tasks from one another by focusing on narrower and narrower scopes of research. This was operationalized by assigning each individual denomination a value according to the specificity of its scope, where 1 is corresponding to a subject area in the WoS subject category, 2 is demarcating a narrower research topic fully contained by a subject category and 0 signaling that a denomination demarcates an area of research that is broader than the subject categories it connects. Here again, as opposed to the procedure of the cluster analysis, mean values for specialization variables were calculated for each observation (institution *i* at point in time *t*). In accordance with economic theory, we would expect a higher depth in specialization to be favorable for efficiency. Analogously to task division though, specialization depth increases coordination costs. So here too, specialization depth might be only favorable to a certain extent, where the productivity gains are outweighed by the increased coordination costs and knowledge burden.

The procedure for the construction of the specialization concentration factor variable is identical to the construction of the task coordination factor introduced in the previous section. In table 20 and 21, the results for factor loadings are presented. The overall KMO-criterion (0.73) suggests a similar performance of the factor in representing its components when compared to the task coordination counterpart.

Notably though, here the uniqueness of all three components is more pronounced than in the latter case. This particularly concerns concentration on the level of the research field, which with a variance of 0.96 not shared with the other variables makes it the least influential component of the factor.

Tab. 20: (Unrotated) principal component analysis for the Spec. Concentration variables

Obs. = 166	Eigenvalue	Difference	Proportion	Cumulative
Factor 1	1.377	1.385	1.161	1.161
Factor 2	-0.008	0.175	-0.007	1.154
Factor 3	-0.183	.	-0.154	1.000

Tab. 21: KMO-criterion, factor loadings, unique variances and predicted scores for the Spec. Concentration factor

Obs. = 166	KMO (0.731)	Rotated Factor Loadings (F1)	Unique Variances	Scoring coeff.
Spec. Conc. - Topics	0.515	0.819	0.329	0.467
Spec. Conc. – Subject Areas	0.515	0.815	0.337	0.453
Spec. Conc. – Research Field	0.854	0.206	0.957	0.050

Just like in the case of the task coordination factor, concentration on the level of the discipline cannot be loaded into this factor and would thus have to be considered as own external factor. Here though, we do not obtain positive correlations with all concentration measures on the other levels. Indeed, we find a low (yet not significant) negative correlation with concentration on subject area level (-0.12) and a moderate (significant) negative correlation with concentration on the denomination level (-0.19**). Since there is a strongly positive significant correlation with concentration on research fields (0.72***), we presume that the information on disciplinary concentration is to some extent conveyed by the inclusion of the research field level in the factor variable.

Potentially though, here two factors, one based on the concentration on the granular level (topics and subject area) and one on the broad level of disciplinary profiles (research field and discipline) could have been used as an alternative instead. Since this is conflicting with the aim to restrict the curse of dimensionality, we decided to use only one (factor) variable as external factor accounting for specialization concentration. In table 22, descriptive statistics for the two specialization variables are provided. Here too we can obtain that the extreme poles for the concentration measure are skewed towards the maximum signaling that it is more likely that institutions concentrate highly on all levels of granularity than not concentrating at all.

Tab. 22: Descriptive statistics for external factors representing Spec. in the full sample

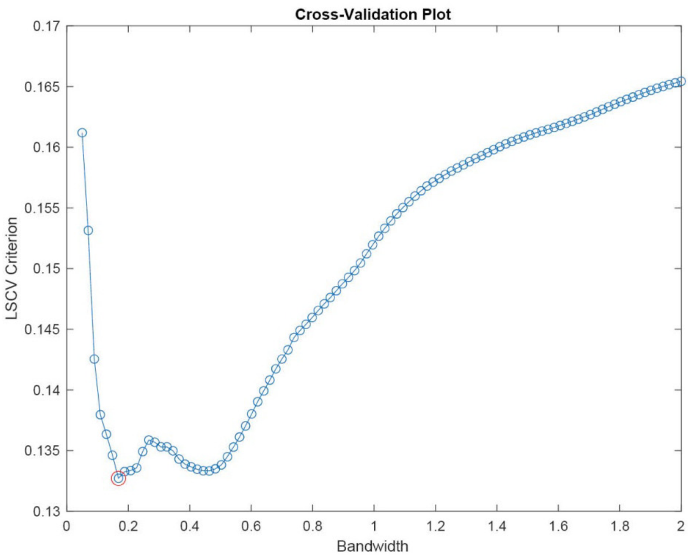
Obs. = 166	Mean	Std. Dev.	Min	Max
f. Spec. Concentration	0.00	0.87	-0.96	6.98
Spec. Gravity	1.13	0.18	0.64	1.60

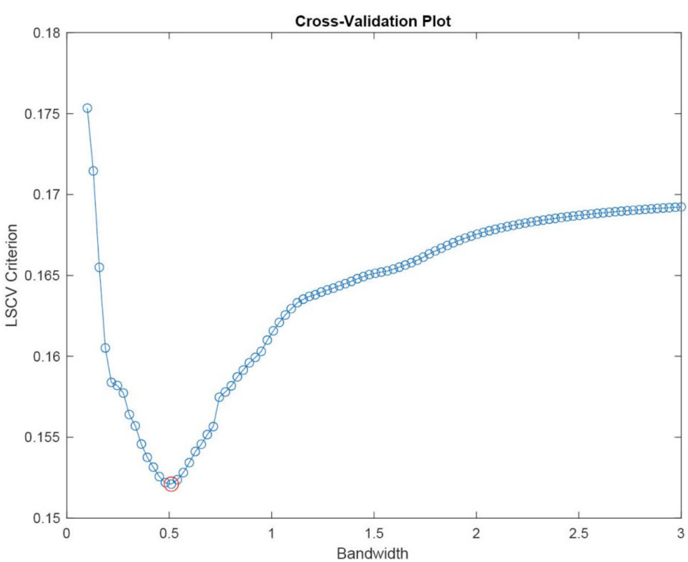
Regarding the depth of specialization, we can obtain that on average a denomination is slightly more specialized than the WoS subject area it was attributed to. In addition, the distribution of the values for Spec.-Gravity seem to be relatively evenly distributed with a moderate spread.

4.2.2.3 Bandwidth selection

As outlined in the methodology section, an appropriate bandwidth needs to be specified for each external factor entering the conditional models. Bădin et al. (2019) proposed an approach that builds on the bandwidth selection process proposed by Li et al. (2013) adapting it for the case of conditional distribution functions. (See 4.1.3.2 for details) Here, their implementation is used to generate cross validation plots for each external factor considered in the efficiency analysis modeling the least squares cross validation criterion as a function of a range of bandwidth values. Even given the cross validation plots, choice

Fig. 17: Cross validation plots for PPM (upper) and CPM (lower) conditional model with task division as external factor.





of bandwidth is not entirely straightforward. Very low or high values, which minimize the LSCV criterion could result in under- or over-smoothing respectively. Looking at the cross validation plots for task division in the conditional publication and citation model provided in figure 18, it can be obtained that the bandwidth that would minimize the LSCV criterion lies relatively close to zero. In order to obtain suitable values, the efficiency models were iteratively calculated using different values selected in accordance with good empirical practice. For all external factors a moderate amount of (two to four) bandwidth values were tested, if they either lied at local minima of the LSCV criterion or at points before sudden increases. In case of the conditional PPM model in figure 17, examples for reasonable values would thus be the absolute minimum marked with a red circle and the local minimum at 0.44. For the conditional CPM variant, the absolute minimum and the two sudden increases in 0.72 and 0.9 were considered as candidates. Finally, bandwidth values were then determined by (1) how good a bandwidth minimizes the LSCV criterion and simultaneously (2) minimizes the number of efficient observations in the conditional models. The latter is of interest, because the analysis of the ratios is mainly informative for inefficient observations of the conditional model.

Here, for the conditional (task division) PP model the latter procedure resulted in the choice of the bandwidth 0.444 (first local minimum). For the CP variant, a bandwidth of 0.715 was selected, which corresponds to the first sudden increase in LSCV criterion obtainable in the bottom cross validation plot. In table 23, all bandwidths selected employing this procedure are reported according to model variant and external factor.

Tab. 23: Summary of bandwidths selected according to external factor and model variant

		PPM		CPM	
		Bandwidth	LSCV	Bandwidth	LSCV
Task Division	pre	0.444	0.133	0.715	0.156
	post	5.394	0.143	5.300	0.141
	full	5.379	0.126	5.326	0.125
Task Coordination	pre	1.636	0.157	1.090	0.156
	post	1.773	0.147	1.888	0.152
	full	2.179	0.158	4.914	0.167
Spec. Concentration	pre	2.000	0.162	2.194	0.164
	post	1.090	0.164	3.970	0.167
	full	2.129	0.166	6.212	0.168
Spec. Gravity	pre	0.455	0.170	0.657	0.169
	post	0.970	0.168	0.292	0.162
	full	1.125	0.168	3.687	0.168

All corresponding cross validation plots are provided in Appendix S2 – S7.

4.2.3 Descriptive statistics

Table 23 presented at the end of the previous section also provides an overview of all model variants that will be analyzed in the upcoming section. An unconditional efficiency measure incorporating publications as output will be calculated for pre-war, post-war and full sample. (The latter serves for the analysis of the university types or rather university clusters, which due to the limited number of observations must be examined in the full frontier case.) Three order-α variants are calculated for each unconditional model (FDH frontier ($\alpha = 1$), robust frontier ($\alpha = 0.99$) and partial frontier ($\alpha = 0.90$) of the inefficiency distribution). Analogously, the same procedure is repeated incorporating citations as output.

Afterwards, the conditional models were calculated, first incorporating the four components as external factors individually. Finally, conditional models for evaluating the joint effects of DoL and Spec.

respectively were calculated employing two external factors (task division and task coordination; Spec. Conc. and Spec. Grav.). In total, this amounts to 54 different sets of efficiency estimates calculated for all sample institutions.[49] Descriptive statistics will now be provided for the pre-war and post-war sample as well as according to clusters. Further, we will take a look at mean values of inputs, outputs and external factors according to institution.

Pre-War Sample

Pre-war sample descriptive statistics for inputs, outputs and external factors are provided in table 24. Clearly, the portrayed values convey the idea of the growth phase of modern science. We see moderate mean values for numbers of professors, publications and citations. Here, on average a sample professor produced 5 publications and 27 citations each decade (, which today would certainly be considered very low values for natural sciences' departments of ordinary universities). The values of the external factors equally reflect the small size of universities in the pre-war sample. In particular, the value of task division corresponds to our theoretical understanding of division of labor.

Tab. 24: Pre-war sample descriptive statistics

Obs. = 55	Mean	Std. Dev.	Min	Max
Sum of Professors (Nr.)	55.78	44.74	3	247
Publications (Nr.)	247.95	307.35	3	1,319
Citations (Nr.)	1,491.27	2,332.84	2	13,193
Task Division	1.39	0.35	1	2.84
f. Task Coordination	-0.07	1.02	-1.58	5.09
f. Spec. Concentration	0.05	1.06	-0.68	6.98
Spec. Gravity	1.15	0.12	0.90	1.37

[49] Effectively, the number of model variants calculated was substantially higher since different efficiency models had to be evaluated in the bandwidth's selection procedure. Further, different specifications of α had to be tested to find a suitable value for the robust frontiers and the frontier of the inefficiency distribution. Finally, for robustness reasons, potential interactions between the DoL and Spec. components were tested.

Here, the institutionalized division of tasks is close to the maximum value (of 1) with the biggest mean 'team' size observed for an observation lying at 2.84 professors per denomination. The mean value for task coordination is slightly negative, potentially indicating that given the overall small size of the universities in the pre-war period, controlling for coordination costs is subordinate to the gains in productivity induced by higher institutionalization of task division.

The specialization concentration measure on the other hand is slightly positive on average. Supposedly, diversity of domains requires for sufficient institutional size, forcing a certain concentration of resources on domains upon the smaller institutions in the pre-war period. Finally, looking at the specialization depth we can obtain a moderate mean value close to 1. Indeed, in the pre-war sample the maximum value for specialization gravity lies below 1.5, meaning that even in the most specialized institution the average denomination does not demarcate a narrower range than a WoS subject area.

When looking at the mean values according to institutions (see table 25), we can further see that comparatively high specialization gravity values tend to belong to polytechnic universities (ETH and MIT) and larger institutions (LMU and HAR) except for the University of Auckland. The latter accounts for extreme values in inputs, outputs and external factors (with the maximum values of institutionalized task division, specialization concentration and gravity). Further, the University of Auckland has only one appearance in the sample. In conjunction with the extremely low input values, this should increase the probability of the latter serving as its own peer, which would be desirable given our aim of only making measuring performance under comparable technologies. This will thus be evaluated when interpreting the results of the efficiency estimates.

Tab. 25: Pre-war values for inputs, outputs and external factors according to institutions

i	X	Y_PUB	Y_CIT	task div.	f.task coord.	f.spec. conc.	spec. grav.	app.
AUC	3	5	15	1.00	0.92	6.98	1.33	1
UTA	17	17	76	1.13	0.09	0.91	0.95	3
CAL	19	308	7441	1.36	0.26	1.25	1.00	1
UOW	23	92	177	1.41	-0.31	0.55	1.00	3
UPP	27	53	190	1.09	0.28	-0.14	1.16	3
LEE	31	165	812	1.12	-0.36	-0.19	1.10	3
UZH	34	197	1398	1.06	1.94	-0.27	1.31	4
UOS	35	34	71	1.21	-0.44	-0.09	1.06	4
ETH	43	109	1643	1.22	0.57	-0.28	1.22	3
RFW	46	347	489	1.57	-0.12	0.02	1.06	4
GAU	47	506	2462	1.32	-0.29	-0.37	1.10	3
STA	50	392	1265	1.80	-1.11	0.38	1.05	3
MIT	63	211	1798	1.49	-0.72	-0.10	1.29	4
UCB	74	192	1778	1.32	0.19	-0.47	1.19	4
LMU	86	264	617	1.51	0.51	-0.29	1.24	4
HAR	117	755	4729	1.56	-0.59	-0.47	1.25	4
COL	128	703	2794	1.93	-1.18	-0.31	1.10	4

Quite surprisingly, we find that regardless of the assumed dominance of pre-war science by European institutions, this is not necessarily expressed by the size of the institutions. Just like in the post-war period the two biggest observations are US institutions with the Harvard University and the Columbia University in New York. This might indicate, if only concerning sheer size, that the hegemony of European institutions in science already began to crumble in the first half of the 20th century.

For sure though, it should be noted that the latter two are among the oldest institutions in the US. Also, the gap between the largest European institution (LMU Munich) in the sample is still moderate when compared to the two. In addition, as was pointed out before we are unfortunately not able to consider the numbers of students enrolled, which of course could be drivers of institutional size in (then already) huge metropolitan areas like Boston and New York City. Finally, the smallest European institution is the Uppsala University, which still dominates four other institutions in size, confirming the idea of the European institutions being the established and dominating ones in the pre-war period.

Post-war sample

Post-war sample descriptive statistics for inputs, outputs and external factors are provided in table 26. In comparison to the pre-war sample, we see the results of the enormous growth in modern science. Mean sum of professors here is 385, which is about one hundred professors more than the mean value of the biggest pre-war institution. We can further note that the spread of minimum and maximum observation of number of professors here too lies at a factor of ca. 100, again validating the idea of splitting the sample in two subsets divided by the event of the second world war.

Tab. 26: Post-war sample descriptive statistics

Obs. = 111	Mean	Std. Dev.	Min	Max
Sum of Professors (Nr.)	384.68	316.07	22	2575
Publications (Nr.)	21,460	24,022	160	111,186
Citations (Nr.)	1,171,480	1,504,299	2,042	6,688,846
Task Division	8.21	8.94	1.06	44.40
f. Task Coordination	0.03	0.96	-1.63	2.88
f. Spec. Concentration	-0.03	0.76	-0.96	4.47
Spec. Gravity	1.12	0.21	0.64	1.60

Further, when looking at the partial publication and citation productivity ratios, we obtain an increase by the factors of about 10 and about 100 respectively. That citation productivity equals the squared value for increase in publication productivity reveals the shift in publication behavior and reflects the absolute growth of the scientific community, which may recognize a publication and cite it.

When looking at the summary statistics for the external factors, we may also obtain significant differences to the pre-war period. For one, task division has clearly shifted towards task division in teams with an average of 8 professors and a maximum of 44 professors sharing an identical denomination. In turn, we can obtain that even in the post-war period there are still observations operating close to maximum institutional task division (1.06). For task coordination we find a positive mean value and a moderate spread supporting the idea that the latter is more important in enlarged institutions.

Accordingly, the specialization concentration value is slightly negative, supporting the idea that larger institutions enable greater diversity. Finally, the value for specialization gravity marginally decreased. This is in total opposition to the theoretical remarks, since we would have expected the depth of specialization to increase over the course of time. It should be noted though that while the mean value decreased, the standard deviation and maximum value substantially increased in comparison with the pre-war values. This at least supports the idea that with time, more specialization depth is feasible. Ideally, we find structural differences for the latter when looking at the different institutions.

Indeed, looking at the results provided in table 27, we find increases in specialization depth for all European institutions when compared to the pre-war values. For non-European institutions we can obtain the opposite trend. Whereas in the pre-war period high values for specialization depth could mainly be found in big institutions, in the post-war sample it seems to be the other way around with the three largest universities operating close to a value of 1 and the smallest three (European) institutions nearly approximating the threshold of 1.5. Further, whereas in the pre-war sample specialization depth of the universities of technology, MIT (1.29) and ETH (1.22) was particularly high, it radically decreased for the former (0.89) while shooting up for the latter (1.54). Presumably, this comparison reveals the effect of the shift from the professorial chair system to the department system in the US institutions after the second world war.

Tab. 27: Post-war values for inputs, outputs and external factors according to institutions

i	X	Y_PUB	Y_CIT	task div.	f.task coord.	f.spec. conc.	spec. grav.	app.
UZH	127	10855	525799	1.28	0.55	-0.61	1.48	6
ETH	128	13732	833525	1.19	0.86	-0.82	1.54	3
LEE	135	8333	361356	1.80	0.59	-0.73	1.35	4
UPP	153	13831	701269	1.78	0.22	-0.53	1.36	4
CAL	172	17289	1004088	3.57	-0.88	0.40	1.05	7
AUC	198	9261	335174	6.48	-0.69	0.94	0.95	7
GAU	280	9142	317546	2.05	0.61	-0.71	1.31	6
UCS	288	24493	1729816	8.05	0.15	1.37	0.90	5
UOS	302	13278	536502	5.26	-0.75	-0.23	1.10	6
HAR	320	29120	2225218	4.06	-0.64	-0.13	1.09	5
RFW	414	9839	315799	2.30	0.43	-0.63	1.28	6
STA	423	38148	2333598	14.27	-0.15	0.21	0.91	7
LMU	447	16686	619755	2.76	0.66	-0.52	1.25	6
MIT	454	30564	1836493	8.13	-0.18	0.55	0.89	7
UTA	469	22471	963359	24.89	0.06	0.45	0.89	5
COL	474	5703	170046	5.31	-1.13	-0.33	1.06	2
UCD	514	18202	934578	10.46	0.20	-0.09	1.27	5
UCB	601	30479	1809983	20.48	0.40	0.05	0.99	7
UCL	611	44181	2301450	15.39	1.11	-0.01	0.99	6
UOW	891	37904	2105974	14.46	-0.53	-0.22	1.07	7

The latter is confirmed when looking at the values of task division for the latter two institutions. While the ETH still operates nearly at the maximum of institutional task division (1.28) (, with their professorial staff input increased by about three times), the task division value of the MIT has risen to about 8 professors per denomination on average, which corresponds to a middle-sized department. Supposedly, a lot of specialization depth of US institutions has been relocated from the institutional to the team-level in the post-war period.

There is another notable difference when comparing the polytechnic universities of pre- and post-war period. While for the ETH we find a low value for specialization concentration in the pre-war period and an even lower one in the post-war period, the low concentration value for the MIT in the pre-war period shifted to one of the highest in the post-war period. The latter could be understood as two different strategies of universities. While the ETH used its growth for further diversifying in narrower demarcated topics, the MIT focused its resources more on selected subject areas or research fields yet at a lesser degree of depth. In comparison, for the California Institute of Technology we find a comparatively high concentration in both periods and stable values for specialization depth. One could argue in favor of any one of those three variants and it will be interesting to reconsider them when the joint effect of the specialization variables on efficiency is assessed.

Full sample according to university clusters

Finally, variables according to university clusters are provided in table 28. The remarks regarding the differences between European and non-European institutions are supported for the full sample results according to the three university clusters. Here, we find a structural reproduction for the differences outlined in inputs, outputs and external factors.

For one, the significant difference in specialization depth of the European cluster (1) is confirmed. Further we find the institutions of cluster 1 to operate at the highest institutionalized task division (and thus at the smallest team size). Interestingly, we do also find the highest value of task coordination for cluster 1, challenging the idea motivated in this work that task coordination is not sufficiently accounted for by the self-governed scientific community. Apparently, institutions operating at higher institutionalized task division and specialization depth do also put more effort into coordinating those tasks. The results of the conditional analysis will reveal whether those levels of coordination are sufficient and if the latter equally applies for all institutions of the cluster.

Analogously to the differentiated strategies of the two universities of technology, we find a general trend of the European cluster to diversify and specialize in depth, whereas the other clusters concentrate their resources more on particular subject areas or research fields, yet at a broader scope. Finally, it can be noted that the lowest value for task coordination is not found at the highest but at the medium value of task division. It will be interesting to remember this when the joint effect of division of labor (accounting for both task division and task coordination) on the efficiency ratios is examined.

Tab. 28: Values for inputs, outputs and external factors according to university clusters

i	X	Y_PUB	Y_CIT	task div.	f.task coord.	f.spec. conc.	spec. grav.	
1	174	7029	295759	1.68	0.50	-0.47	1.28	59
2	221	12377	710779	4.17	-0.63	0.29	1.07	52
34	436	24344	1376473	12.22	0.06	0.23	1.01	55

Full sample descriptive statistics can be retraced in sections 4.2.1-4.2.2, where the choice of inputs, outputs and the construction of external factors was motivated.

4.3 Empirical Results

4.3.1 Pre-war frontier

4.3.1.1 Efficiency results

The presentation of the empirical results will be structured in three sections dedicated to efficiency estimated under pre-war (4.3.1) and post-war (4.3.2) technology, as well as for the full frontier case differentiated according to different university clusters (4.3.3). In each of the latter, first efficiency results obtained from the unconditional case are discussed, before the influence of DoL and Spec. on the production process is addressed. In each of those subsections results for both publication productivity model and citation productivity model will be presented at the same time to check whether the obtained effects are robust for two important outputs of the scientific production process.

In table 29, mean values of inputs, outputs and efficiency estimates for both PP and CP model for the pre-war frontier according to each institution are provided. Also, in the last column the number of observations of each institution in the pre-war sample is given. As we can see, most institutions have observations for either three or all four periods. The only two observations that appear only once in the pre-war sample are the University of Auckland and the California Institute of Technology.

Tab. 29: Pre-war period institutional mean of inputs, outputs and efficiency estimates

i	X	Yp	Yc	FDHp	FDHc	Oa99p	Oa99c	Oa90p	Oa90c	APP
AUC	3	5	15	0.00	0.00	0.00	0.00	-0.29	-0.29	1
UTA	17	17	76	0.05	0.03	0.05	0.03	-0.09	-0.09	3
CAL	19	308	7441	0.00	0.00	0.00	0.00	-0.38	-2.53	1
UOW	23	92	177	0.10	0.10	0.10	0.10	-0.08	-0.05	3
UPP	27	53	190	0.14	0.14	0.14	0.14	-0.02	0.00	3
LEE	31	165	812	0.17	0.22	0.17	0.22	-0.07	-0.08	3
UZH	34	197	1398	0.20	0.26	0.20	0.26	-0.12	-0.30	4
UOS	35	34	71	0.29	0.29	0.29	0.29	0.12	0.11	4
ETH	43	109	1643	0.42	0.42	0.42	0.42	0.13	-0.14	3
RFW	46	347	489	0.24	0.48	0.24	0.48	-0.25	0.14	4
GAU	47	506	2462	0.08	0.51	0.08	0.51	-0.49	-0.43	3
STA	50	392	1265	0.26	0.56	0.26	0.56	-0.29	0.05	3
MIT	63	211	1798	0.58	0.79	0.58	0.79	0.13	-0.11	4
UCB	74	192	1778	0.70	0.99	0.70	0.99	0.25	-0.03	4
LMU	86	264	617	0.66	1.19	0.66	1.19	0.23	0.48	4
HAR	117	755	4729	0.34	1.12	0.34	1.12	-0.76	-1.23	4
COL	128	703	2794	0.28	1.31	0.28	1.31	-0.72	0.01	4

Since this makes them potentially distorting observations, they are interesting examples for what can and what cannot be accounted for by the methodology. Both observations are found to be efficient in full and robust frontier model of both PP and CP variant. Looking at the numbers of inputs and outputs this might seem odd. The University of Auckland produced a total number of only 5 publications and 15 citations respectively, whereas the Caltech accounts for 308 publications and 7,441 citations respectively. Taking the partial productivity ratios into account, the University of Auckland would range among the least productive in producing publications and average at best in citational productivity whereas the Caltech's partial productivity would certainly be among the best performers.

Nonetheless, both institutions are classified as efficient, dominating units, which constitute the efficient hull. That is because, operating at an input level of only 3 professors, there is neither in the input dimension nor in the output dimension any other empirical observation available against which the University of Auckland in this period could be compared with. In simpler terms, the institution operates under its own technology. The consequence is that it won't affect the analysis of the external factors in the upcoming section.[50] Clearly, this is desirable since the institution is supposedly just at the beginning of a growth phase and drawing valid general conclusions based on its current task division level is unlikely at best.

The efficiency of the Caltech on the other hand accounts for the highest citations an institution generated on average in the pre-war sample with a very high partial productivity ratio for citation productivity. Intuitively, we might think that this observation should have been classified as an outlier in section 4.2.1.3 (, at least when considering institutional averages). Even though its citational return is certainly high, remember that in table 29 institutional mean values are given and for the Caltech there is only one observation in the sample. (See Appendix S8 for a full table with efficiency estimates of all 55 observations) Higher values for citations do exist,[51] and the latter observation does not exceed either threshold defined for outlier classification. In addition, the partial publication productivity ratio falls in line with other institutional mean values like e.g., the Georg-August Universität Göttingen. All-things-considered, the observation is certainly no outlier in the sense of a measurement error.

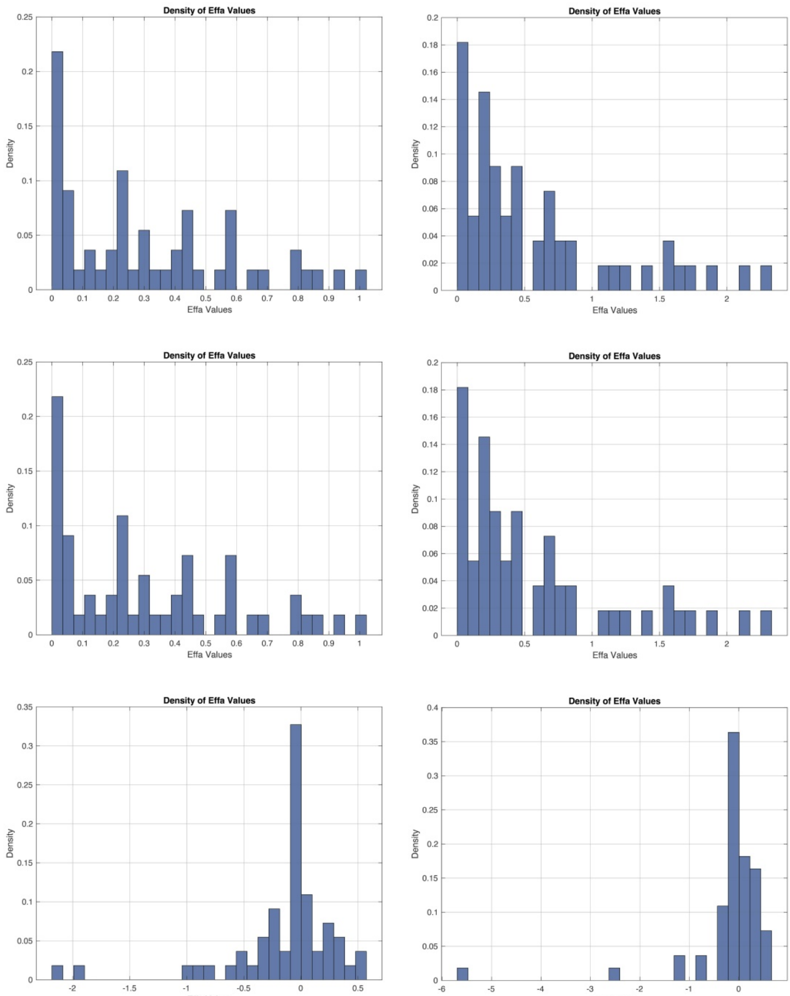
[50] The efficiency ratios calculated to assess the effect of external factors are only informative for inefficient observations of the unconditional measure. They may nonetheless affect the frontier when a consideration of the external factor helps an observation becoming a dominating unit itself.

[51] The Harvard university e.g., produced 13,193 citations in the same period (1920) and is also classified as efficient yet at a different input level.

On the contrary though, in the methodology section the idea was strongly promoted that all observations need to be benchmarked against suitable reference technologies. Consequently, the idea of partial frontiers was introduced to permit a certain number of observations to lie beyond the boundary, which are then classified as superefficient. Given its extraordinarily high citational partial productivity, the Caltech could in principle qualify as an observation against which other observations (i.e. of earlier periods) cannot reasonably be compared. Classifying the latter as superefficient would allow to measure efficiency under differently specified technologies (frontiers) and enable us to account for potential mismatches in comparison groups when evaluating the effect of external factors. Looking at the histograms portraying the distribution of efficiency values (in figure 18) though, we find that in the pre-war sample the robust frontier is identical to the full frontier. Here, there is no conceptual outlier excluded from the boundary. Apparently, the chosen percentile for alpha (= 0.99) is not low enough to permit observations to lie beyond the pre-war frontier.[52] The inefficiency distribution in turn, portrayed in the lower two figures shifts half of the observations beyond the frontier. This seems to be a good fit. Looking at table 29 again, in the inefficiency frontier model (alpha = 0.5) the Caltech is classified as highly superefficient with a value of -2.53. Here, the remaining observations are benchmarked against a technology that is not constituted by the Caltech (, but by less productive observations). So, when evaluating the effect of external factors on efficiency special attention will be paid to a comparison of the full and robust frontier case with the partial frontier case. If we find those results to be consistent, then the Caltech observations did not decisively distort the obtained effect. (We will see that in the post-war and full frontier case according to clusters the procedure works equally in case of the robust frontier (alpha = 0.9) enabling three different layers of protection against observations embodying improbable feasibility levels.)

[52] In order to guarantee for a good comparability of the effect of the external factors on efficiency when comparing pre-war and post-war results, the models were always calculated for the full frontier, a robust frontier and an inefficiency frontier with the same values for alpha (1, 0.99 and 0.9 respectively).

Fig. 18: Densities for PP (left) and CP (right) full (upper), robust (middle) and partial (lower) efficiency estimates (pre-war period)



Taking a second look at the histograms, we can further obtain that citation efficiency is lower than publication efficiency. In the full frontier case, the spread of inefficient values in the PP model ranges from 0 (efficient) to about one with a relatively high density of above 0.2 close to the boundary. This means that the least efficient observations use the same amount of input as a peer which produces about 248 more paper publication output per decade.[53] For the CP model in turn, the spread is wider, with the highest inefficiency values approximating 2.5 (~3000 citations less than technically feasible) and the density of observations, lying close to the frontier, below 0.2. The greater volatility and spread in citational inefficiency confirm the theoretical remarks on the properties of citations, which are subject to cumulative advantages (Matthew effect) and are less loosely connected to the input of the scientific production process in comparison to publications (see section 4.2.1.2).

The latter is further affirmed by the densities of efficiency values in the lower histograms, where the partial frontier is depicted. Here we see a relatively even distribution of values beyond and below the boundary in the PP model, suggesting that the partial frontier defines a technology based on observations operating at the middle of the inefficiency distribution. Now, the least efficient observations lack about half of publications of the sample mean to reach the frontier, which suggests that the technologies estimated under full, robust and partial frontier are similar in shape and proportionately distributed over the output dimension.

Finally, when looking at the inefficiency frontier estimated in the CP model, we find a more skewed distribution of efficiency estimates. Here, when the efficiency frontier is constructed at the middle of the inefficiency distribution, we find an extreme observation which has a nearly 6 times higher citational return when compared to an average technology. The latter observation refers to the Harvard university in between 1920 and 1930, which produced 13,193 citations with a given number of 163 professors. We can also visually inspect the distance of the above addressed observation of the Caltech to the inefficiency

[53] Remember that here output-oriented efficiency is measured, by setting the distance function in the input dimension to zero and the distance function in the output dimension to the (pre-war) sample mean of about 248 publications (See table 24 for descriptive statistics of the pre-war sample)

frontier, which even though slightly deviating from the bulk of observations lies a lot closer to the frontier when compared to the Harvard observation.

Recapitulated, about one third of observations is producing publications either efficiently or close to the boundary. The number of inefficient observations is higher in the CP model, with about two times greater distances of the least productive observations to the frontier. The latter is driven by extraordinarily productive observations (e.g., Harvard university, 1920s), which are classified as superefficient in the partial frontier model, yet not in the robust frontier model. The fading densities in the inefficiency dimension portrayed in all histograms suggests that most universities considered in the pre-war sample operates close to the frontier. This is in line with the image of those institutions, which regularly occupy high positions in university rankings. The reasonable distributions and the findings validate the idea promoted in the methodology section to find reasonable boundaries for all observations with the exception of the robust frontier capturing no outliers here.

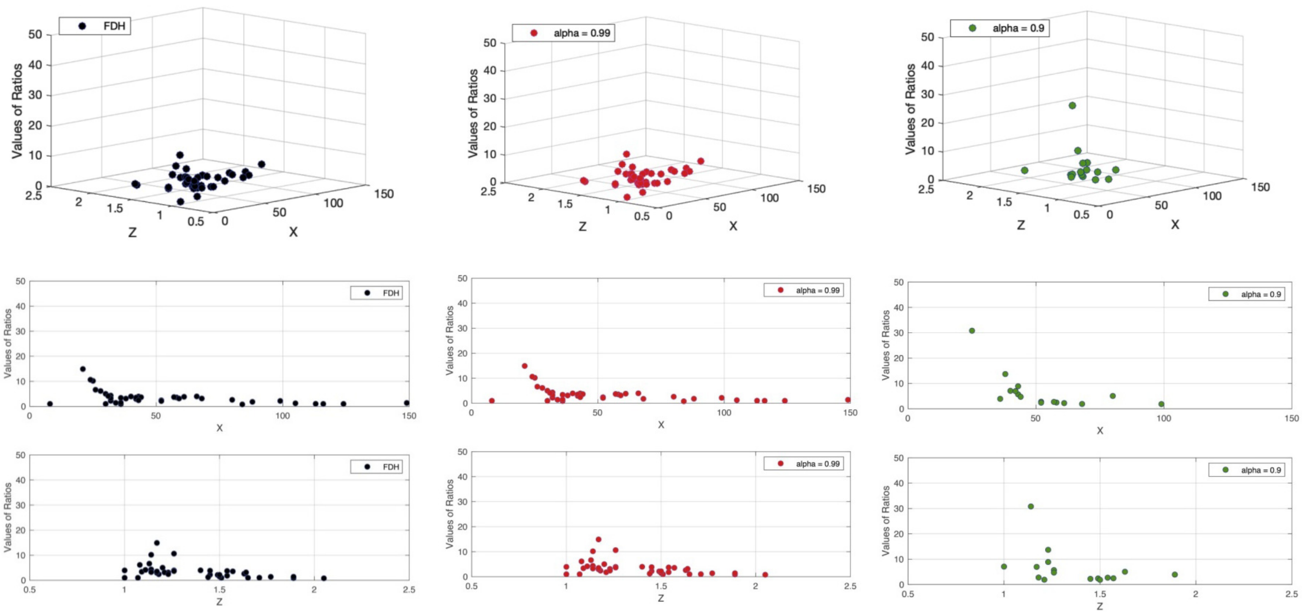
4.3.1.2 Effect of DoL and Spec. on pre-war efficiency

In figure 19, the ratios of conditional to unconditional PP efficiency are provided as a function of the inputs X, the external factor Z (here: task division) and as a joint effect of X and Z in the three-dimensional scatter plots. Again, the latter are differentiated according to the full, robust and partial frontier case. The interaction between size and external factor is comparable for all four components of DoL and Spec. Further, for the pre-war sample we find no difference in the effects of the external factors according to full, robust and partial frontier. This means that for the pre-war sample, we can reject the separability condition, meaning that a two-stage analysis, using a truncated regression model to examine the effect of external factors on efficiency would have not been methodologically sound (Daraio et al. 2018).

Since the interaction of size and external factor holds for all external factors and the effects obtained for the latter are consistent for different technologies, the interpretation can be limited here to the example of task division, for which the interaction with size is most pronounced

Fig. 19: Ratios of (task division) conditional to unconditional model for full, robust and partial PP efficiency estimates (pre-war period)

β_α (X,Y|Z) / β_α (X,Y) vs. X vs. Z (task division)



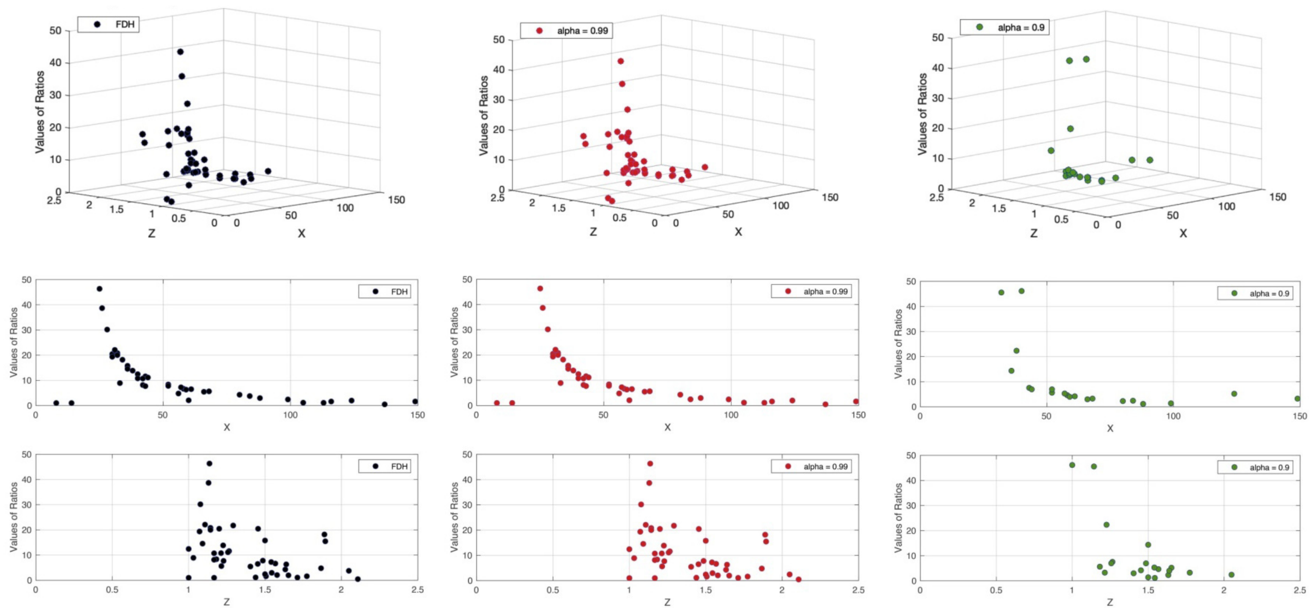
(See Appendix S11-S16 for plots of the other external factors). All three marginal plots suggest an effect of size and task division on efficiency. Since as discussed in the section above, full frontier and robust frontier are identical in the unconditional case, it is not surprising that the shapes for the effects of size and task division are similar here as well. The effect of size in case of the inefficiency distribution is slightly more pronounced with higher values of ratios concentrated in the input range of 25 to 100 (professors) yet the overall pattern is consistent for all three frontiers. Indeed, the ratios suggest a favorable effect of more professorial input on efficiency with an input of 50 (professors) being a critical threshold. Observations lying below the latter are found to be particularly inefficient.

This corresponds to the ratios plotted against the values of task division, which is also found to impact efficiency. In the PP model we find a similar pattern for the effect of task division on the efficiency ratios, with higher values (, corresponding to more professors per denomination) being favorable for efficiency, whereas moving towards a fully institutionalized task division (one professor per denomination) decreases inefficiency. It should be noted though, that approximating the value 1, this effect gets mitigated a little suggesting that committing to a fully institutionalized task division is not necessarily harmful in case of the pre-war sample and that the positive effect of team-based science requires for a certain minimum team size. Clearly though, we can see an interaction of size and task division. In the pre-war period, smaller universities profited from greater values in task division, meaning employing their scarce professorial input on fewer denominations.

Looking at the marginal plots of the CP model in figure 20, we can see identical patterns for the interaction of task division with size. Indeed, in the case of full and robust frontier, we obtain a greater vulnerability of citational productivity to size. Here, clearly the size effect reveals an even more consistent effect on efficiency when compared with the external factor (task division). This indicates that the positive effect of size on citational productivity cannot fully be reduced to the effect of task division. This is in line with the expected cumulative advantages larger institutions might have in generating citations. Here the size of the institution might also to some extent reflect its capacity to draw attention (and consequently recognition) to its publications favoring its efficiency in the CP model.

Fig. 20: Ratios of (task division) conditional to unconditional for full, robust and partial CP efficiency estimates (pre-war period)

β_{α}(X,Y|Z) / β_{α}(X,Y) vs. X vs. Z (task division)



Figures for the nonparametrically fitted regressions of the marginal effects, as well as nonparametric fits of the joint effects are provided for division of labor and specialization in the PP model case in figure 21 and 23 as well as in the CP model case in figure 22 and 24 respectively. For the marginal effect fitted for task division we can confirm the interpretation of the visual inspection of the ratios. We find a positive effect for higher values of task division in both models. Interestingly, here in the pre-war case, the supposed positive effect of coordinating tasks must be rejected. There is a monotonous linear negative effect of high task coordination on efficiency in the publication model and a negative effect of task coordination in the citational model below the sample mean value. Afterwards it seems to be stagnating and decreasing eventually for very high values of task coordination. The latter though is only backed by one observation of 1.5 for the factor variable and should thus be considered as unclear.

The joint effect depicted in the GPR based surface plot aligns with the interpretation of the marginal effects.[54] A combination of low values for task division in conjunction with a high value of task coordination seems particularly unfavorable and the other way around. Indeed, universities with higher institutionalized task division seem to benefit in particular from low values of task coordination. This indicates that in pre-war science there is a favorable effect of larger size and more professors per denomination that are effectively strictly separated. A possible explanation for this could be that high task coordination in the pre-war science period could come about when denominations are covering broad research domains (e.g., engineering, biology, medicine, chemistry) instead of measuring the coordination of specialized topics or subject areas. In any case, the results clearly suggest that differentiation is beneficial for efficiency in the pre-war period.

[54] Keep in mind that the fitted marginal effects are not just a cross section of the surface plots. The joint effect refers to a model estimated incorporating both external factors, whereas the marginal effects are based on models incorporating each external factor individually. Considering the fitted marginal and joint effects thus also serves the purpose of cross validating the patterns of the effects obtained.

Fig. 21: Plots for Marginal and Joint Effects of (the Components of) ***Division of Labor*** *on*
<u>Pre-War</u> Period ***Publication*** *Productivity Model (GPR model: MSE = 0.731, RMSE = 0.855, R-squared = 0.269)*

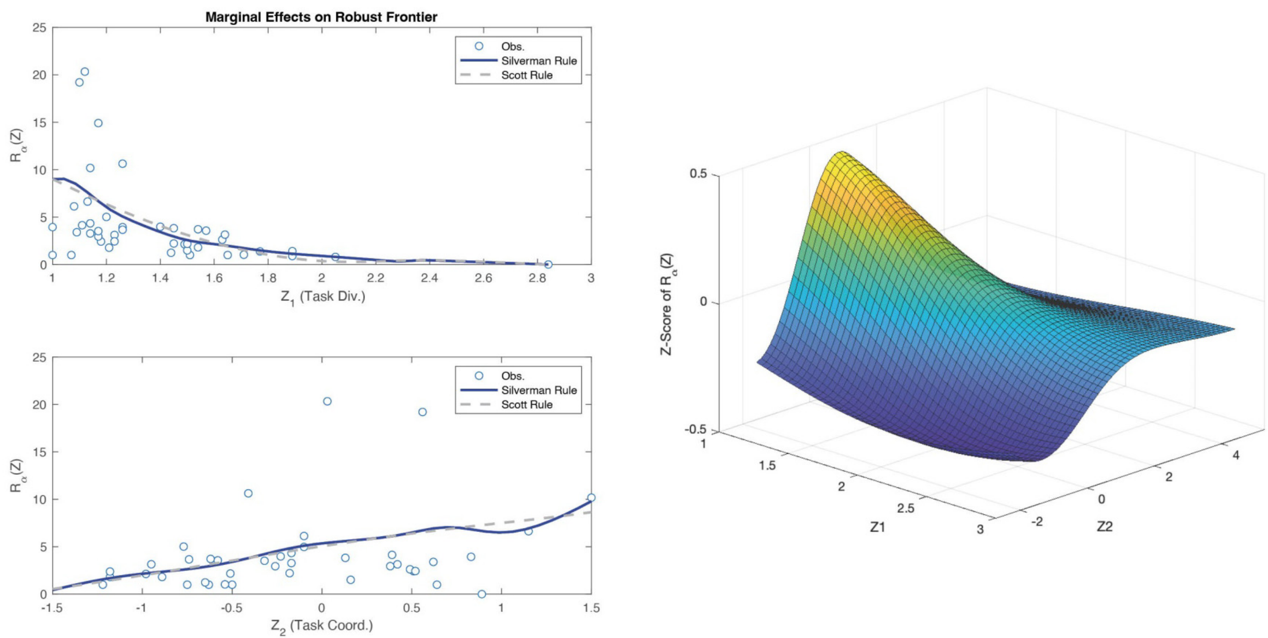


Fig. 22: Plots for Marginal and Joint Effects of (the Components of) ***Division of Labor*** *on* <u>*Pre-War*</u> *Period* ***Citation*** *Productivity Model (GPR model: MSE = 0.803, RMSE = 0.896, R-squared = 0.197)*

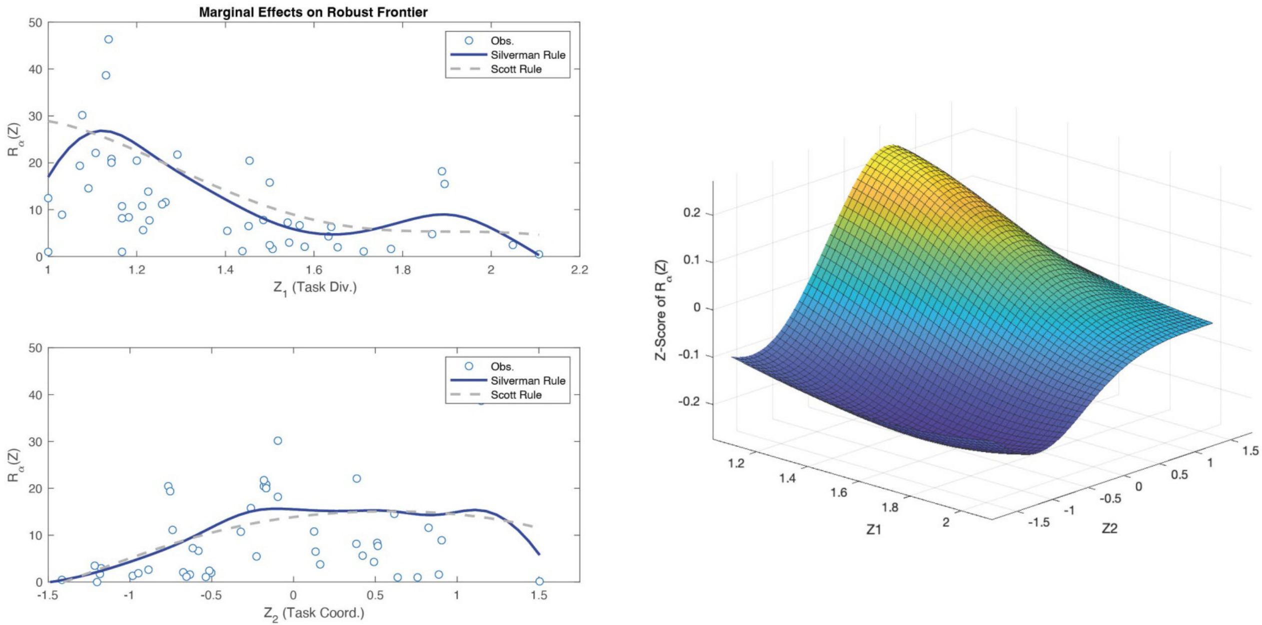


Fig. 23: Plots for Marginal and Joint Effects of (the Components of) ***Specialization*** *on* *Pre-War Period* ***Publication*** *Productivity Model (GPR model: MSE = 0.594, RMSE = 0.771, R-squared = 0.406)*

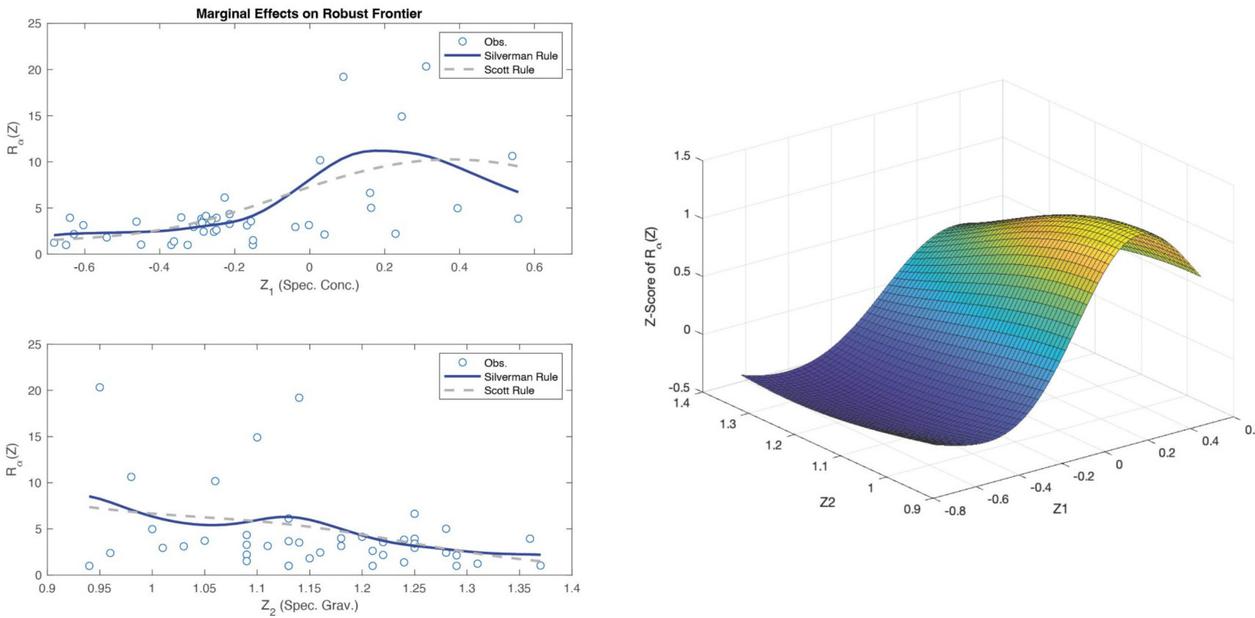
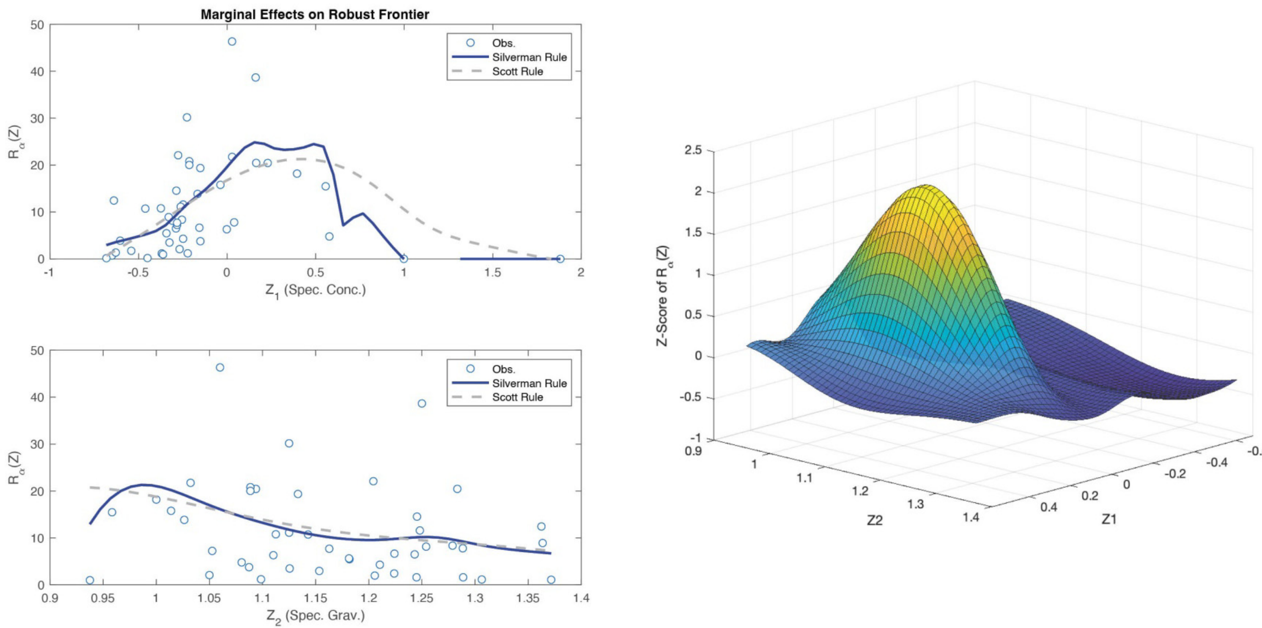


Fig. 24: Plots for Marginal and Joint Effects of (the Components of) ***Specialization*** *on*
Pre-War Period ***Citation*** *Productivity Model (GPR model: MSE = 0.354, RMSE = 0.595, R-squared = 0.646)*



Looking at the effects of specialization concentration, we find for the first-time slight differences for the effects in the PP and CP models. In the publication model, we find a slowly increasing negative impact of higher concentration, which is stagnating at higher values for the regression fitted using the Scott rule and slightly decreasing for very high values when employing the Silverman rule. Since the two regression lines fitted are deviating for high values of concentration according to plug-in rule chosen, we can conclude that in this section we are confronted with higher uncertainty in the data. The latter is equally applicable for the curve in the CP model. In the CP model we obtain a more rapidly increasing negative effect of concentration on efficiency until higher values of concentration are attained, where the two deviating curves hint at an inverted U-shaped effect. Just like in the case of task coordination, this suggests that opposed to the idea that more concentration is necessarily more productive like for example in the cases perpetuated where polytechnic universities were found to be more efficient than all-sciences universities, this cannot be confirmed for the pre-war period.

Finally, the marginal effects of specialization gravity are assessed. The pattern obtained here is congruent for both productivity models and suggests that deeper specialization in the sense of a denomination demarcating narrower scopes of research is favorable to efficiency. The joint effect of the two specialization variables reflects the uncertainty addressed for higher values of concentration, where the fitted curves where deviating. In particular in the CP model, the joint effect is not really smooth and hard to interpret. In the PP model on the other hand, the joint effect of specialization supports the idea that lower concentration and higher specialization in depth are favorable for efficiency and the other way around.

Interestingly, marginal and joint effects support the idea that in pre-war science, universities were more efficient in case their tasks were more isolated, scope of research more diverse and those isolated tasks specialized more in depth. The positive effect of less institutionalized task division and higher values of professors per denomination is not necessarily conflicting with the latter, since those numbers are still comparatively low ranging from 1 to 3 professors per denomination and partially could be due to a generally beneficial size effect. The combination of these effects could certainly be pictured by imagining a relatively big all-sciences university with established chairs across different research fields, which are comparatively more specialized than peers in other institutions.

4.3.2 Post-war frontier

4.3.2.1 Efficiency results

In table 30, the efficiency estimates are provided for the post-war frontier. (See Appendix S9 for a full table with efficiency estimates of all 111 observations) Regarding the (sample mean) institutional size it can be documented that while in the pre-war sample European institutions ranked high to average, they are now clearly dominated by US institutions, in particular by institutions belonging to state university systems (e.g., University of Washington, University of California) The LMU Munich for example moved from the institution with the third biggest mean input in the pre-war sample to the eights rank with half of professorial staff when compared to the biggest institution of the post-war period, the University of Washington at Seattle.

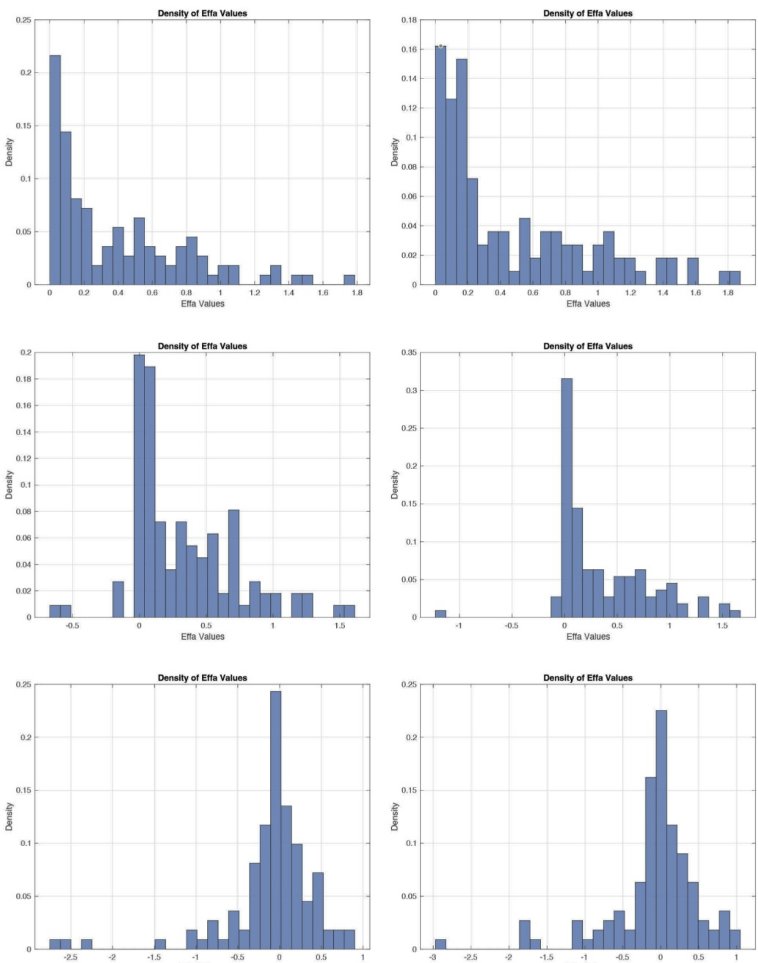
In general, the mean performance in production and citation efficiency are highly correlated. Indeed, only the University of California at San Diego and the Harvard University perform better in citational efficiency than publication efficiency in both full and robust frontier case. In the partial frontier case, they also reveal the highest values for superefficiency together with the Stanford university. Furthermore, in the robust frontier case, of which the efficiency estimates are taken for the nonparametric fit of marginal and joint effects, the Harvard university is the only institution that is classified as an outlier on average (not just for a single observation). Since we are looking here at average scores according to institutions over a period of seven decades, the estimates do not just reflect efficiency in a certain period but rather reflect continuity in performance (in a sample that is already limited to the most productive universities in the world). The results for citational efficiency thus suggest an overarching importance of the Harvard university for post-war science.

Tab. 30: Post-war period institutional mean of inputs, outputs and efficiency estimates

i	X	Y	Y2	FDH	FDHc	Oa99	Oa99c	Oa90p	Oa90c	APP
UZH	127	10855	525799	0.03	0.05	0.00	0.02	-0.20	-0.20	6
ETH	128	13732	833525	0.05	0.07	0.02	0.03	-0.24	-0.26	3
LEE	135	8333	361356	0.07	0.07	0.05	0.06	-0.13	-0.13	4
UPP	153	13831	701269	0.13	0.14	0.09	0.10	-0.22	-0.21	4
CAL	172	17289	1004088	0.06	0.07	0.02	0.02	-0.24	-0.28	7
AUC	198	9261	335174	0.17	0.22	0.13	0.18	-0.05	-0.02	7
GAU	280	9142	317546	0.36	0.38	0.33	0.34	0.07	0.10	6
UCS	288	24493	1729816	0.21	0.16	0.16	0.10	-0.21	-0.34	5
UOS	302	13278	536502	0.37	0.39	0.33	0.34	0.03	0.07	6
HAR	320	29120	2225218	0.27	0.22	0.20	-0.05	-0.28	-0.66	5
RFW	414	9839	315799	0.61	0.66	0.56	0.61	0.28	0.34	6
STA	423	38148	2333598	0.32	0.45	0.19	0.27	-0.49	-0.53	7
LMU	447	16686	619755	0.62	0.70	0.56	0.64	0.19	0.31	6
MIT	454	30564	1836493	0.54	0.53	0.40	0.46	-0.11	-0.13	7
UTA	469	22471	963359	0.57	0.66	0.48	0.58	0.05	0.27	5
COL	474	5703	170046	0.81	0.83	0.78	0.79	0.45	0.47	2
UCD	514	18202	934578	0.70	0.73	0.61	0.66	0.16	0.25	5
UCB	601	30479	1809983	0.74	0.89	0.67	0.78	0.02	0.02	7
UCL	611	44181	2301450	0.51	0.80	0.37	0.48	-0.51	-0.28	6
UOW	891	37904	2105974	0.68	1.04	0.58	0.66	-0.24	-0.11	7

When looking at the distributions of the efficiency estimates according to full, robust and partial frontier case given in figure 25, the high correlation of publication and citation efficiency is confirmed, and we can obtain that the difference obtained in patterns of publication and citation efficiency in the pre-war results disappeared. The latter suggests that if we consider the feasibility of attaining certain levels of publications and citations by defining suitable boundaries (technologies) the existing differences in numbers of citations per publication are marginalized (The results of publication and citation efficiency reveal a similar distribution of efficiency estimates). Since the latter is regularly considered as a performance indicator in scientometrics, this is an interesting finding and further highlights the importance of moving from an exclusive consideration of scientific outputs ('bibliometric hypothesis') to accounting for institutional inputs.

Fig. 25: Densities for PP (left) and CP (right) full (upper), robust (middle) and partial (lower) efficiency estimates (post-war period)



The patterns of publication and citation model lie so close to one another that permitting a small number of observations to lie beyond the boundary (, moving from full to robust frontier) causes the marginally lower citational and marginally higher publication efficiency obtainable in the full frontier case to flip over in the robust frontier case. We can further see that here in the post-war period the efficiency estimates are calculated using three different technologies, because as expected, observations are classified as superefficient in the robust frontier case. This concerns e.g., recent observations of the Stanford university and the MIT in 2010, but also observations of different periods like the Harvard University in 1990 or the University of Leeds in 1960. This shows that the methodology can get a grasp of conceptual outliers and (, perhaps surprisingly) the latter are neither found to be strictly time or size dependent validating the idea of measuring a singular joint frontier for the post-war period.

The histograms in the bottom panels are confirmatory of the choice of alpha, portraying nearly exactly the middle of the inefficiency distribution. Further, here the conceptual outliers detected in the robust frontier case indeed reveal a significant distance from their peers signaled by the gaps in between the densities depicted in the negative dimension. On the contrary, the observations classified as inefficient using the robust frontier as benchmark, now shifted beyond the boundary (due to the low alpha value) are grouping relatively dense and close to the frontier, which validates that those observations have rightfully been considered as regular inefficient observations of the robust frontier case.

Notably, while the inefficiency spread in publication productivity increased from 1 to 1.5 for the least efficient observation, average citation productivity moved closer together (decline from 2 to 1.5 of least efficient observation). Of course, for the pre-war frontier, where a lot of recently founded universities were still at the start of institutional growth, it is reasonable to assume that productivity in acquiring recognition could vary. On the contrary though, the pre-war period only covers a time period of 40 years, whereas the post-war period observation spans over 70 years. That indeed, given the latter, the spread in citational productivity is nonetheless greater in the pre-war period, is an indicator that feasibility depends more on an institution’s size and how long it is established than on the point in time it operates at.

4.3.2.2 Effect of DoL and Spec. on post-war efficiency

Just like in the previous section, the analysis of the ratios according to full, robust and partial frontier will be examined for the sole example of task division. Here, analogous to the pre-war results, the interaction of size with the external factors is most pronounced for task division and the pattern of the size effect obtained here is similar for all four components of DoL and Spec. (See Appendix S17-S22 for plots of the other external factors).

When comparing the interaction of task division and size portrayed in figure 26 for PP and figure 27 for CP model with the one in the pre-war period, we notice that the interaction of size and efficiency is less clear. Evidently, observations operating at an input size of 150 to 300 could profit from increased size. Outside of this range, the interaction seems to be rather random.

For the robust frontier in the CP model, we see that the inefficiency estimates around 200 are shifted upwards. Supposedly, a lot of observations exist that lie in this input range and due to the greater number of peers they are more vulnerable towards the inclusion of task division in the conditional model. The latter could point at the limits of the methodology employed when there is an accumulation of observations around a certain input level, where there might be a small portion of observations benchmarked against a technology that is not suitable for them. On the contrary, the observations concentrated at a particular size level could have substantially different levels of e.g., task division, explaining that the effect of the external variable is more pronounced here. The potential interactions of size with task division and size with the remaining external factors will be addressed when limitations of the work are discussed in chapter 5.

When looking at the marginal effects of task division, we find a more consistent pattern. Increased values for task division, in particular institutions with 10 or more professors working on the same denomination are unlikely to be highly production or citation efficient respectively. The decreasing trend in inefficiency for higher task division in teams is consistent for both PP and CP model as well as the three different technologies that serve as benchmark for the observations.

Fig. 26: Ratios of (task division) conditional to unconditional for full, robust and partial PP efficiency estimates (post-war period)

$\beta_\alpha\,(X,Y|Z)\,/\,\beta_\alpha\,(X,Y)$ vs. X vs. Z (task division)

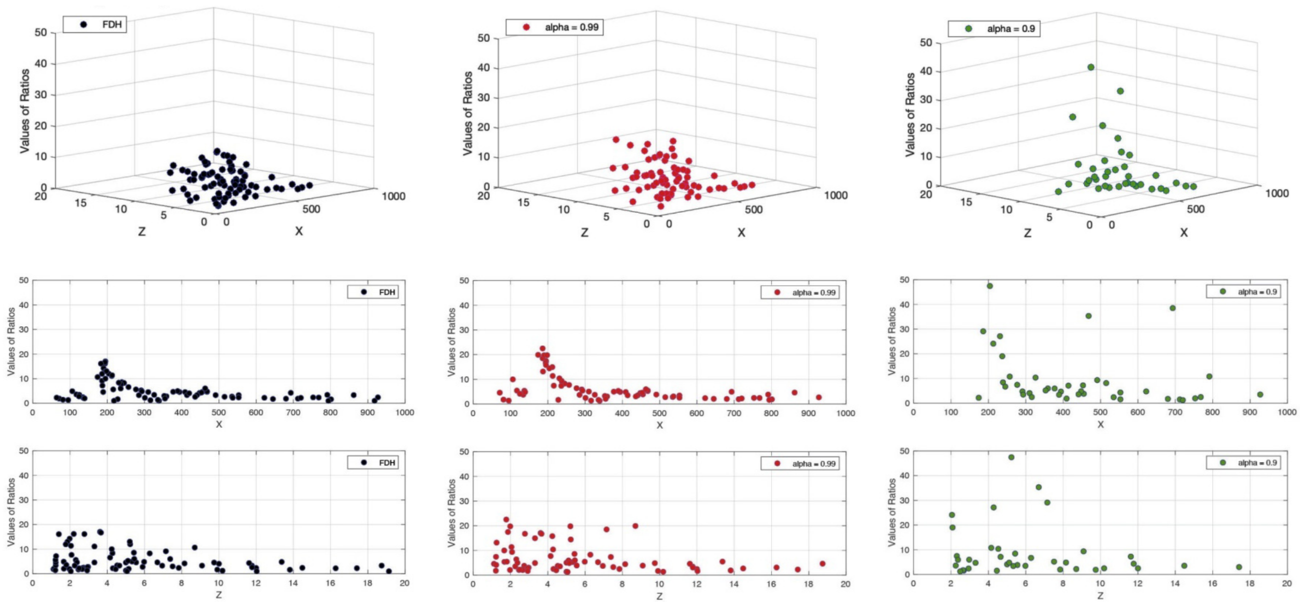
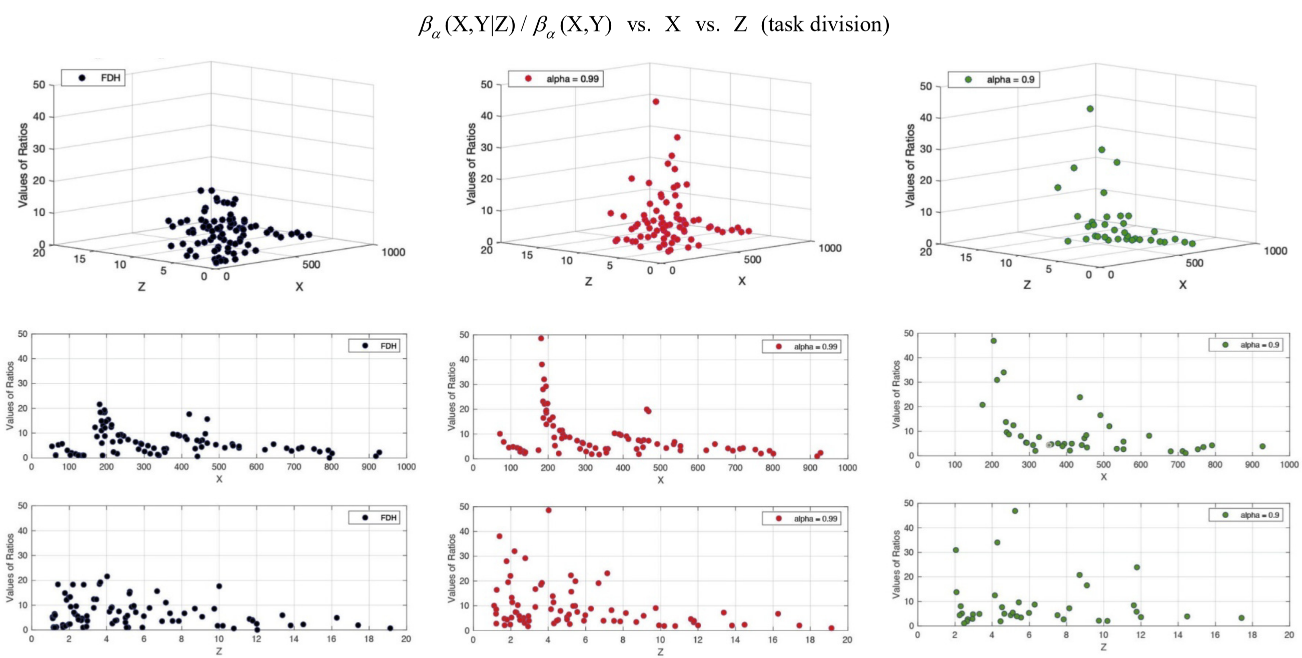


Fig. 27: Ratios of (task division) conditional to unconditional for full, robust and partial CP efficiency estimates (post-war period)



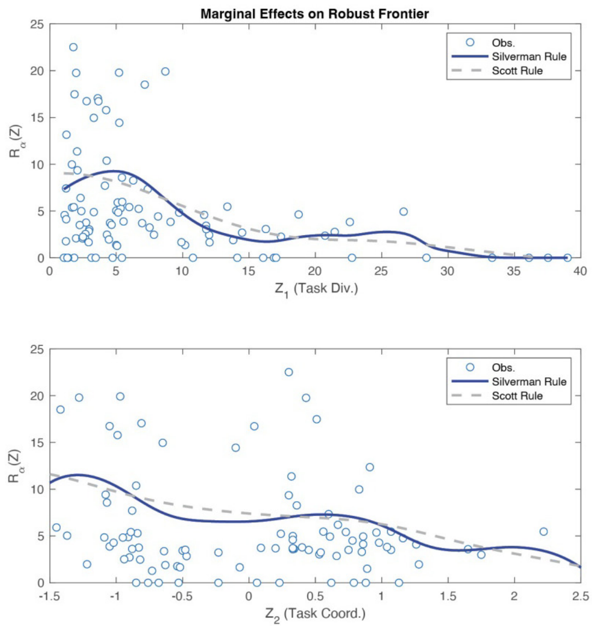
When turning to the fitted effects provided for division of labor in the PP and CP model case in figure 28 and 29 respectively, as well as for specialization in the PP and CP model case in figure 30 and 31 respectively, we can rediscover that in the regression fitted employing the Silverman rule of thumb, full task differentiation (, institutionalized task division) is less inefficient than low to medium task division, where inefficiency peaks around a value of five professors per denomination. Alternatively, when employing the Scott rule the fitted effect suggests a monotonously increasing efficiency with greater values for task division.

As opposed to the results of the pre-war sample, we find a monotonous positive effect of higher task coordination on efficiency in the publication model. In the citational model, the fitted regression indicates no effect of task coordination for low and medium values yet a positive effect for very high task coordination on efficiency. When looking at the joint effects we can easily obtain that this positive effect of task coordination only applies to institutions with highly differentiated tasks and low values for task division. The latter is perfectly in line with the theoretical expectation that highly institutionalized task division creates higher coordination costs. Institutions addressing these coordination costs by investing in higher task coordination are found to be more efficient than their differentiating yet not coordinating peers. Even though the marginal effect of task division and task coordination are both a lot less pronounced in the CP model the joint effect on efficiency is identical to the one of the publication model, which suggests that the joint components of division of labor influence both epistemic outcomes considered in the post-war period.

The joint effect for division of labor further suggests that the positive effect of task division (in teams) fades for very high values of (20 or more) professors per denomination. The latter could be explained by the same mechanism applicable to highly institutionalized task division, only on the level of the department or team. Here, coordination costs might also increase with increasing team size that requires for enhanced effort of coordinating subtasks. (See section 2.1.2 for literature dealing with division of labor in scientific teams.)

When assessing the marginal effects of (specialization) concentration in both models, for the first time we obtain two different patterns. Whereas in the PP model, the fitted regression suggests an inverted

Fig. 28: Plots for Marginal and Joint Effects of (the Components of) ***Division of Labor*** *on* <u>*Post-War*</u> *Period* ***Publication*** *Productivity Model (GPR model: MSE = 0.809, RMSE = 0.899, R-squared = 0.191)*



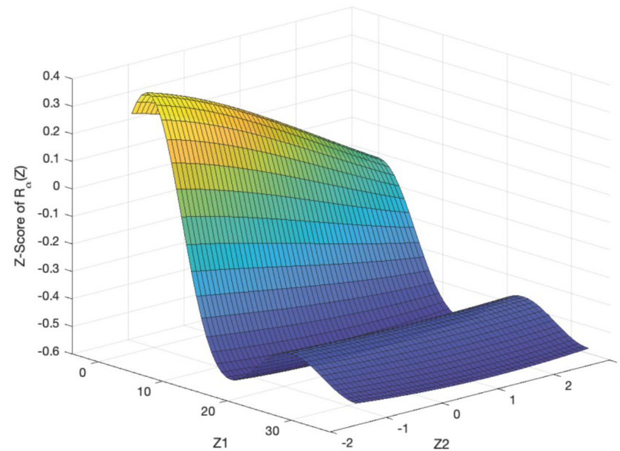


Fig. 29: Plots for Marginal and Joint Effects of (the Components of) ***Division of Labor*** *on* *<u>Post-War</u> Period* ***Citation*** *Productivity Model (GPR model: MSE = 0.801, RMSE = 0.895, R-squared = 0.199)*

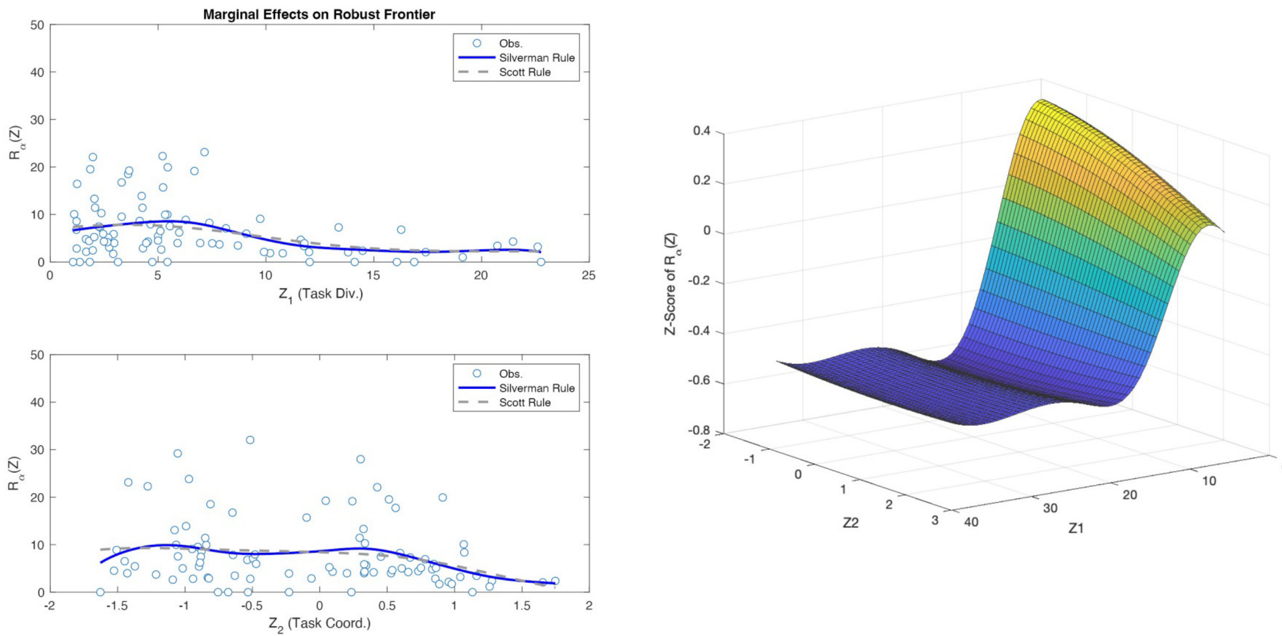


Fig. 30: Plots for Marginal and Joint Effects of (the Components of) ***Specialization*** *on*
<u>*Post-War*</u> *Period* ***Publication*** *Productivity Model (GPR model: MSE = 0.956, RMSE = 0.978, R-squared = 0.04)*

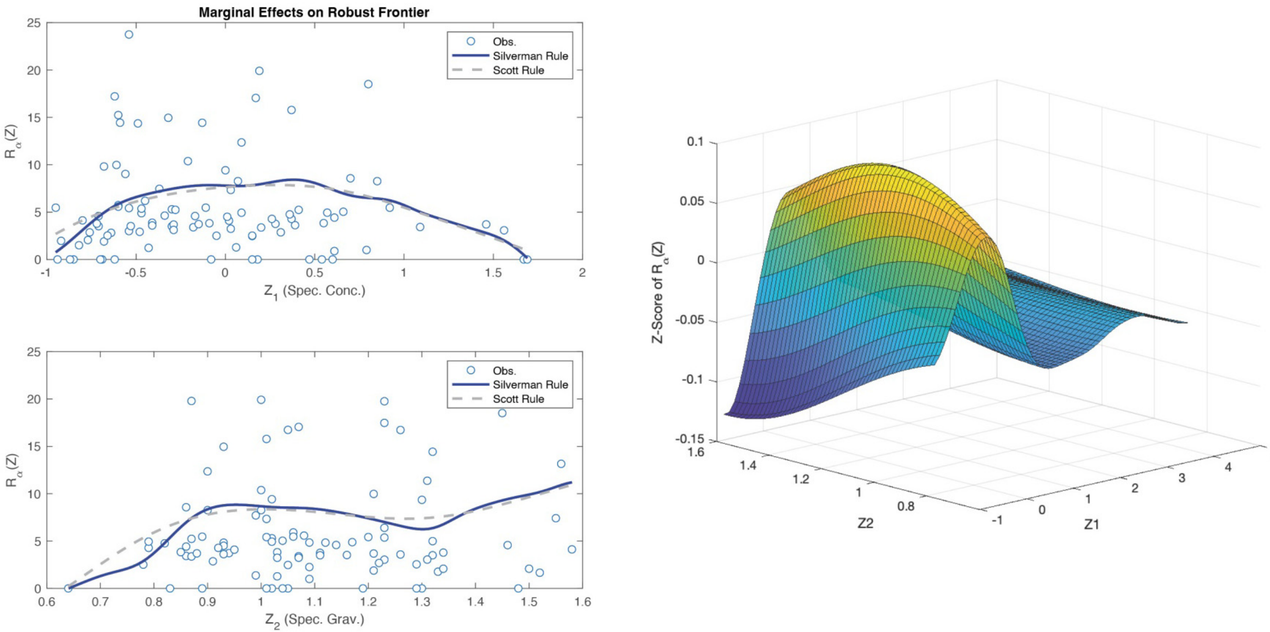
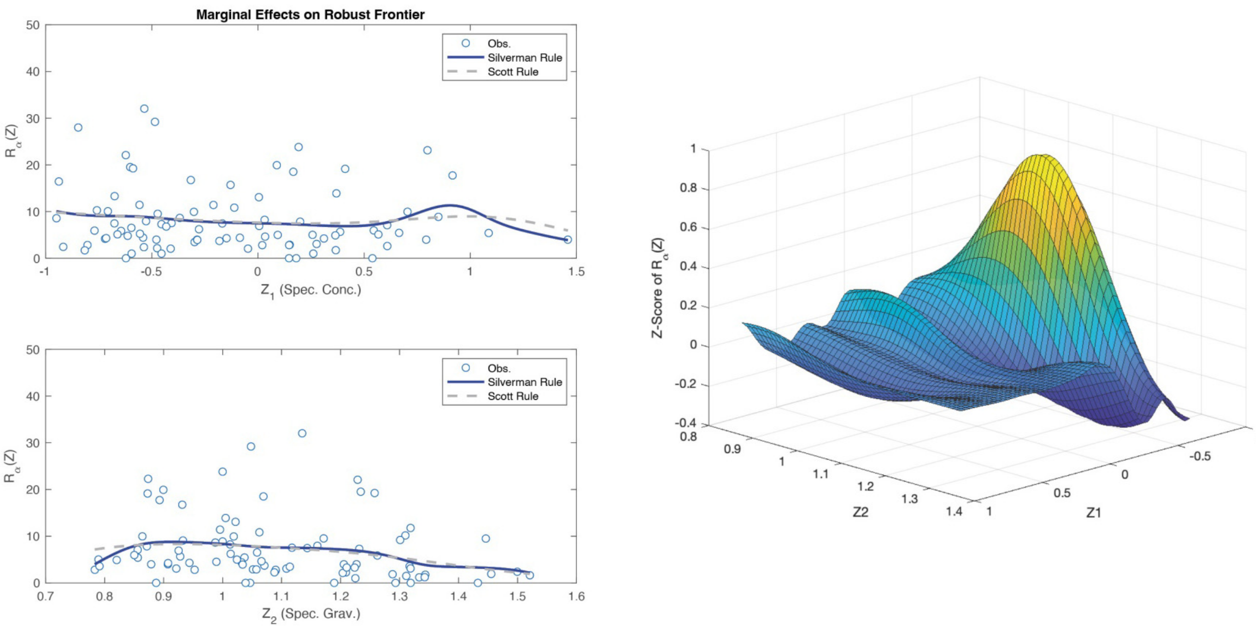


Fig. 31: Plots for Marginal and Joint Effects of (the Components of) ***Specialization*** *on* *<u>Post-War</u> Period* ***Citation*** *Productivity Model (GPR model: MSE = 0.741, RMSE = 0.861, R-squared = 0.259)*



U-shaped effect of concentration on efficiency, in the CP model the ratios are distributed rather randomly implicating no effect on efficiency. This is in contrast to the prevailing idea in scientometrics that citations must be field-normalized. Granted, here only citations of the natural sciences and engineering disciplines are considered, which could on the aggregate level of the institution and the course of a decade indeed be less vulnerable to potential distortions according to field. Nonetheless, a really profound effect of concentration on citational productivity cannot be stated for the post-war period.

Interestingly though, we find an effect of concentration on publication efficiency, which one could have supposed for recognition instead. Here we can confirm the before anticipated threshold for concentration to be favorable for efficiency. While medium levels of concentration, which could be interpreted as institutions without a strategical orientation towards a certain domain, are detrimental for publication efficiency, very low and very high concentration of professorial staff on denominations, subject areas and research fields is found to have a positive effect on publication efficiency.[55]

Finally, specialization gravity also seems to have a different effect on PP and CP model. Common denominator of the effect in the two models is that it has a rather moderate effect in the range of 0.9 to 1.3. Focusing exclusively on this range, we would probably conclude that the variable has no effect on efficiency. Considering the tail ends, which are consistent for both alternate bandwidths specifications, we might get the impression of a negative effect of specialization depths on efficiency in the publication model. In any case, the positive effect of specializing in depth found in the pre-war model, must be rejected for the post-war sample.

Considering the joint effects of specialization, we can conclude that its effect on post-war citation efficiency might indeed be complicated to entangle. Here the fitted surface plot suggests a negative impact on efficiency for medium specialization gravity in conjunction with medium low concentration. For the PP variant, the joint effect reveals a similar pattern, where medium values of concentration in conjunction

[55] Please keep in mind when interpreting the inverted U-shaped effect here that values approximating zero signal efficiency, whereas higher values indicate inefficiency.

with medium values of gravity produce the most inefficient ratios. In conjunction with the interpretation of the marginal plots, it can be concluded that combining average levels of concentration and gravity enhances inefficiency. On the other hand, following any kind of strategy by either strongly or not concentrating at all and strongly or not specializing in depth at least seems less detrimental. Potentially, the effect of specialization can further be entangled when considering the effects for the different university types analyzed in the upcoming section.

4.3.3 Full frontier according to university clusters

4.3.3.1 Efficiency results

Finally, the results of the mean efficiency estimates (for the full sample frontier) according to clusters are provided in table 31. (See Appendix S10 for a full table with efficiency estimates of all 166 observations) For the whole period we find that in the FDH case (the European) cluster 1 is both slightly more efficient than the second cluster in both publication and citation productivity. Interestingly, for the whole period of 1890 to 2020 the larger institutions bundled in clusters 3 and 4 are found to be on average substantially less efficient. This affirms the above proposed idea that the positive effect of size obtained in the pre-war period rather reflects the positive effect of being an established institution than a big one. Additionally, it supports the finding for the post-war frontier that the positive effect of increased task division saturates for very high values.

Tab. 31: Mean of inputs, outputs and efficiency estimates according to cluster

i	X	Y	Y2	FDH	FDHc	Oa99	Oa99c	Oa90p	Oa90c	APP
Cluster 1	174	7029	295759	0.27	0.29	0.23	0.25	-0.06	-0.02	59
Cluster 2	221	12377	710779	0.32	0.32	0.25	0.23	-0.18	-0.27	52
Cluster 34	436	24344	1376473	0.60	0.75	0.49	0.56	-0.42	-0.44	55

Certainly, the most interesting finding here is the difference of the results in the FDH case and for the robust frontier. Permitting a small amount of very productive observations to lie beyond the frontier, the institutions in the second cluster become equally efficient on average as their European counterparts. This suggests that except for the larger institutions part of state university systems, when considered over the whole period of modern science excellent universities are on average equally (in-)efficient.

When looking at the densities of the efficiency estimates portrayed in figure 32 for the PP model and figure 33 for the CP model, it becomes clear why the European cluster performs very well when a joint frontier for the whole period is measured. While the spread in FDH in-

Fig. 32: Densities of cluster 1 (left), cluster 2 (middle) and cluster 3 (right) for full (upper), robust (middle) and partial (lower) PP model efficiency estimates

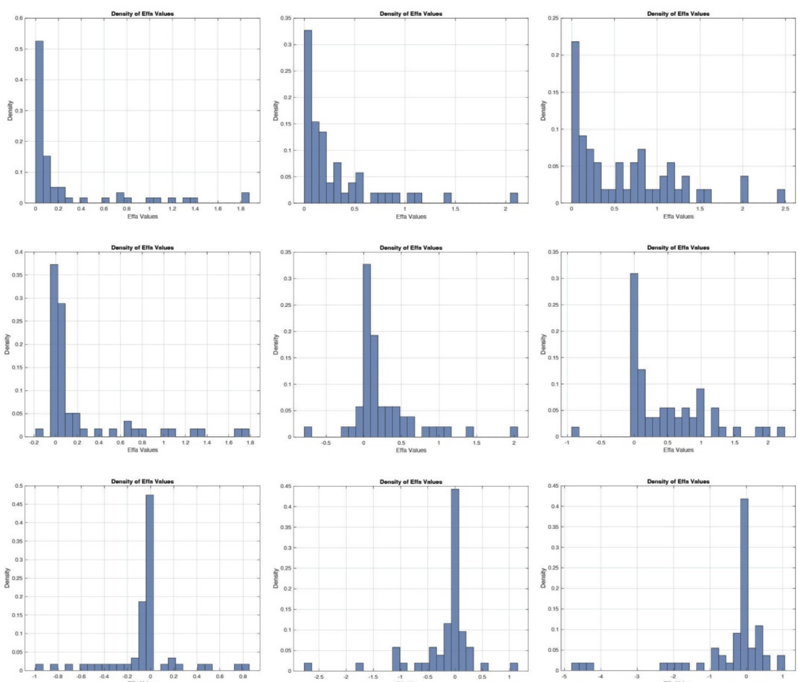
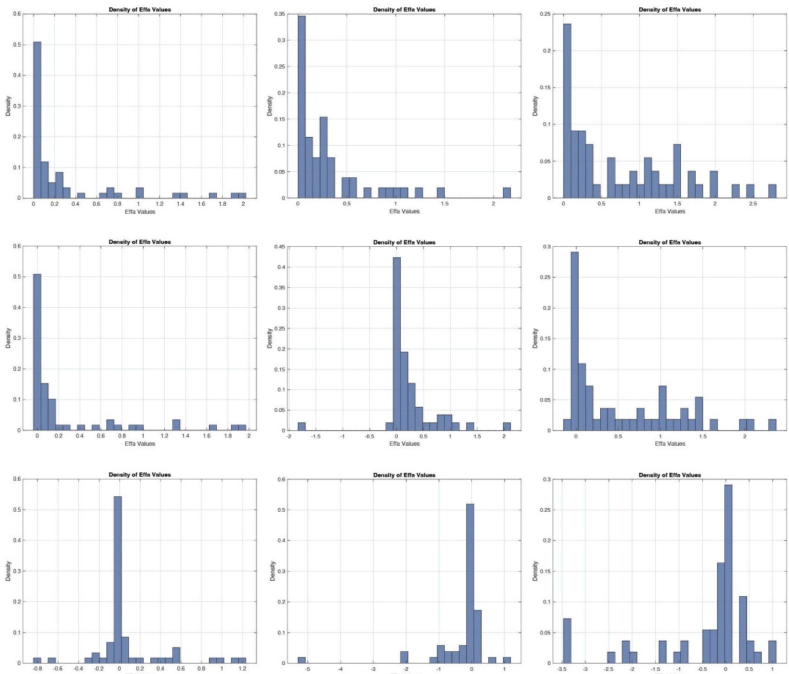


Fig. 33: Densities of cluster 1 (left), cluster 2 (middle) and cluster 3 (right) for full (upper), robust (middle) and partial (lower) CP model efficiency estimates



inefficiency is comparable to the one of the second cluster, the density of efficient observations is particularly high. Apparently, when a joint frontier for the whole period is estimated, a lot of the smaller European institutions are operating under their own technology, shifting their mean efficiency upwards. This indicates two things. For one, the idea that the methodology can be used to guarantee for universities being benchmarked against suitable technologies works out quite fine. The latter though means that the comparability of the mean values of clusters is very limited since when disproportionately more institutions in one cluster serve as dominating units than in other clusters, the cluster mean efficiency needs to be interpreted very carefully.

The idea that the methodology can be used to define different technologies not only for different points in time, but also different university types is further affirmed by the histograms at the bottom of the panel. When separating the densities according to clusters, we find that the partial frontiers reflect a realistically distributed and well-centered inefficiency distribution for each cluster showing that different technologies serve as benchmark according to cluster.

Finally, the histograms point at some similarities with the results of the pre- and post-war frontier. We obtain higher values for superefficiency in cluster 2 (, where Harvard university and Caltech are located) and a wider spread in inefficiency for the citational models when compared with their publication model counterpart.

4.3.3.2 Effect of DoL and Spec. on university types

The fitted marginal effects according to clusters are provided in figure 34 for the PP model and in figure 35 for the CP model. Naturally, the effects for the full sample, not differentiated according to pre- and post-war sample reflect the combined patterns obtained in the previous section resulting in more tailed effects describing the trade-offs of the DoL and Spec. components. Further the fitted regressions here differ in scale of the external factors, making the differences in between the clusters' division of labor and specialization visible.

Overall, a lot of the effects are too random to derive functional dependencies that allow to describe the general relationship of the external factors and efficiency of the whole period. A few of the effects described in the previous sections can be rediscovered e.g. that full institutionalized task division (differentiation) is more favorable to efficiency than medium low values of task division. Unfortunately, given the low number of observations according to cluster an interaction with task coordination separated for pre- and post-war frontier cannot be provided here.

It might thus be reasonable to draw the attention of the reader's eye to the effect of specialization gravity, which was found to be favorable for pre-war efficiency yet unclear for post-war efficiency. In this cluster specific analysis, the fitted effect for the latter variable returns the most pronounced effect, consistent for both PP and CP model. For the

Fig. 34: Marginal Effects of DoL and Spec. on PP efficiency estimates for cluster 1 (left), cluster 2 (middle) and cluster 3 (right)

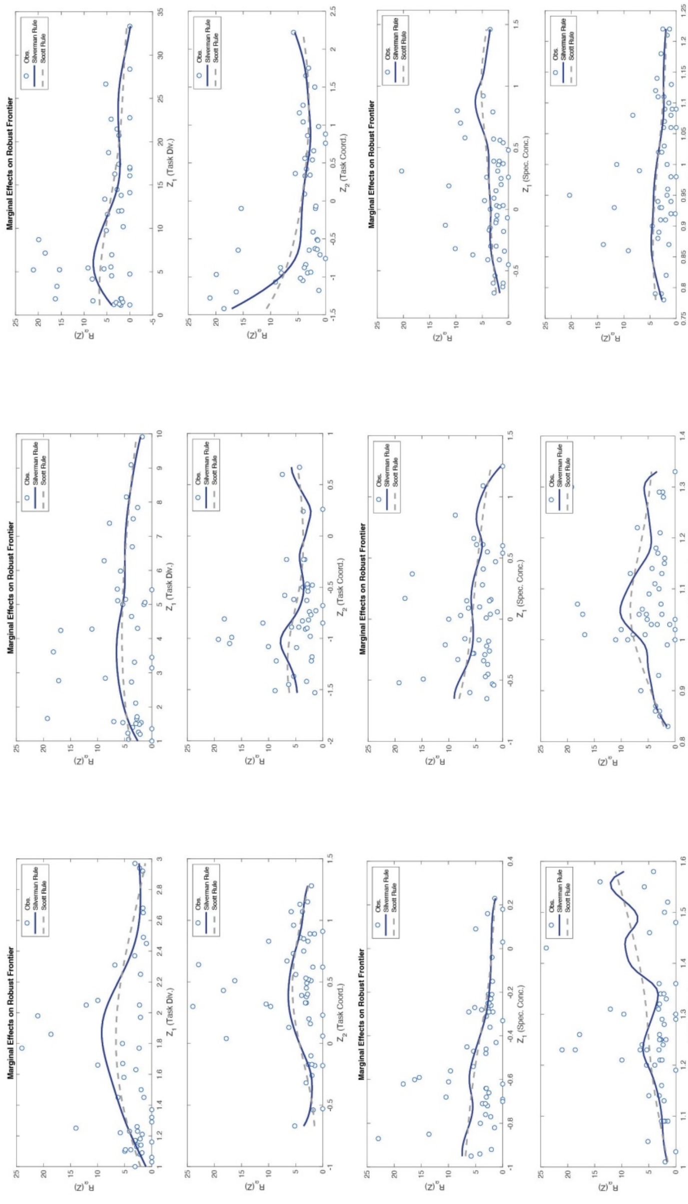
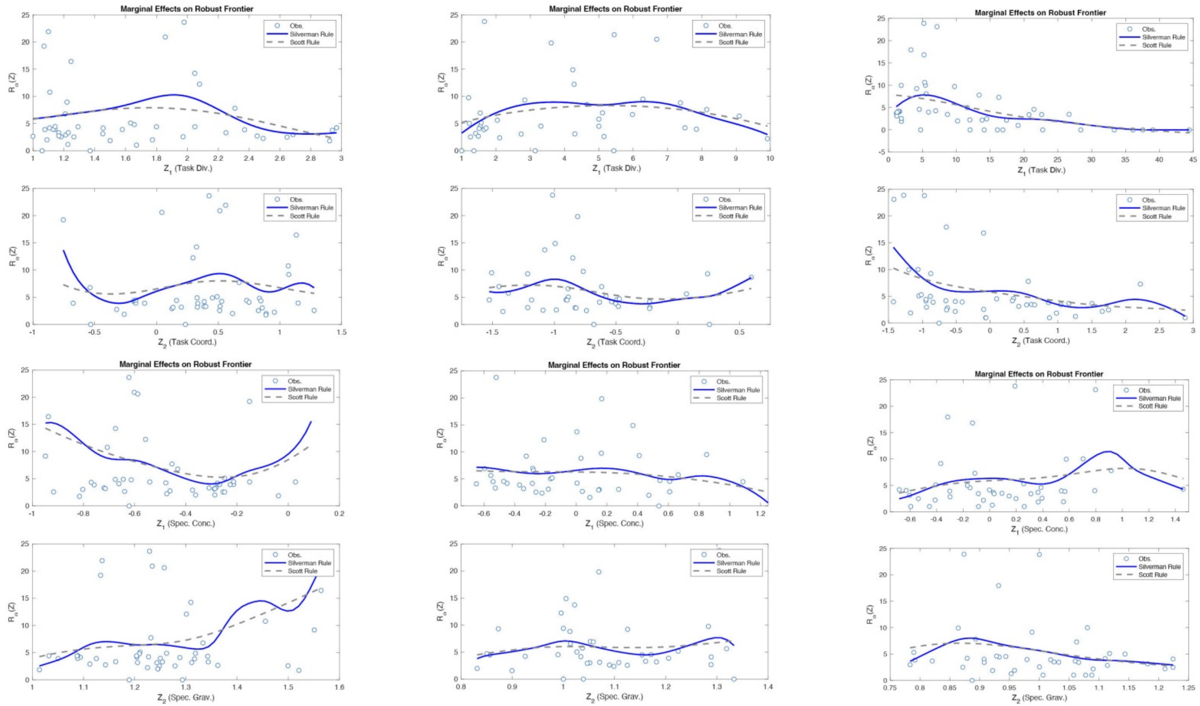


Fig. 35: Marginal Effects of DoL and Spec. on CP efficiency estimates for cluster 1 (left), cluster 2 (middle) and cluster 3 (right)



(European) cluster 1, the fitted curve suggests a negative effect of higher specialization in depth when compared to both other clusters, which reveal rather stagnating trends over the two different models. This negative effect of specialization model for cluster 1 holds for both PP and CP model. In addition, we can see at the scales of the specialization gravity variable according to cluster that, whereas cluster 2 operates at a specialization depth ranging from about 0.8 to 1.4 and clusters 3 and 4 in a range of 0.75 to 1.25, the variable ranges from 1 to a maximum of 1.6 for cluster 1. In conjunction with the positive effect of specialization gravity in the pre-war period, we may conclude that the institutions in cluster 1 are overspecialized in depth and might seriously benefit from lower values of specialization gravity. Since we know that cluster 1 is exclusively constituted by European universities this suggests that the professorial chair system sets an incentive for ongoing specialization in depth, which was a winning formula in the pre-war period, but led to overspecialization in the post-war period negatively influencing both its publication and citation efficiency.

5. Discussion of Results

The key results of the conditional efficiency framework presented in the previous chapter can be summarized as follows:

1. Efficiency in the pre-war period is found to be more volatile for citations than publications, whereas in the post-war period it is the other way round. Overall, the high share of dominating units and the dense distribution of inefficiency values confirm that the excellent universities considered here operate close to the efficient boundary and their productivities are generally quite comparable.
2. In pre-war science, we find a positive relationship of efficiency with increased task division (less differentiation) yet a negative one with task coordination. In post-war science, we find positive effects of both higher task division and coordination.
3. The marginal effects for specialization concentration follow an inverted U-shaped form except for the post-war CP model, where no clear effect is obtained. For specialization gravity we find a positive effect on efficiency in the pre-war period yet an unclear (or considering the tail ends supposedly even negative) one in the post-war period.
4. The joint effect for division of labor suggests for both CP and PP model for both pre-war and post-war frontier that differentiation (low task division value) profits from more coordination. The positive effect for task coordination vanishes for high values of task division (less differentiation, more team-based). The joint effect for specialization varies according to model, as well as pre- and post-war period. Robust for all models except pre-war publication efficiency is that configurations of average concentration and average depth are the least efficient.
5. The graphical inspection of ratios suggests an interaction of DoL and Spec. with size, which exists for all models and is most pronounced for task-division. Institutions with an input-size of 25 to

100 in the pre-war and 150 to 300 professors in the post-war period are most affected by the inclusion of DoL and Spec. in the conditional models.

6. The analysis according to university types reveals that on average efficiency of longer established institutions (cluster 1 and 2) over the whole period 1890 to 2020 is nearly identical. Recently established larger universities (cluster 3 and 4) are found to be more inefficient. Regarding the influence of DoL and Spec. on efficiency according to clusters, the clearest effect is found for specialization gravity in (European) cluster 1. Here, a negative relationship of specialization depth and efficiency is obtained.

Those findings confirm some of the theoretical expectations derived from the theory of division of labor. The task division variable employed here accounts for both differentiation on the institutional level and task division within an institutionalized unit, a denomination demarcates. The positive effect of the latter in all models coincides with the positive effects of task division in scientific teams obtained in the literature. It could be further confirmed that in later periods, for which we expect an increased 'knowledge burden' and higher coordination costs, task coordination has a positive effect on productivity. In addition, it can be shown that institutions with highly differentiated tasks profited from more task coordination in general. The negative effect for task coordination in the pre-war period might above all be attributable to the here implemented measure of task coordination, which relies on a classification of denominations according to the Web of Science subject categories. The latter led to very broadly defined denominations (e.g., 'natural sciences' or 'sciences') being defined as connecting different subject areas, research fields and disciplines. This resulted in high values for task coordination in the first sample decades, which are quite comparable to those of the last sample period's denominations (see Fig. 5: 71). Evidently though, those universal denominations can hardly be considered as coordinative. Much rather they are relicts of a pre-specialized science, where any form of delineation or demarcation of specialties led to higher productivity regardless of coordination efforts.

The results for specialization concentration suggest an inverted U-shaped relationship with publication efficiency. Effectively, this indicates that institutions which either concentrate on all granularity levels or do not concentrate at all have the highest publication productivity ratios. This effect is similar yet less robust for pre-war citation efficiency. Admittedly, from a theoretical point of view we would have rather expected this effect to be the other way around. Assuming that citational returns vary according to field (as the field-normalization literature suggests), not specializing in one domain at all could mitigate the negative effect linked to the citation customs in a particular field, whereas high concentration could increase citational return due to an outstanding role of the institution in this domain. That this relationship exists for the PP model yet is unclear for the CP model might point at the limits of deriving a functional dependency of concentration on research domains based on observations of 20 institutions.

The mechanism attributed to specialization gravity is best represented by the results for the European institutions. Indeed, it is striking that their degree of specialization depth is substantially higher when compared to sample peers. This may be explained by European universities being the oldest, and still operating with the professorial chair system, as opposed to the academic department system. Parsons and Platt (1990) claimed that the latter was the biggest innovation of US universities, which enabled more agile forms of coordination of research domains. As predicted, the task division value for European universities remained relatively stable over the course of time, which means that they invested their growth in size above all in differentiation, supposedly driven by the gravitational force of their high specialization in depth. Their non-European peers grew equally or even more in size and further reveal substantial concentration levels, yet their specialization depth remained equal or occasionally even decreased, supposedly favoring the less differentiated more team-based task division. Future research should thus validate whether the two systems indeed provide different prerequisites for handling coordination costs by controlling or not controlling the mechanism of concentration leading to narrower specialization, which in turn sets an incentive for more differentiation.

Policy implications need to build on the finding that the characterization of the scientific community as efficient self-governed sphere needs to be amended by acknowledging institutional prerequisites as additional determinants of epistemic outcomes. That the organically grown European (Humboldtian) university is found to be affected most by the self-reinforcing mechanism of DoL and Spec. fits the picture and should provoke an open discussion on how the configurations of DoL and Spec. can be optimized in the future given the evidence for pathologies due to overspecialization. European HEI policies could for example set incentives that professorial resources are invested less in differentiation and more in coordination of areas of research in the future. This particularly concerns the more granular levels of subject areas and research fields, which cannot be sufficiently accounted for by broadly defined initiatives for interdisciplinarity. One way to achieve this organizationally, is by moving from the professorial chair-based system to a department-based system (or a hybrid form), where task division is organized collectively instead of being institutionalized. Alternatively, a centralized decision-making unit could be installed to monitor the coordination costs incurred and to prevent new appointments from defining narrower research areas (than their predecessors) by default. Also, the composition of research areas (based on a requirement analysis of the gaps in the subject areas covered and connected) could be guided by such a centralized unit. This should help facilitate collaboration between the institutionalized tasks within the department or faculty as a whole and assure that coordination costs are kept in check.

The distributions of the efficiency results for the full, robust and partial frontier in both publication and citation productivity models suggest that defining joint benchmarks for the pre- and post-war period provides meaningful results. This might have been supported by the purposefully introduced sampling bias considering only highly ranked universities. As expected, efficiency estimates vary moderately within the sample, with an outstanding outlier role of only very few institutions (e.g., Harvard university in post-war citation productivity model). The efficiency results do also reflect the shift from science being dominated by European institutions in the pre-war period to US institutions dominating in the post-war period. This concerns both efficiency and sheer size.

Nonetheless, as already pointed out, a few limitations of this approach exist, which need to be addressed at this point. Even though the here introduced dataset comprises microdata for 46,394 professors based on 10,167 entries, they effectively concern only 20 different institutions. Certainly, the number of observations for pre-war (N = 55) and post-war (N = 111) period, as well as the number of observations for the clusters in the full-frontier analysis (N = 59, 52 and 55 respectively), pose the biggest limitation of this work. Also, even though it could be shown that the way the methodology is set up enables to define very targeted technologies (which mainly vary according to time and institutional type through the different size dimension) and thus a meaningful benchmark for the joint frontiers spanning over large time periods, the interactions of DoL and Spec. components with input size point at the limits of the analysis given the restricted sample size. Certainly, in an ideal world with unlimited access to data one would increase the number of institutions, points in time and define narrower time periods for the measurement of joint frontiers. Also, as already addressed, employing student numbers to calculate joint research and teaching models to investigate the universities' production process would have been desirable. Another issue that was raised above is that it cannot be ruled out that some of the effects observed for the components of DoL and Spec. in the different models might stem from information that could not be accounted for due to restricted access to data. This above all concerns factors known to influence universities' performance such as financial means and numbers of other academic staff, which may explain heterogeneity in professorial staff's quality or productivity levels. The latter seems to be particularly true for the effects obtained for task division, which revealed the greatest interaction with input size. In particular in the citation productivity model, one might question whether it is indeed task division that explains the positive effect of size or if it is the other way round and the positive effect of size is caused by another factor not accounted for, which is reflected by the task division. Finally, for the reasons motivated in section 3.1.1, the empirical analysis is limited to engineering and natural sciences disciplines and thus the conclusions regarding the hypothesis are not valid for social sciences or humanities.

Regardless of those limitations, we are confident in stating that the empirical analysis provides sufficient evidence that DoL and Spec. are important determinants of epistemic outcomes. Not only do we find local effects of the components considered on efficiency estimates, but we find reasonable and consistent relationships in the sense of the nonparametric fits either approximating linear or inverted U-shaped functions, which are not random. Both marginal and joint effects obtained, coincide with the theoretical expectations defined in chapter 2 and are robust for both a model incorporating publications and one employing citations as output, which are both considered to be important epistemic outcomes regularly assessed in all science studies' branches.

6. Conclusions

Regarding the research question of this work, it can be concluded that institutional division of labor and specialization are neglected determinants of epistemic outcomes, which bear the potential to explain pathologies (such as declining productivity levels) within the scientific production process. The latter phenomena should be accounted for in the science studies to ensure that institutions are designed to promote efficient forms of scientific collaboration and mitigate potential path dependencies and increasing coordination costs due to an ongoing topical concentration and specialization depth. The empirical findings of this work cast doubt on the plausibility of the rational theory-based models dominating the science studies, which promote the idea of the efficiently functioning self-governed scientific community. The results rather suggest that the latter is impeded by the institutional arrangements with which scientific institutions are confronted. In particular the findings of the descriptive analysis suggest that the here promoted idea for specialization concentration (mainly depending on institutional prerequisites), have more explanatory power for predicting institutional specialization than the credit maximization rationale. Finally, in the cluster analysis it was shown that institutional types with the highest task differentiation are indeed the ones with the highest specialization depth and vice versa.

The analysis based on the conditional efficiency framework reveals that the effect of DoL and Spec. on epistemic outcomes follows functional relationships and structurally differs according to time and university type. The effects obtained are consistent with economic theory, as we find higher productivity levels for universities that keep their coordination costs in check, either by organizing their researchers in institutional settings that mitigate coordination costs (lower specialization depth, academic department) or by promoting coordination efforts (higher specialization depth, boundary-spanning chairs).

Apart from the latter insights, this work's contribution to the literature is threefold:

I. First, in the theoretical line of thought, it was argued that topics concerned with scientific collaboration could benefit from integrating the theory on institutional division of labor and it was shown how it can be operationalized in a quantitative empirical analysis. An analysis of institutional DoL and Spec. may be seen as a more concrete and granular equivalent to the analysis of interdisciplinarity, which is often operationalized with indicators based on (bibliometric) output data and for which the trends obtained in this work deviated from the ones obtained on the more granular levels and conveyed less clear information (see 3.3.1.1). Further, it closes the research gap documented between the micro-level of collaboration e.g., in the laboratory or team and the macro-level perspective of the overall institutional specialization or broadly defined disciplines.

II. Second contribution is the here introduced new dataset based on documentation of 2,549 semantically distinct denominations in 10,167 entries representing 46,394 university professors affiliated with 20 regularly top-ranked universities. This dataset allows to measure division of labor and specialization for the period 1890 to 2020 (, therefore accounting for a large share of all of modern science), employing an institutional, input-based perspective. The latter is valuable insofar as it allows to complement studies based on the 'bibliometric hypothesis' with an institutional perspective, as demanded by authors in the science studies (e.g., Bornmann et al. 2023). The denominations were coded according to the frequently employed *Web of Science* classification scheme to guarantee for a good compatibility with existing bibliometric data and to assess DoL and Spec. on different levels of granularity. Further, the dataset may be used to analyze different aspects not covered in this work, like for example approaching the denominations from a linguistic perspective or analyzing semantic diversity in this context.

III. Finally, third, this work employed the advanced nonparametric conditional efficiency framework to overcome the issues linked to employing traditional efficiency methods like the Data Envelopment Analysis and the flaws of the so-called two stage approach (Simar and Wilson 2011). Applications of the latter are still scarce (, supposedly because an implementation in readily available statistical packages is still missing) and even scarcer in context of performance measurement of higher education institutions. Also, operationalization of division of labor and specialization in four conceptually distinct indicators goes beyond the traditionally employed diversity measures.

Taking this work as point of departure, future research could overcome some of the limits addressed by expanding the dataset and increasing the number of observations in institutions. In order to provide targeted policy implications, the scope could for example be limited to European institutions for a period concerning only recent years. Further, existing micro data e.g., collected in the ETER and Aquameth project could be integrated to model the effect of DoL and Spec. on scientific productivity more holistically, considering other important determinants of epistemic outcomes such as student numbers, funding, etc. This way, the science studies might be able to make a huge contribution towards resolving some of the pathologies in science observed and supporting that disruptiveness and innovativeness is maintained in the upcoming years.

C. References

Abramo, G. and D'Angelo, C.A. (2023): Warnings of declining research productivity: Does Italy buck the trend? *Research Evaluation*, *32*(3), 557-565.

Adams, J.D., Black, G.C., Clemmons, J.R. and Stephan, P.E. (2005): Scientific teams and institutional collaborations: Evidence from US universities, 1981-1999. *Research Policy*, 34(3), 259-285.

Al-Hamdi, M.T. (2006): Ibn Khaldun: The father of the division of labor. International Conference on Ibn Khaldun, Madrid (Vol. 23, 11/2016).

Alexander, J.M., Himmelreich, J. and Thompson, C. (2014): Epistemic Landscapes, Optimal Search, and the Division of Cognitive Labor. *Philosophy of Science*, 82, 424-453.

Apostel, L., Berger, G., Briggs, A., et al. (eds.) (1972): *Interdisciplinarity: Problems of teaching and research in universities.* Paris: Organization for Economic Cooperation and Development.

Aragon, Y., Daouia, A. and Thomas-Agnan, C. (2005): Nonparametric frontier estimation: a conditional quantile-based approach. *Econometric Theory*, 21(2), 358-389.

Bacon, F. (1902): *Novum organum. Edited by Joseph Devey*. New York: P.F. Collier.

Bădin, L., Daraio, C. and Simar, L. (2010): Optimal bandwidth selection for conditional efficiency measures: A data-driven approach. *European Journal of Operational Research*, 201(2), 633-640.

Bădin, L., Daraio, C. and Simar, L. (2012): How to measure the impact of environmental factors in a nonparametric production model. *European Journal of Operational Research*, 223(3), 818-833.

Bădin, L., Daraio, C. and Simar, L. (2014): Explaining inefficiency in nonparametric production models: the state of the art. *Annals of Operations Research*, 214(1), 5-30.

Bădin, L., Daraio, C. and Simar, L. (2019): A bootstrap approach for bandwidth selection in estimating conditional efficiency measures. *European Journal of Operational Research*, 277(2), 784-797.

Bammer, G. (2017): Toward a new discipline of integration and implementation sciences. In: Frodeman, R. (ed.), *The Oxford Handbook of Interdisciplinarity*. Oxford: Oxford University Press, 525-529.

Banker, R.D., Charnes, A. and Cooper, W.W. (1984): Some models for estimating technical and scale inefficiencies in data envelopment analysis. *Management Science*, 30(9), 1078-1092.

Bechtel, W. (1993): Integrating sciences by creating new disciplines: The case of cell biology. *Biology and Philosophy*, 8, 277-299.

Becker, G.S. and Murphy, K.M. (1992): The division of labor, coordination costs, and knowledge. *The Quarterly Journal of Economics*, 107(4), 1137-1160.

Bedessem, B. (2019): The division of cognitive labor: Two missing dimensions of the debate. *European Journal for Philosophy of Science*, 9, 1-16.

Ben-David, J. and Collins, R. (1974): Soziale Faktoren im Ursprung einer neuen Wissenschaft: der Fall der Psychologie. In: Weingart, P. (ed.), *Wissenschaftssoziologie II. Determinanten wissenschaftlicher Entwicklung*. Frankfurt am Main: Athenäum Verlag, 122-152.

Bird, A. (2007): What is scientific progress? *Noûs*, 41(1), 64-89.

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

Bonaccorsi, A. and Daraio, C. (2007): *Universities and strategic knowledge creation. Specialization and performance in Europe*. Cheltenham: Edward Elgar Publisher.

Bonaccorsi, A., Daraio, C., Fantoni, S., Folli, V., Leonetti, M. and Ruocco, G. (2017): Do social sciences and humanities behave like life and hard sciences? *Scientometrics*, 112, 607-653.

Bongioanni, I., Daraio, C. and Ruocco, G. (2014): A quantitative measure to compare the disciplinary profiles of research systems and their evolution over time. *Journal of Informetrics*, 8, 710-727.

Bornmann, L. and Mutz, R. (2014): Growth rates of modern science: A bibliometric analysis. *Journal of the Association for Information Science and Technology*, 66(11), 2215-2222.

Bornmann, L. and Haunschild, R. (2019): Societal impact measurement of research papers. In: Glänzel, W., Moed, H.F., Schmoch, U. and Thelwall, M. (eds.): *Springer Handbook of Science and Technology Indicators*, Cham: Springer Nature, 609-632.

Bornmann, L., Gralka, S., de Moya Anegón, F. and Wohlrabe, K. (2023): Efficiency of universities and research-focused institutions worldwide: The introduction of a new input indicator reflecting institutional staff numbers. *Journal of Informetrics*, 17(2), 1-17.

Boulakia, J.D.C. (1971): Ibn Khaldun: a fourteenth-century economist. *Journal of Political Economy*, 79(5), 1105-1118.

Bourdieu, P. (1975): The specificity of the scientific field and the social conditions of the progress of reason. *Social Science Information*, 14(6), 19-47.

Bourdieu, P. (1988): *Homo Academicus*. Translated by Bernd Schwibs. 1st ed. Frankfurt am Main: Suhrkamp Verlag.

Bruni, R., Daraio, C. and Aureli, D. (2021): Information reconstruction in educational institutions data from the European tertiary education registry. *Data in Brief*, 34, 1-5.

Bücher, K. (1922): *Die Entstehung der Volkswirtschaft: Vorträge und Aufsätze*. 16th ed. Tübingen: Laupp'sche Buchhandlung.

Catalano, G., Daraio, C. and Quaglia, G. (2019): Matching education and scientific specialization of European universities: A Micro-based Country Level Analysis. *17th International Conference on Scientometrics and Informetrics - ISSI 2019-Proceedings*, 2, 2314-2319.

Cauwels, P. and Sornette, D. (2022): Are 'flow of ideas' and 'research productivity' in secular decline? *Technological Forecasting and Social Change*, *174*, 1-17.

Cazals, C., Florens, J.P. and Simar, L. (2002). Nonparametric frontier estimation: a robust approach. *Journal of Econometrics*, 106(1), 1-25.

Chambers, R.G., Chung, Y. and Färe, R. (1996): Benefit and distance functions. *Journal of Economic Theory*, 70(2), 407-419.

Charnes, A., Cooper, W.W. and Rhodes, E. (1978): Measuring the efficiency of decision-making units. *European Journal of Operational Research*, 2(6), 429-444.

Chu, J.S. and Evans, J.A. (2021): Slowed canonical progress in large fields of science. *Proceedings of the National Academy of Sciences*, 118(41), 1-5.

Clark, T.N. (1974): Die Stadien wissenschaftlicher Institutionalisierung. In: Weingart, P. (ed.): *Wissenschaftssoziologie II. Determinanten wissenschaftlicher Entwicklung*. Frankfurt am Main: Athenäum Verlag, 105-121.

Cohen, E. (2021): *The University and its Boundaries: Thriving or Surviving in the 21st Century*. Oxfordshire: Routledge.

Cooper, W.W., Seiford, L.M. and Tone, K. (2007): Data envelopment analysis: a comprehensive text with models, applications, references and DEA-solver software. New York: Springer.

Crow, M.M. and Dabars, W.B. (2017): Interdisciplinarity and the Institutional Context of Knowledge in the American Research University. In: Frodeman, R. and Thompson Klein (eds.), *The Oxford Handbook of Interdisciplinarity*. Oxford: Oxford University Press, 471-484.

Cui, H., Wu, L. and Evans, J.A. (2022): Aging scientists and slowed advance. *arXiv preprint*, https://doi.org/10.48550/arXiv.2202.04044, 1-37.

Cummings, J.N. and Kiesler, S. (2014): Organization theory and the changing nature of science. *Journal of Organization Design*, 3(3), 1-16.

CWTS [Centre for Science and Technology Studies] (2024): *NOWT-WoS classification.* Available at: https://www.cwts.nl/pdf/nowt_classification_sc.pdf [Last accessed: 03/04/2024]

D'Agostino, F. (2009): From the organization to the division of cognitive labor. *Politics, Philosophy & Economics*, 8(1), 101-129.

Daouia, A. and Simar, L. (2007): Nonparametric efficiency analysis: A multivariate conditional quantile approach. *Journal of Econometrics*, 140(2), 375-400.

Daraio, C. (2019): Econometric approaches to the measurement of research productivity. In: Glänzel, W., Moed, H.F., Schmoch, U. and Thelwall, M. (eds.): *Springer Handbook of Science and Technology Indicators*, Cham: Springer Nature, 633-666.

Daraio, C. and Bonaccorsi, A. (2016): Beyond University Rankings? Generating New Indicators on Universities by Linking Data in Open Platforms. *Journal of the Association for Information Science and Technology*, 68(2): 508-529.

Daraio, C. and Simar, L. (2007): *Advanced Robust and Nonparametric Methods in Efficiency Analysis: Methodology and Applications*. New York: Springer.

Daraio, C. and Simar, L. (2014): Directional distances and their robust versions: Computational and testing issues. *European Journal of Operational Research*, 237(1), 358-369.

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

Daraio, C., Bonaccorsi, A. and 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., Geuna, A., Lepori, B., Bach, L. et al. (2011): The European university landscape: A micro characterization based on evidence from the Aquameth project, *Research Policy*, 40(1), 148-164.

Daraio, C., Fabbri, F., Gavazzi, G., Grazia Izzo, M., Leuzzi, L. et al. (2018): Assessing the interdependencies between scientific disciplinary profiles. *Scientometrics*, 116: 1785-1803.

Daraio, C., Lenzerini, M., Leporelli, C., Moed, H.F., Naggar, P. et al. (2015c): Sapientia: the Ontology of Multi-dimensional Research Assessment. Proceedings of the 15th International Society for Scientometrics and Informetrics Conference (06/2015).

Daraio, C., Simar, L. and Wilson, P.W. (2020): Fast and efficient computation of directional distance estimators. *Annals of Operations Research*, 288(2), 805-835.

Daraio, C., Simar, L. and Wilson, P.W. (2018): Central limit theorems for conditional efficiency measures and tests of the 'separability' condition in non-parametric, two-stage models of production. *The Econometrics Journal*, 21(2), 170-191.

De Langhe, R. (2010): The division of labour in science: the tradeoff between specialization and diversity. *Journal of Economic Methodology*, 17(1), 37-51.

De Langhe, R. and Greiff, M. (2009): Standards and the distribution of cognitive labour. *Logic Journal of the IGPL*, 18(2), 278-293.

De Oliveira, V.M. and Campos, P.R. (2019): The emergence of division of labor in a structured response threshold model. *Physica A: Statistical Mechanics and Its Applications*, 517, 153-162.

De Ridder, J. (2014): Epistemic dependence and collective scientific knowledge. *Synthese*, 191, 37-53.

Debreu, G. (1951): The Coefficient of Resource Utilization. *Econometrica*, 19(3), 273-292.

Dellsén, F., Lawler, I. and Norton, J. (2022): Thinking about progress: From science to philosophy. *Noûs*, 56(4), 814-840.

Durkheim, E. (1966): *The Division of Labor in Society. Translated by George Simpson*. 4th ed. Toronto: Collier-Macmillan.

Eaton, C. and Stevens, M.L. (2020): Universities as peculiar organizations. Sociology Compass, 14(3), 1-13.

Ervin, D.J. (1987): Interdependence and Differentiation as Components of the Division of Labor. *Social Science Quarterly*, 68(1), 177.

Farrell, M.J. (1957) The measurement of productive efficiency. *Journal of the Royal Statistical Society: Series A (General)*, 120(3), 253-281.

Ferguson, A. (1767) *An Essay on the History of Civil Society*. 5th ed. London and Edinburgh: Cadell, Creech and Bell.

Furman, J.L. and Stern, S. (2011): Climbing atop the shoulders of giants: The impact of institutions on cumulative research. *American Economic Review*, 101(5), 1933-1963.

Gibbs, J.P. and Poston, D.L. (1975): The division of labor: Conceptualization and related measures. *Social Forces*, 53(3), 468-476.

Gibbs, J.P. and Martin, W.T. (1962): Urbanization, technology, and the division of labor: International patterns. *American Sociological Review*, 27(5), 667-677.

Goldman, A. and Blanchard, T. (2016): Social epistemology. In E.N. Zalta (ed.), The Stanford encyclopedia of philosophy, winter 2016 edition. https://plato.stanford.edu-/entries/epistemology-social/. Accessed 24 Sept 2018.

Goldman, A. and O'Connor, C. (2021) Social Epistemology. In: Zalta, E.N. (ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2021 Edition), URL = <https://plato.stanford.edu-/archives/win2021/entries/epistemology-social/>.

Goldman, A.I. and Shaked, M. (1991): An Economics Model of Scientific Activity and Truth Acquisition. *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition*, 63(1), 31-55.

Habgood-Coote, J. (2019): Group knowledge, questions, and the division of epistemic labour. *Ergo: An Open Access Journal of Philosophy*, 6, 1-45.

Häussler, C. and Sauermann, H. (2020): Division of labor in collaborative knowledge production: The role of team size and interdisciplinarity. *Research Policy*, 49(6), 1-20.

Herberholz, L. and Wigger, B.U. (2021): Efficiency of European universities: A comparison of peers. *Research Policy*, 50(9), 1-13.

Hong, L. and Page, S. (2004): Groups of diverse problem solvers can outperform groups of high-ability problem solvers. *Proceedings of the National Academy of Sciences*, 101(46), 16385-16389.

Jabbehdari, S. and Walsh, J.P. (2017): Authorship norms and project structures in science. *Science, Technology and Human Values*, 42(5), 872-900.

Jacobs, J.A. (2017): The need for disciplines in the modern research university. In, Frodeman, R. (ed.): *The Oxford handbook of interdisciplinarity*, 2nd ed., Oxford: Oxford University Press, 35-39.

Jansen, D. (1995): Convergence of basic and applied research? Research orientations in German high-temperature superconductor research. *Science, Technology, & Human Values*, 20(2), 197-233.

Johnson, A.L. (2007): Stochastic DEA. The Regression-Based Approach. In: Tone, K. (ed.), *Advances in DEA theory and applications*. Chichester: Wiley, 85-99.

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

Jones, B.F. and Weinberg, B.A. (2011): Age dynamics in scientific creativity. *Proceedings of the National Academy of Sciences*, 108(47), 18910-18914.

Kaiser, A. (2023) Über den Einfluss von Arbeitsteilung auf Wirklichkeitskonstruktion in der Wissenschaft. In: Schaffer, A. and Lang, E. (eds.), *Die Kunst vom Finden und Bauen der Wirklichkeit*. Marburg: Metropolis-Verlag.

Kamijo, Y. and Nakama, D. (2023): Designing division of labor with strategic uncertainty within organizations: Model analysis and a behavioral experiment. *Journal of Economics & Management Strategy*, 32(2), 257-272.

Kitcher, P. (1990): The division of cognitive labor. *The Journal of Philosophy*, 87(1), 5-22.

Knorr-Cetina, K. (1984) *Die Fabrikation von Erkenntnis. Zur Anthropologie der Naturwissenschaft. Vorwort von Rom Harré.* 1st. edition. Frankfurt am Main: Suhrkamp.

Koopmans, T.C. (1951): Efficient allocation of resources. *Econometrica: Journal of the Econometric Society*, 455-465.

Kuhn, T.S. (1976): *Die Struktur wissenschaftlicher Revolutionen*. 2nd ed. Frankfurt am Main: Suhrkamp.

Kukula, R. (ed.) (1891-1930) *Minerva. Jahrbuch der gelehrten Welt.* 1-30. Berlin; Leipzig: De Gruyter (0076-8960).

Kwiek, M. (2018): *Changing European Academics: A Comparative Study of Social Stratification, Work Patterns and Research Productivity*. 1st ed. Oxfordshire: Routledge.

Larivière, V., Desrochers, N., Macaluso, B., Mongeon, P., Paul-Hus, A. and Sugimoto, C.R. (2016): Contributorship and division of labor in knowledge production. *Social Studies of Science*, 46(3), 417-435.

Latour, B. and Woolgar, S. (1986): Laboratory Life – The Construction of Scientific Facts. Introduction by Jonas Salk. With a new postscript by the authors. Princeton: Princeton University Press.

Leahey, E. and Reikowsky, R.C. (2008): Research specialization and collaboration patterns in sociology. *Social Studies of Science*, 38(3), 425-440.

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

Lee, Y.N., Walsh, J.P. and Wang, J. (2015): Creativity in scientific teams: Unpacking novelty and impact. *Research Policy*, 44(3), 684-697.

Lepori, B., Geuna, A. and Mira, A. (2019): Scientific output scales with resources. A comparison of US and European universities. *PloS one*, 14(10), 1-18.

Leydesdorff, L. (2021): *The evolutionary dynamics of discursive knowledge: Communication-theoretical perspectives on an empirical philosophy of science.* Cham: Springer Nature.

Leydesdorff, L. and Milojević, S. (2012): Scientometrics. *arXiv preprint*, https://doi.org/-10.48550/arXiv.1208.4566, 1-18.

Leydesdorff, L. and Rafols, I. (2011): Indicators of the interdisciplinarity of journals: Diversity, centrality, and citations. *Journal of Informetrics*, 5(1), 87-100.

Leydesdorff, L., Wagner, C.S. and Bornmann, L. (2019): Interdisciplinarity as diversity in citation patterns among journals: Rao-Stirling diversity, relative variety, and the Gini coefficient. *Journal of Informetrics*, 13(1), 255-269.

Li, J., Yin, Y., Fortunato, S. and Wang, D. (2019): A dataset of publication records for Nobel laureates. *Scientific Data*, 6(1), 33.

Li, Q. and Racine, J.S. (2023): *Nonparametric econometrics: theory and practice*. Princeton: Princeton University Press.

Li, Q., Lin, J. and Racine, J.S. (2013): Optimal bandwidth selection for nonparametric conditional distribution and quantile functions. *Journal of Business & Economic Statistics*, 31(1), 57-65.

Lin, Y., Evans, J.A. and Wu, L. (2022): New directions in science emerge from disconnection and discord. *Journal of Informetrics*, 16(1), 1-12.

Liu, J.S., Lu, L.Y., Lu, W.M. and Lin, B.J. (2013): Data envelopment analysis 1978-2010: A citation-based literature survey. *Omega*, 41(1), 3-15.

Liu, X., Yi, B., Li, M. and Li, J. (2021): Is interdisciplinary collaboration research more disruptive than monodisciplinary research? *Proceedings of the Association for Information Science and Technology*, 58(1), 264-272.

López-Illescas, C., de Moya-Anegón, F. and Moed, H.F. (2011): A ranking of universities should account for differences in their disciplinary specialization. *Scientometrics*, 88(2), 563-574.

Lu, C., Zhang, Y., Ahn, Y.Y., Ding, Y., Zhang, C. and Ma, D. (2020): Co-contributorship network and division of labor in individual scientific collaborations. *Journal of the Association for Information Science and Technology*, 71(10), 1162-1178.

Macaluso, B., Larivière, V., Sugimoto, T. and Sugimoto, C.R. (2016): Is science built on the shoulders of women? A study of gender differences in contributorship. *Acad. Med.*, 91(8), 1136-1142.

Mäki, U. (2005): Economic Epistemology: Hopes and Horrors. *Episteme*, 1(3), 211-222

Mastromarco, C., Simar, L. and Wilson, P.W. (2019): Nonparametric Statistical Analysis of Production. In: Ten Raa, T. and Greene, W.H. (eds.), *The Palgrave Handbook of Economic Performance Analysis*, London: Palgrave Macmillan, 301-382.

Mayo-Wilson, C., Zollman, K. and Danks, D. (2011): The Independence Thesis: When Individual and Social Epistemology Diverge. *Philosophy of Science*, 78, 653-677.

Meier, S., Stephenson, M. and Perkowski, P. (2019): Culture of trust and division of labor in nonhierarchical teams. *Strategic Management Journal*, 40(8), 1171-1193.

Merton, R.K. (1973e): The normative structure of science. In: Merton, R.K. and Storer N.W. (eds.), *The sociology of science: theoretical and empirical investigations. Edited and with an introduction by Norman W. Storer.* Chicago and London: Chicago University Press, 267-280.

Merton, R.K. (1973) Priorities in Scientific Discovery. In: Merton, R.K. and Storer N.W. (eds.), *The sociology of science: theoretical and empirical investigations. Edited and with an introduction by Norman W. Storer.* Chicago and London: Chicago University Press, 286-324.

Merton, R.K. (1973a) Behavior Patterns of Scientists. In: Merton, R.K. and Storer N.W. (eds.), *The sociology of science: theoretical and empirical investigations. Edited and with an introduction by Norman W. Storer.* Chicago and London: Chicago University Press, 325-342.

Merton, R.K. (1973b) Singletons and Multiples in Scientific Discovery: A Chapter in the Sociology of Science. In: Merton, R.K. and Storer N.W. (eds.), *The sociology of science: theoretical and empirical investigations. Edited and with an introduction by Norman W. Storer*. Chicago and London: Chicago University Press, 343-370.

Merton, R.K. (1973c) Multiple Discoveries as Strategic Research Site. In: Merton, R.K. and Storer N.W. (eds.), *The sociology of science: theoretical and empirical investigations. Edited and with an introduction by Norman W. Storer*. Chicago and London: Chicago University Press 371-382.

Merton, R.K. (1973d) The Matthew Effect in Science. In: Merton, R.K. and Storer N.W. (eds.), *The sociology of science: theoretical and empirical investigations. Edited and with an introduction by Norman W. Storer*. Chicago and London: Chicago University Press, 439-450.

Merton, R.K. and Zuckerman, H. (1973): Institutionalized Patterns of Evaluation in Science. In: Merton, R.K. and Storer N.W. (eds.), *The sociology of science: theoretical and empirical investigations. Edited and with an introduction by Norman W. Storer*. Chicago and London: Chicago University Press, 460-496.

Merton, R.K. and Zuckerman, H. (1973a): Age, Aging and Age Structure in Science. In: Merton, R.K. and Storer N.W. (eds.), *The sociology of science: theoretical and empirical investigations. Edited and with an introduction by Norman W. Storer*. Chicago and London: Chicago University Press, 497-560.

Moed, H.F., de Moya-Anegón, F., López-Illescas, C. and Visser, M. (2011): Is concentration of university research associated with better research performance? *Journal of Informetrics*, 5(4), 649-658.

Morris, R.L. (2020): Increasing Specialization: Why We Need to Make Mathematics More Accessible. *Social Epistemology*, 35, 37-47.

Moschini, U., Fenialdi, E., Daraio, C., Ruocco, G. and Molinari, E. (2020): A comparison of three multidisciplinarity indices based on the diversity of Scopus subject areas of authors' documents, their bibliography, and their citing papers. *Scientometrics*, 125, 1145-1158.

Muldoon, R. (2013): Diversity and the Division of Cognitive Labor. *Philosophy Compass*, 8(2), 117-125.

Muldoon, R. and Weisberg, M. (2011): Robustness and idealization in models of cognitive labor. *Synthese*, 183, 161-174.

Mullins, N.C. (1974): Die Entwicklung eines wissenschaftlichen Spezialgebiets: die Phagen-Gruppe und die Ursprünge der Molekularbiologie. In, Weingart, P.: *Wissenschaftssoziologie II. Determinanten wissenschaftlicher Entwicklung.* Frankfurt am Main: Athenäum Verlag, 184-222.

Oestreich, G. and Degener, F. (1952): *Minerva Jahrbuch der gelehrten Welt*. 34 [1] Berlin: De Gruyter. (9783112300381)

Park, M., Leahey, E. and Funk, R.J. (2023): Papers and patents are becoming less disruptive over time. *Nature*, 613(7942), 138-144.

Parsons, T. and Platt, G.M. (1990): *Die amerikanische Universität. Ein Beitrag zur Soziologie der Erkenntnis. Übersetzt von Michael Bischoff.* 1st ed. Frankfurt am Main: Suhrkamp.

Pastor, J.M. and Serrano, L. (2016): The determinants of the research output of universities: specialization, quality and inefficiencies. *Scientometrics*, 109(2), 1255-1281.

Petrella, R. (1992): Internationalization, multinationalization and globalization of R&D: Toward a new division of labor in science and technology? *Knowledge and Policy*, 5, 3-25.

Pierce, S.J. (1992): On the origin and meaning of bibliometric indicators: Journals in the social sciences, 1886-1985. *Journal of the American Society for Information Science*, 43(7), 477-487.

Polanyi, M., Ziman, J. and Fuller, S. (2000): The Republic of Science: Its Political and Economic Theory. *Minerva*, 1(1), 54-73.

Popper, K. (2002): *Conjectures and refutations: The growth of scientific knowledge*. 2nd ed. Oxfordshire: Routledge.

Pöyhönen, S. (2017): Value of cognitive diversity in science. *Synthese*, 194, 4519-4540.

Rafols, I. and Meyer, M. (2010): Diversity and network coherence as indicators of interdisciplinarity: case studies in bionanoscience. *Scientometrics*, 82(2), 263-287.

Raveendran, M., Puranam, P. and Warglien, M. (2016): Object salience in the division of labor: Experimental evidence. *Management Science*, 62(7), 2110-2128.

Reif, F. (1961): The Competitive World of the Pure Scientist: The quest for prestige can cause conflict between the goals of science and the goals of the scientist. *Science*, 134(3494), 1957-1962.

Romero, F. (2020): The Division of Replication Labor. *Philosophy of Science*, 87, 1014-1025.

Rousseau, R., Zhang, L. and Hu, X. (2019): Knowledge integration: Its meaning and measurement. In: Glänzel, W., Moed, H.F., Schmoch, U. and Thelwall, M. (eds.): *Springer Handbook of Science and Technology Indicators*, Cham: Springer Nature, 69-94.

Rueffler, C., Hermisson, J. and Wagner, G.P. (2012): Evolution of functional specialization and division of labor. *Proceedings of the National Academy of Sciences*, 109(6), 326-335.

Oestreich, G. and Degener, F. (1952): *Minerva Jahrbuch der gelehrten Welt*. 34 [1] Berlin: De Gruyter.

Schuder, W. (ed.) (1956-1969) *Minerva. Jahrbuch der gelehrten Welt.* 34 [2], 35 [1;2] Berlin: De Gruyter. (9783111264806)

Schulz, E., Speekenbrink, M. and Krause, A. (2018): A tutorial on Gaussian process regression: Modelling, exploring, and exploiting functions. *Journal of Mathematical Psychology*, 85, 1-16.

Schumpeter, J.A. (1994): *History of Economic Analysis. Edited from manuscript by Elizabeth B. Schumpeter and with an introduction by Mark Perlman.* 1st edition. London: Routledge.

Seiford, L.M. and Thrall, R.M. (1990): Recent developments in DEA: the mathematical programming approach to frontier analysis. *Journal of Econometrics*, 46(1-2), 7-38.

Shibayama, S., Baba, Y. and Walsh, J.P. (2015): Organizational design of university laboratories: Task allocation and lab performance in Japanese bioscience laboratories. *Research Policy*, 44 (3), 610-622.

Simar, L. and Vanhems, A. (2012): Probabilistic characterization of directional distances and their robust versions. *Journal of Econometrics*, 166(2), 342-354.

Simar, L. and Wilson, P.W. (2011): Two-stage DEA: caveat emptor. *Journal of Productivity Analysis*, 36, 205-218.

Simar, L. and Wilson, P.W. (2015): Statistical approaches for non-parametric frontier models: a guided tour. *International Statistical Review*, 83(1), 77-110.

Singer, D. (2019): Diversity, Not Randomness, Trumps Ability. *Philosophy of Science*, 86, 178-191.

Shkliarevsky, G. (2022): Is Our Research Productivity In Decline? A New Approach in Resolving the Controversy. *arXiv preprint arXiv,* 1-35. DOI: 10.48550/arXiv.2203.01235

Smith, A. (1804) *An Inquiry into the Nature and Causes of the Wealth of Nations.* Volume I. 3rd edition. Hartford: Cooke, Lincoln and Gleason printers.

Smith, D.L. and Snow, R.E. (1976): The division of labor: conceptual and methodological issues. *Social Forces*, 55(2), 520-528.

Smith, M.R. (1978): Coordinating Divided Labor: A Re-Examination. *Sociological Focus*, 11(2), 143-159.

Stephan, P. (2012): How economics shapes science. Cambridge (Mass.) and London: Harvard University Press.

Storer, N.W. (1973) The reward system of science. Prefatory Note. In: Merton, R.K. and Storer N.W. (eds.), *The sociology of science: theoretical and empirical investigations. Edited and with an introduction by Norman W. Storer.* Chicago and London: Chicago University Press, 281-285.

Storer, N.W. (1973a) The processes of evaluation in science. Prefatory Note. In: Merton, R.K. and Storer N.W. (eds.), *The sociology of science: theoretical and empirical investigations. Edited and with an introduction by Norman W. Storer.* Chicago and London: Chicago University Press, 415-418.

Tabner, I. (2007): A review of concentration, diversity or entropy metrics in economics, finance, ecology and communication science. *International Journal of Interdisciplinary Social Sciences: Annual Review*, 2(4), 53-60.

Thoma, J. (2015): The epistemic division of labor revisited. *Philosophy of Science*, 82(3), 454-472.

Van Raan, A. (2019): Measuring science: Basic principles and application of advanced bibliometrics. In: Glänzel, W., Moed, H.F., Schmoch, U. and Thelwall, M. (eds.): *Springer Handbook of Science and Technology Indicators*, Cham, CH: Springer Nature, 237-280.

Viola, M. (2015): Some Remarks on the Division of Cognitive Labor. *A Journal on Research Policy and Evaluation*, 3, 1-14.

Walsh, J.P. and Lee, Y.N. (2015): The bureaucratization of science. *Research Policy*, 44(8), 1584-1600.

Walsh, J.P., Lee, Y.N. and Tang, L. (2019): Pathogenic organization in science: Division of labor and retractions. *Research Policy*, 48(2), 444-461.

Waltman, L. and Van Eck, N.J. (2019): Field normalization of scientometric indicators. In: Glänzel, W., Moed, H.F., Schmoch, U. and Thelwall, M. (eds.): *Springer Handbook of Science and Technology Indicators*, Cham: Springer Nature, 281-300.

Wang, D., Song, C. and Barabási, A.L. (2013): Quantifying long-term scientific impact. *Science*, 342(6154), 127-132.

Wang, J. (2023): An intuitive tutorial to Gaussian processes regression. *Computing in Science & Engineering*, 25(4). 4-11.

Weber, M.M. (1989): Rassenhygiene und nationalsozialistische Ideologie: Ansätze und Probleme der Forschung: Anmerkungen zu Emil Peter Beckers Werk: „Zur Geschichte der Rassenhygiene. Wege ins Dritte Reich“. *Sudhoffs Archiv*, 73(2), 223-232.

Weingart, P. (1974): Wissenschaftlicher Wandel als Institutionalisierungsstrategie. In, Weingart, P. (ed.): *Wissenschaftssoziologie II. Determinanten wissenschaftlicher Entwicklung.* Frankfurt am Main: Athenäum Verlag.

Weisberg, M. and Muldoon, R. (2009): Epistemic landscapes and the division of cognitive labor. *Philosophy of Science*, 76(2), 225-252.

West, A. (1999): The flute factory: an empirical measurement of the effect of the division of labor on productivity and production cost. *The American Economist*, 43(1), 82-87.

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

Zeng, A., Shen, Z., Zhou, J., Fan, Y., Di, Z. et al. (2019): Increasing trend of scientists to switch between topics. *Nature Communications*, 10(1), 1-11.

Zhang, J.H., Htike, K.W., Oad, A. and Zhang H. (2014): Decision Making Problem in Division of Cognitive Labor with Parameter Inaccuracy: Case Studies. *American Journal of Information Systems*, 2(1), 6-10.

Zitt, M., Lelu, A., Cadot, M. and Cabanac, G. (2019): Bibliometric delineation of scientific fields. In: Glänzel, W., Moed, H.F., Schmoch, U. and Thelwall, M. (eds.): *Springer Handbook of Science and Technology Indicators*, Cham: Springer Nature, 25-68.

Zollman, K. (2018): The Credit Economy and the Economic Rationality of Science. *The Journal of Philosophy*, 65(1), 5-33.

D. Annex

Supplementary Material (Printed)

Digitalized Supplementary Material – Permanently available at: https://drive.google.com/drive/folders/1L1A6BF324gej4eVzKwi8lfTkrn9IqrEz?usp=sharing

S1. Sum of all (P_sum), full (P_full), associate (P_asso) and assistant (P_assi) professors as well as sum of denominations (DEN), mean nr. of topics covered (TOPICS), ind. specialization (SPEC), subjects (SUBJ), research fields (FIELD) and disciplines attributable according to period, location and institution

	P_sum	P_full	P_asso	P_assi	DEN	TOPICS	SPEC	SUBJ	FIELD	DISC
1890	**350**	**262**	**290**	**290**	**290**	**1.29**	**1.17**	**1.35**	**1.26**	**1.10**
European	**162**	**120**	**137**	**137**	**137**	**1.45**	**1.26**	**1.46**	**1.34**	**1.16**
ETH Zürich	38	38	31	31	31	1.29	1.32	1.37	1.29	1.11
Ludwig-Maximilians-Universität München	52	33	44	44	44	1.33	1.28	1.31	1.21	1.13
Rheinische Friedrich-Wilhelms-Universität Bonn	32	18	22	22	22	1.22	1.09	1.22	1.13	1.06
Universität Zürich	40	31	40	40	40	1.95	1.36	1.95	1.73	1.33
North American (US)	**167**	**121**	**133**	**133**	**133**	**1.24**	**1.10**	**1.25**	**1.16**	**1.10**
Columbia University	38	34	31	31	31	1.11	1.03	1.11	1.08	1.03
Harvard University	66	44	47	47	47	1.14	1.18	1.21	1.15	1.05
Massachusetts Institute of Technology	30	24	25	25	25	1.20	1.28	1.10	1.10	1.07
University of California - Berkeley	25	11	22	22	22	1.40	1.06	1.44	1.36	1.24
University of Texas at Austin	8	8	8	8	8	1.38	0.94	1.38	1.13	1.13
Oceania	**21**	**21**	**20**	**20**	**20**	**1.19**	**1.14**	**1.33**	**1.29**	**1.05**
University of Sydney	21	21	20	20	20	1.19	1.14	1.33	1.29	1.05

	P_sum	P_full	P_asso	P_assi	DEN	TOPICS	SPEC	SUBJ	FIELD	DISC
1900	**700**	**500**	**510**	**510**	**510**	**1.19**	**1.13**	**1.20**	**1.14**	**1.05**
European	**255**	**167**	**204**	**204**	**204**	**1.29**	**1.16**	**1.29**	**1.19**	**1.05**
ETH Zürich	43	43	34	34	34	1.30	1.25	1.35	1.28	1.05
Georg-August Universität Göttingen	40	24	33	33	33	1.15	1.11	1.25	1.23	1.10
Ludwig-Maximilians-Universität München	61	23	42	42	42	1.33	1.24	1.25	1.16	1.05
Rheinische Friedrich-Wilhelms-Universität Bonn	31	20	24	24	24	1.26	1.03	1.23	1.10	1.03
Universität Zürich	26	16	23	23	23	1.50	1.25	1.38	1.27	1.04
University of Leeds	32	19	28	28	28	1.16	1.09	1.22	1.16	1.09
Uppsala Universitet	22	22	20	20	20	1.32	1.14	1.36	1.14	1.00
North American (US)	**401**	**289**	**271**	**271**	**271**	**1.19**	**1.09**	**1.18**	**1.13**	**1.05**
Columbia University	88	88	57	57	57	1.05	1.15	1.07	1.05	1.00
Harvard University	124	92	75	75	75	1.13	1.21	1.19	1.07	1.02
Massachusetts Institute of Technology	52	22	35	35	35	1.13	1.29	1.08	1.08	1.04
Stanford University	36	15	19	19	19	1.08	0.96	1.06	1.06	1.00
University of California - Berkeley	68	40	56	56	56	1.31	1.18	1.29	1.21	1.09
University of Texas at Austin	19	19	17	17	17	1.26	0.95	1.21	1.21	1.05
University of Washington at Seattle	14	13	12	12	12	1.36	0.90	1.36	1.21	1.14
Oceania	**44**	**44**	**35**	**35**	**35**	**1.09**	**1.13**	**1.14**	**1.09**	**1.05**
University of Sydney	44	44	35	35	35	1.09	1.13	1.14	1.09	1.05

	P_sum	P_full	P_asso	P_assi	DEN	TOPICS	SPEC	SUBJ	FIELD	DISC
1910	**859**	**520**	**571**	**571**	**571**	**1.19**	**1.17**	**1.21**	**1.16**	**1.09**
European	**296**	**186**	**232**	**232**	**232**	**1.25**	**1.15**	**1.25**	**1.17**	**1.05**
ETH Zürich	42	41	36	36	36	1.29	1.25	1.29	1.21	1.05
Georg-August Universität Göttingen	42	26	36	36	36	1.10	1.14	1.19	1.17	1.07
Ludwig-Maximilians-Universität München	80	25	49	49	49	1.34	1.21	1.30	1.21	1.06
Rheinische Friedrich-Wilhelms-Universität Bonn	36	22	24	24	24	1.28	1.01	1.19	1.11	1.03
Universität Zürich	36	25	33	33	33	1.42	1.25	1.31	1.22	1.03
University of Leeds	32	19	28	28	28	1.16	1.09	1.22	1.16	1.09
Uppsala Universitet	28	28	26	26	26	1.21	1.13	1.25	1.11	1.00
North American (US)	**503**	**318**	**299**	**299**	**299**	**1.12**	**1.16**	**1.13**	**1.10**	**1.05**
Columbia University	137	121	65	65	65	1.02	1.10	1.04	1.01	1.01
Harvard University	116	59	77	77	77	1.09	1.29	1.16	1.10	1.03
Massachusetts Institute of Technology	58	31	37	37	37	1.17	1.22	1.14	1.12	1.09
Stanford University	59	36	36	36	36	1.08	1.11	1.10	1.07	1.00
University of California - Berkeley	99	54	66	66	66	1.11	1.22	1.16	1.13	1.07
University of Washington at Seattle	34	17	18	18	18	1.24	1.00	1.21	1.18	1.09
Oceania	**60**	**16**	**40**	**40**	**40**	**1.21**	**1.19**	**1.25**	**1.22**	**1.17**
University of Auckland	3	3	3	3	3	1.33	1.33	1.33	1.33	1.33
University of Sydney	57	13	37	37	37	1.09	1.05	1.16	1.11	1.00

	P_sum	P_full	P_asso	P_assi	DEN	TOPICS	SPEC	SUBJ	FIELD	DISC
1920	**1200**	**746**	**713**	**713**	**713**	**1.21**	**1.15**	**1.20**	**1.14**	**1.08**
European	**430**	**258**	**286**	**286**	**286**	**1.28**	**1.18**	**1.27**	**1.17**	**1.06**
ETH Zürich	43	43	35	35	35	1.30	1.16	1.30	1.23	1.07
Georg-August Universität Göttingen	60	37	38	38	38	1.12	1.05	1.15	1.08	1.03
Ludwig-Maximilians-Universität München	149	52	84	84	84	1.40	1.24	1.36	1.26	1.04
Rheinische Friedrich-Wilhelms-Universität Bonn	84	39	41	41	41	1.31	1.09	1.26	1.15	1.06
Universität Zürich	33	27	32	32	32	1.48	1.36	1.36	1.21	1.03
University of Leeds	30	29	28	28	28	1.07	1.13	1.13	1.10	1.10
Uppsala Universitet	31	31	28	28	28	1.26	1.20	1.32	1.16	1.06
North American (US)	**748**	**466**	**405**	**405**	**405**	**1.13**	**1.16**	**1.12**	**1.09**	**1.02**
CALTECH	19	11	14	14	14	1.26	1.00	1.26	1.21	1.05
Columbia University	247	104	87	87	87	1.03	1.13	1.07	1.04	1.01
Harvard University	163	94	98	98	98	1.05	1.30	1.10	1.06	1.01
Massachusetts Institute of Technology	113	54	66	66	66	1.19	1.37	1.14	1.12	1.03
Stanford University	56	56	30	30	30	1.07	1.08	1.05	1.05	1.04
University of California - Berkeley	105	102	73	73	73	1.13	1.31	1.13	1.11	1.02
University of Texas at Austin	24	24	19	19	19	1.21	0.98	1.17	1.13	1.00
University of Washington at Seattle	21	21	18	18	18	1.05	1.10	1.00	1.00	1.00
Oceania	**22**	**22**	**22**	**22**	**22**	**1.22**	**1.13**	**1.22**	**1.17**	**1.17**
University of Auckland	3	3	3	3	3	1.33	1.33	1.33	1.33	1.33
University of Sydney	19	19	19	19	19	1.11	0.92	1.11	1.00	1.00

	P_sum	P_full	P_asso	P_assi	DEN	TOPICS	SPEC	SUBJ	FIELD	DISC
1950	**4296**	**1844**	**1069**	**1069**	**1069**	**1.12**	**1.06**	**1.15**	**1.10**	**1.03**
European	**586**	**350**	**405**	**405**	**405**	**1.32**	**1.28**	**1.35**	**1.24**	**1.07**
ETH Zürich	71	55	64	64	64	1.38	1.46	1.39	1.27	1.14
Georg-August Universität Göttingen	106	56	64	64	64	1.34	1.21	1.38	1.22	1.10
Ludwig-Maximilians-Universität München	137	64	82	82	82	1.42	1.23	1.36	1.26	1.07
Rheinische Friedrich-Wilhelms-Universität Bonn	117	54	65	65	65	1.33	1.20	1.34	1.24	1.04
Universität Zürich	54	33	48	48	48	1.46	1.40	1.48	1.30	1.07
University of Leeds	64	51	55	55	55	1.19	1.30	1.34	1.25	1.08
Uppsala Universitet	37	37	27	27	27	1.08	1.19	1.16	1.14	1.00
North American (US)	**3443**	**1452**	**613**	**613**	**613**	**1.05**	**1.08**	**1.09**	**1.06**	**1.03**
CALTECH	134	77	49	49	49	1.07	1.11	1.12	1.09	1.04
Columbia University	753	294	96	96	96	1.04	1.07	1.07	1.04	1.03
Harvard University	394	193	119	119	119	1.06	1.17	1.12	1.06	1.03
Massachusetts Institute of Technology	310	100	62	62	62	1.01	1.04	1.05	1.02	1.02
Stanford University	409	165	42	42	42	1.14	0.93	1.11	1.04	1.02
University of California - Berkeley	455	242	82	82	82	1.02	1.12	1.08	1.08	1.05
University of California - Davis	186	66	26	26	26	1.01	1.45	1.03	1.03	1.01
University of California - Los Angeles	239	118	44	44	44	1.03	0.86	1.08	1.08	1.03
University of Texas at Austin	174	78	20	20	20	1.05	1.00	1.10	1.07	1.05
University of Washington at Seattle	389	119	73	73	73	1.12	1.02	1.13	1.06	1.02
Oceania	**267**	**42**	**51**	**51**	**51**	**1.00**	**0.82**	**1.01**	**1.01**	**1.00**
University of Auckland	22	9	12	12	12	1.00	0.64	1.00	1.00	1.00
University of Sydney	245	33	39	39	39	1.00	1.00	1.02	1.01	1.00

	P_sum	P_full	P_asso	P_assi	DEN	TOPICS	SPEC	SUBJ	FIELD	DISC
1960	**4013**	**2804**	**1233**	**1233**	**1233**	**1.13**	**1.11**	**1.15**	**1.10**	**1.04**
European	**986**	**477**	**596**	**596**	**596**	**1.25**	**1.31**	**1.30**	**1.20**	**1.06**
ETH Zürich	125	76	102	102	102	1.26	1.58	1.36	1.29	1.12
Georg-August Universität Göttingen	190	86	96	96	96	1.28	1.23	1.32	1.18	1.06
Ludwig-Maximilians-Universität München	277	97	120	120	120	1.36	1.23	1.33	1.22	1.07
Rheinische Friedrich-Wilhelms-Universität Bonn	195	74	105	105	105	1.30	1.23	1.32	1.19	1.06
Universität Zürich	81	42	67	67	67	1.17	1.33	1.16	1.10	1.04
University of Leeds	64	51	55	55	55	1.19	1.30	1.34	1.25	1.08
Uppsala Universitet	54	51	51	51	51	1.20	1.29	1.30	1.17	1.00
North American (US)	**2641**	**2256**	**561**	**561**	**561**	**1.10**	**0.99**	**1.13**	**1.10**	**1.05**
CALTECH	195	123	54	54	54	1.07	1.07	1.12	1.09	1.06
Columbia University	194	194	70	70	70	1.02	1.05	1.10	1.06	1.03
Harvard University	189	189	86	86	86	1.10	1.13	1.19	1.10	1.06
Massachusetts Institute of Technology	553	240	47	47	47	1.08	0.86	1.10	1.09	1.08
Stanford University	212	212	64	64	64	1.14	0.93	1.13	1.12	1.09
University of California - Berkeley	326	326	72	72	72	1.19	0.94	1.25	1.19	1.09
University of California - Davis	204	204	39	39	39	1.18	1.32	1.22	1.17	1.03
University of California - Los Angeles	316	316	31	31	31	1.15	0.79	1.09	1.04	1.00
University of California - San Diego	71	71	15	15	15	1.03	0.89	1.13	1.13	1.10
University of Texas at Austin	124	124	21	21	21	1.02	0.91	1.03	1.02	1.01
University of Washington at Seattle	257	257	62	62	62	1.11	0.99	1.11	1.08	1.03
Oceania	**386**	**71**	**76**	**76**	**76**	**1.02**	**1.02**	**1.02**	**1.02**	**1.01**
University of Auckland	95	16	19	19	19	1.03	0.99	1.01	1.01	1.01
University of Sydney	291	55	57	57	57	1.01	1.06	1.03	1.02	1.01

	P_sum	P_full	P_asso	P_assi	DEN	TOPICS	SPEC	SUBJ	FIELD	DISC
1970	**4761**	**2355**	**908**	**908**	**908**	**1.14**	**1.11**	**1.17**	**1.11**	**1.04**
European	**1106**	**489**	**470**	**470**	**470**	**1.23**	**1.31**	**1.29**	**1.18**	**1.06**
Georg-August Universität Göttingen	213	95	104	104	104	1.19	1.31	1.30	1.21	1.07
Ludwig-Maximilians-Universität München	359	133	121	121	121	1.31	1.23	1.32	1.20	1.06
Rheinische Friedrich-Wilhelms-Universität Bonn	237	125	114	114	114	1.26	1.30	1.30	1.16	1.04
Universität Zürich	103	73	78	78	78	1.20	1.48	1.26	1.16	1.05
University of Leeds	194	63	53	53	53	1.19	1.26	1.25	1.17	1.07
North American (US)	**3223**	**1778**	**362**	**362**	**362**	**1.10**	**0.98**	**1.14**	**1.09**	**1.07**
CALTECH	173	133	46	46	46	1.03	1.06	1.08	1.06	1.03
Harvard University	301	168	70	70	70	1.07	1.05	1.15	1.04	1.02
Massachusetts Institute of Technology	448	235	55	55	55	1.12	0.86	1.16	1.15	1.13
Stanford University	367	250	26	26	26	1.09	0.78	1.09	1.09	1.09
University of California - Berkeley	515	389	24	24	24	1.26	0.93	1.28	1.22	1.16
University of California - Davis	453	196	39	39	39	1.11	1.14	1.13	1.03	1.02
University of California - San Diego	198	113	38	38	38	1.01	0.87	1.06	1.05	1.03
University of Washington at Seattle	768	294	64	64	64	1.07	1.11	1.15	1.12	1.04
Oceania	**432**	**88**	**76**	**76**	**76**	**1.09**	**1.05**	**1.10**	**1.05**	**1.01**
University of Auckland	139	17	27	27	27	1.11	1.09	1.06	1.06	1.00
University of Sydney	293	71	49	49	49	1.08	1.01	1.13	1.04	1.01

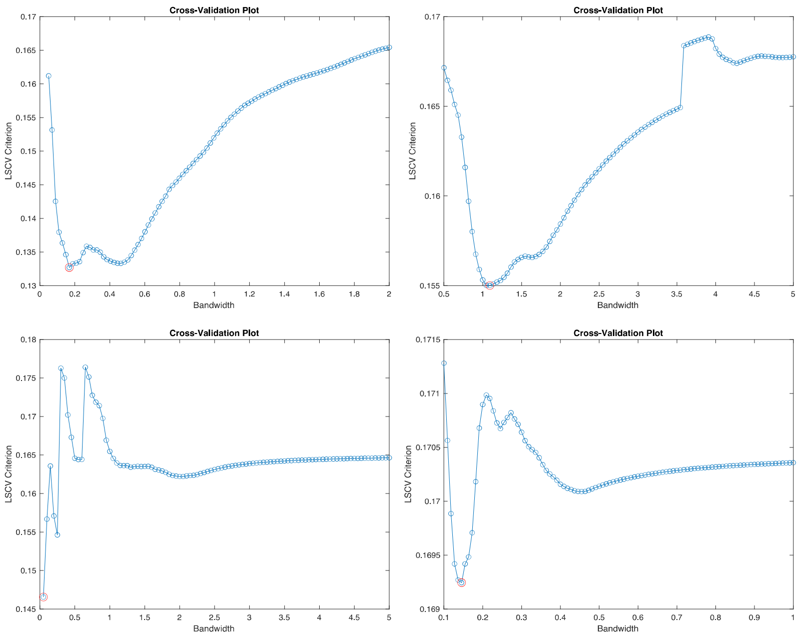
	P_sum	P_full	P_asso	P_assi	DEN	TOPICS	SPEC	SUBJ	FIELD	DISC
1980	**6535**	**3807**	**1040**	**1040**	**1040**	**1.18**	**1.12**	**1.22**	**1.15**	**1.06**
European	**1778**	**1143**	**613**	**613**	**613**	**1.21**	**1.34**	**1.30**	**1.19**	**1.06**
Georg-August Universität Göttingen	377	253	159	159	159	1.21	1.32	1.30	1.20	1.08
Ludwig-Maximilians-Universität München	712	470	160	160	160	1.23	1.21	1.27	1.19	1.07
Rheinische Friedrich-Wilhelms-Universität Bonn	535	319	182	182	182	1.25	1.26	1.30	1.17	1.05
Universität Zürich	154	101	112	112	112	1.14	1.56	1.34	1.19	1.04
North American (US)	**4176**	**2558**	**357**	**357**	**357**	**1.17**	**0.98**	**1.21**	**1.17**	**1.11**
CALTECH	181	134	45	45	45	1.05	1.04	1.11	1.09	1.08
Harvard University	339	183	67	67	67	1.10	1.03	1.16	1.06	1.05
Massachusetts Institute of Technology	491	300	54	54	54	1.13	0.85	1.16	1.15	1.14
Stanford University	405	277	24	24	24	1.11	0.87	1.14	1.11	1.11
University of California - Berkeley	622	463	30	30	30	1.26	1.07	1.31	1.25	1.14
University of California - Los Angeles	791	540	35	35	35	1.35	1.03	1.36	1.32	1.09
University of California - San Diego	285	201	41	41	41	1.21	0.88	1.27	1.25	1.22
University of Washington at Seattle	1062	460	61	61	61	1.12	1.09	1.15	1.14	1.03
Oceania	**581**	**106**	**70**	**70**	**70**	**1.16**	**1.03**	**1.14**	**1.09**	**1.01**
University of Auckland	228	32	23	23	23	1.14	1.02	1.06	1.06	1.00
University of Sydney	353	74	47	47	47	1.17	1.03	1.22	1.13	1.01

	P_sum	P_full	P_asso	P_assi	DEN	TOPICS	SPEC	SUBJ	FIELD	DISC
1990	**7810**	**4661**	**1493**	**1493**	**1493**	**1.20**	**1.16**	**1.26**	**1.17**	**1.07**
European	**2043**	**1454**	**895**	**895**	**895**	**1.20**	**1.39**	**1.34**	**1.21**	**1.06**
Georg-August Universität Göttingen	441	242	196	196	196	1.20	1.32	1.33	1.19	1.09
Ludwig-Maximilians-Universität München	553	511	206	206	206	1.26	1.29	1.37	1.23	1.08
Rheinische Friedrich-Wilhelms-Universität Bonn	680	398	257	257	257	1.23	1.34	1.31	1.19	1.06
Universität Zürich	183	117	131	131	131	1.18	1.54	1.37	1.21	1.03
Uppsala Universitet	186	186	105	105	105	1.11	1.43	1.31	1.20	1.07
North American (US)	**5122**	**3056**	**455**	**455**	**455**	**1.27**	**1.01**	**1.30**	**1.22**	**1.14**
CALTECH	195	142	46	46	46	1.06	1.01	1.10	1.07	1.06
Harvard University	375	175	69	69	69	1.17	1.05	1.24	1.09	1.01
Massachusetts Institute of Technology	468	301	70	70	70	1.22	0.87	1.26	1.24	1.22
Stanford University	434	269	27	27	27	1.18	0.90	1.30	1.22	1.12
University of California - Berkeley	693	491	26	26	26	1.43	1.02	1.34	1.15	1.10
University of California - Davis	927	501	64	64	64	1.46	1.22	1.47	1.38	1.22
University of California - Los Angeles	537	336	33	33	33	1.48	1.06	1.56	1.41	1.12
University of California - San Diego	468	308	35	35	35	1.25	0.89	1.29	1.29	1.27
University of Texas at Austin	633	388	19	19	19	1.28	0.91	1.32	1.27	1.27
University of Washington at Seattle	392	145	66	66	66	1.12	1.13	1.13	1.10	1.02
Oceania	**645**	**151**	**143**	**143**	**143**	**1.14**	**1.08**	**1.14**	**1.10**	**1.01**
University of Auckland	231	39	54	54	54	1.15	1.00	1.10	1.08	1.02
University of Sydney	414	112	89	89	89	1.13	1.16	1.18	1.11	1.01

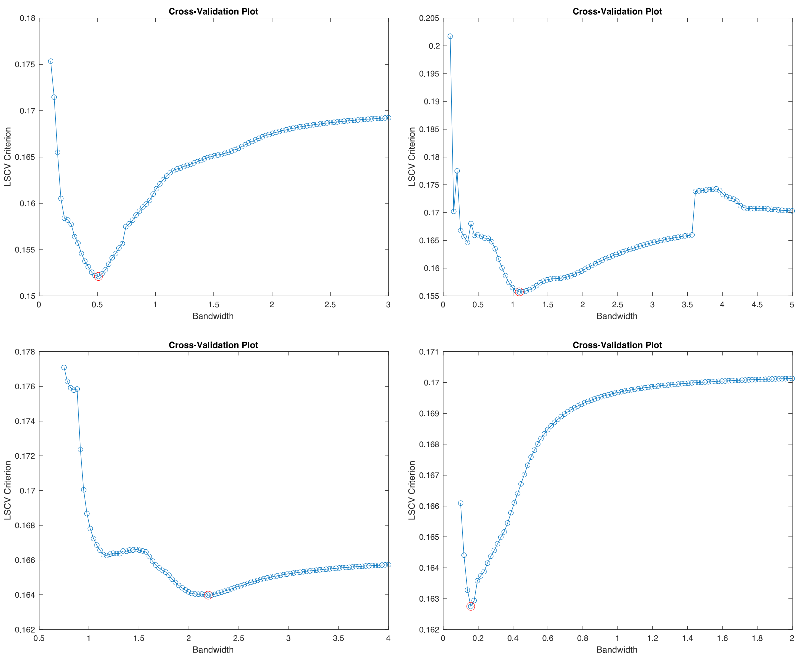
	P_sum	P_full	P_asso	P_assi	DEN	TOPICS	SPEC	SUBJ	FIELD	DISC
2000	**8831**	**6566**	**1975**	**1975**	**1975**	**1.28**	**1.22**	**1.36**	**1.25**	**1.12**
European	**2648**	**2037**	**1328**	**1328**	**1328**	**1.22**	**1.48**	**1.39**	**1.26**	**1.10**
ETH Zürich	189	161	154	154	154	1.17	1.60	1.34	1.28	1.15
Georg-August Universität Göttingen	355	230	180	180	180	1.32	1.50	1.46	1.27	1.14
Ludwig-Maximilians-Universität München	645	590	263	263	263	1.27	1.34	1.38	1.24	1.08
Rheinische Friedrich-Wilhelms-Universität Bonn	720	425	289	289	289	1.22	1.31	1.30	1.19	1.05
Universität Zürich	187	123	150	150	150	1.19	1.56	1.46	1.26	1.07
University of Leeds	216	172	177	177	177	1.14	1.55	1.42	1.32	1.12
Uppsala Universitet	336	336	115	115	115	1.23	1.52	1.38	1.27	1.08
North American (US)	**5729**	**4232**	**508**	**508**	**508**	**1.35**	**1.01**	**1.38**	**1.30**	**1.20**
CALTECH	154	128	49	49	49	1.10	1.04	1.14	1.11	1.11
Harvard University	290	228	89	89	89	1.31	1.11	1.38	1.28	1.17
Massachusetts Institute of Technology	463	350	85	85	85	1.34	0.90	1.34	1.32	1.30
Stanford University	477	359	28	28	28	1.26	0.92	1.36	1.31	1.16
University of California - Berkeley	766	561	27	27	27	1.53	0.95	1.40	1.27	1.19
University of California - Davis	801	592	58	58	58	1.50	1.21	1.46	1.36	1.18
University of California - Los Angeles	862	698	46	46	46	1.52	1.09	1.63	1.44	1.12
University of California - San Diego	420	362	42	42	42	1.42	0.98	1.50	1.39	1.35
University of Texas at Austin	702	396	18	18	18	1.32	0.79	1.35	1.28	1.28
University of Washington at Seattle	794	558	66	66	66	1.20	1.08	1.21	1.19	1.12
Oceania	**454**	**297**	**139**	**139**	**139**	**1.26**	**1.17**	**1.30**	**1.19**	**1.05**
University of Auckland	236	79	32	32	32	1.34	1.01	1.32	1.22	1.06
University of Sydney	218	218	107	107	107	1.17	1.32	1.27	1.16	1.04

	P_sum	P_full	P_asso	P_assi	DEN	TOPICS	SPEC	SUBJ	FIELD	DISC
2010	**6936**	**3944**	**365**	**365**	**365**	**1.25**	**1.18**	**1.33**	**1.29**	**1.18**
European	**56**	**56**	**55**	**55**	**55**	**1.13**	**1.71**	**1.32**	**1.30**	**1.16**
ETH Zürich	56	56	55	55	55	1.13	1.71	1.32	1.30	1.16
North American (US)	**6444**	**3781**	**273**	**273**	**273**	**1.34**	**0.90**	**1.35**	**1.26**	**1.16**
CALTECH	169	139	49	49	49	1.05	1.02	1.11	1.10	1.08
Columbia University	137	96	6	6	6	1.32	0.47	1.32	1.28	1.04
Massachusetts Institute of Technology	442	295	41	41	41	1.33	0.83	1.32	1.31	1.31
Stanford University	660	392	29	29	29	1.25	1.01	1.36	1.29	1.17
University of California - Berkeley	830	580	23	23	23	1.53	0.92	1.37	1.22	1.17
University of California - Los Angeles	918	720	48	48	48	1.62	1.09	1.66	1.46	1.16
University of Texas at Austin	713	387	19	19	19	1.39	0.82	1.42	1.27	1.27
University of Washington at Seattle	2575	1172	58	58	58	1.21	1.06	1.22	1.18	1.06
Oceania	**436**	**107**	**37**	**37**	**37**	**1.29**	**0.93**	**1.32**	**1.31**	**1.21**
University of Auckland	436	107	37	37	37	1.29	0.93	1.32	1.31	1.21
Sample	**46291**	**28009**	**10167**	**10167**	**10167**	**206.72**	**192.76**	**212.52**	**200.37**	**184.70**

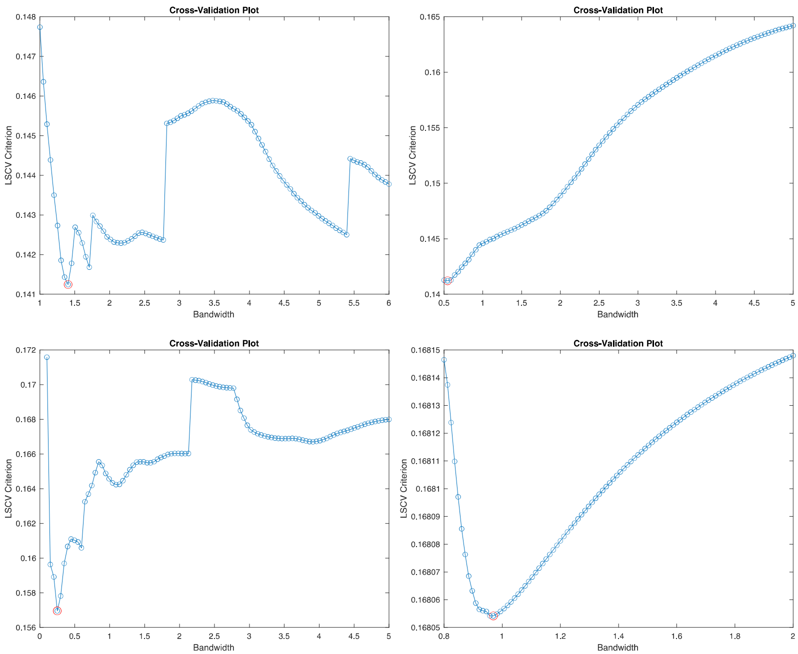
S2. Cross validation plots of pre-war PPM model for task division (upper left), task coordination (upper right), Spec. concentration (lower left) and Spec. gravity (lower right)



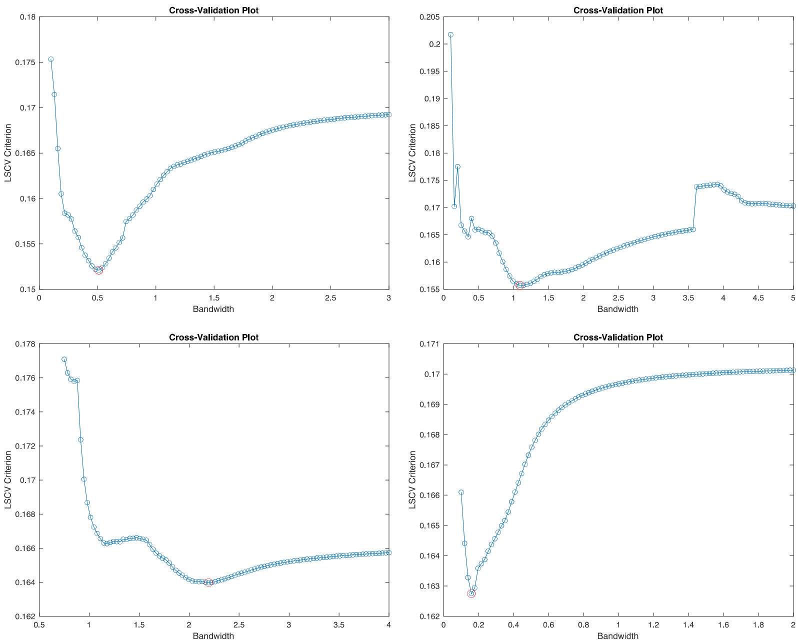
S3. Cross validation plots of pre-war CPM model for task division (upper left), task coordination (upper right), Spec. concentration (lower left) and Spec. gravity (lower right)



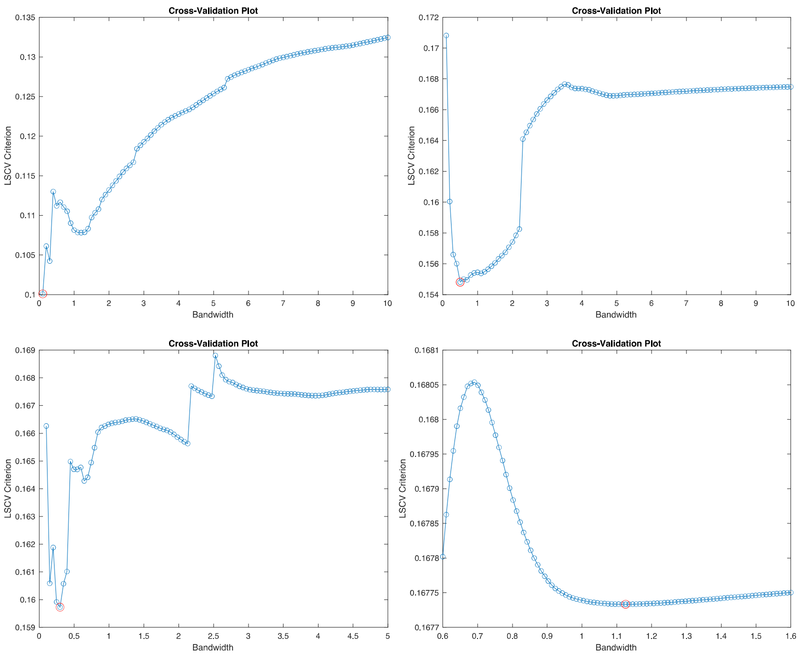
S4. Cross validation plots of post-war PPM model for task division (upper left), task coordination (upper right), Spec. concentration (lower left) and Spec. gravity (lower right)



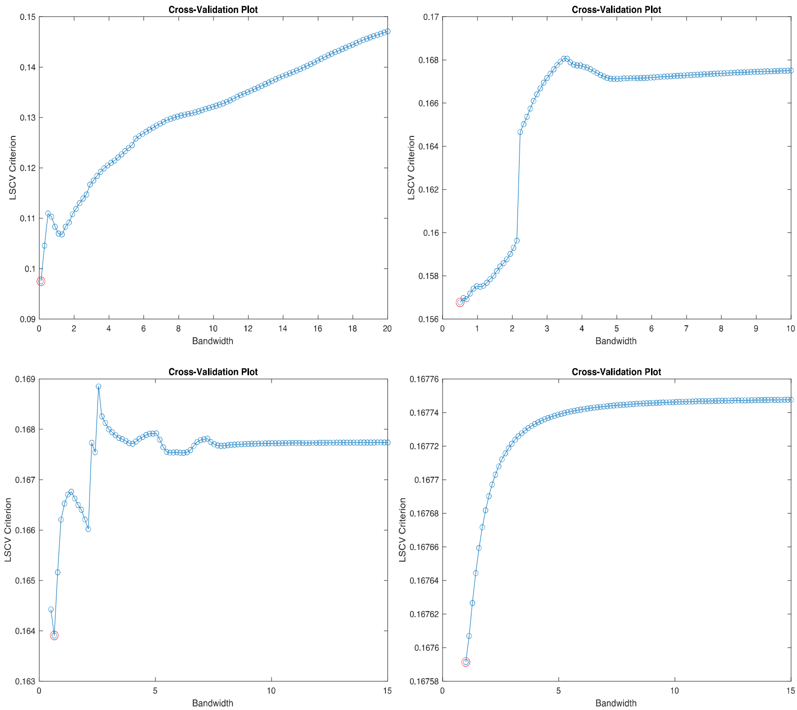
S5. Cross validation plots of post-war CPM model for task division (upper left), task coordination (upper right), Spec. concentration (lower left) and Spec. gravity (lower right)



S6. Cross validation plots of full frontier PPM model for task division (upper left), task coordination (upper right), Spec. concentration (lower left) and Spec. gravity (lower right)



S7. Cross validation plots of full frontier PPM model for task division (upper left), task coordination (upper right), Spec. concentration (lower left) and Spec. gravity (lower right)



S8. Pre-war frontier values of inputs, outputs and efficiency estimates

i	X	Yp	Yc	FDHp	FDHc	Oa99p	Oa99c	Oa90p	Oa90c
1890_col	38	173	405	0.34	0.34	0.34	0.34	0.04	0.04
1890_har	66	361	570	0.40	0.84	0.40	0.84	-0.07	0.43
1890_lmu	52	78	62	0.59	0.59	0.59	0.59	0.30	0.28
1890_mit	30	36	1425	0.20	0.20	0.20	0.20	0.00	-0.22
1890_rfw	32	201	334	0.23	0.23	0.23	0.23	-0.07	0.01
1890_ucb	25	7	9	0.11	0.11	0.11	0.11	0.02	0.00
1890_uos	21	12	9	0.04	0.04	0.04	0.04	-0.02	0.00
1890_uta	8	3	2	0.01	0.01	0.01	0.01	-0.20	-0.20
1890_uzh	40	64	454	0.38	0.38	0.38	0.38	0.14	0.02
1900_col	88	517	1962	0.57	1.24	0.57	1.24	-0.19	0.31
1900_eth	43	6	4	0.43	0.43	0.43	0.43	0.22	0.22
1900_gau	40	472	1386	0.00	0.38	0.00	0.38	-0.47	-0.04
1900_har	124	587	1887	0.79	1.88	0.79	1.88	-0.20	0.36
1900_lee	32	65	207	0.23	0.23	0.23	0.23	0.02	0.02
1900_lmu	61	133	143	0.64	0.75	0.64	0.75	0.34	0.38
1900_mit	52	125	565	0.59	0.59	0.59	0.59	0.28	0.18
1900_rfw	31	343	465	0.00	0.22	0.00	0.22	-0.36	-0.04
1900_sta	36	143	407	0.30	0.30	0.30	0.30	0.05	0.04
1900_ucb	68	43	54	0.88	0.88	0.88	0.88	0.57	0.50
1900_uos	44	20	3	0.45	0.45	0.45	0.45	0.22	0.23
1900_uow	14	18	11	0.00	0.00	0.00	0.00	-0.09	-0.09
1900_upp	22	24	32	0.05	0.05	0.05	0.05	-0.04	-0.01
1900_uta	19	3	12	0.05	0.00	0.05	0.00	0.00	-0.01
1900_uzh	26	79	206	0.13	0.13	0.13	0.13	-0.09	-0.04
1910_col	137	862	3675	0.00	2.12	0.00	2.12	-0.78	0.00
1910_eth	42	137	2427	0.41	0.41	0.41	0.41	0.11	-0.32
1910_gau	42	243	845	0.24	0.41	0.24	0.41	0.00	0.00
1910_har	116	753	3264	0.18	1.74	0.18	1.74	-0.59	0.00
1910_lee	32	133	482	0.23	0.23	0.23	0.23	0.00	-0.02
1910_lmu	80	196	436	0.84	1.09	0.84	1.09	0.38	0.66
1910_mit	58	148	503	0.58	0.70	0.58	0.70	0.29	0.29
1910_rfw	36	218	267	0.30	0.30	0.30	0.30	-0.06	0.07
1910_sta	59	315	684	0.47	0.72	0.47	0.72	-0.02	0.30
1910_ucb	99	191	2242	1.02	1.43	1.02	1.43	0.55	0.12

1910_uos	57	23	86	0.68	0.68	0.68	0.68	0.43	0.30
1910_uow	34	97	150	0.27	0.27	0.27	0.27	0.04	0.05
1910_upp	28	36	184	0.16	0.16	0.16	0.16	0.00	0.00
1910_uzh	36	180	829	0.30	0.30	0.30	0.30	0.00	-0.07
1920_auc	3	5	15	0.00	0.00	0.00	0.00	-0.29	-0.29
1920_cal	19	308	7441	0.00	0.00	0.00	0.00	-0.38	-2.53
1920_col	247	1258	5135	0.22	1.55	0.22	1.55	-1.96	-0.29
1920_eth	43	183	2498	0.43	0.43	0.43	0.43	0.07	-0.30
1920_gau	60	803	5154	0.00	0.74	0.00	0.74	-1.00	-1.27
1920_har	163	1319	13193	0.00	0.00	0.00	0.00	-2.18	-5.70
1920_lee	30	298	1746	0.04	0.20	0.04	0.20	-0.22	-0.23
1920_lmu	149	648	1827	0.56	2.33	0.56	2.33	-0.09	0.59
1920_mit	113	533	4699	0.95	1.69	0.95	1.69	-0.05	-0.69
1920_rfw	84	624	888	0.43	1.17	0.43	1.17	-0.52	0.50
1920_sta	56	719	2703	0.00	0.66	0.00	0.66	-0.90	-0.19
1920_ucb	105	528	4807	0.81	1.54	0.81	1.54	-0.14	-0.76
1920_uos	19	81	184	0.00	0.00	0.00	0.00	-0.16	-0.11
1920_uow	21	161	370	0.04	0.04	0.04	0.04	-0.20	-0.13
1920_upp	31	98	355	0.22	0.22	0.22	0.22	-0.02	0.00
1920_uta	24	46	213	0.09	0.09	0.09	0.09	-0.07	-0.07
1920_uzh	33	465	4104	0.00	0.25	0.00	0.25	-0.55	-1.12

S9. Post-war frontier values of inputs, outputs and efficiency estimates

i	X	Yp	Yc	FDHp	FDHc	Oa99p	Oa99c	Oa90p	Oa90c
1950_auc	22	160	2042	0.00	0.00	-0.04	-0.04	-0.22	-0.22
1950_cal	134	1857	60129	0.08	0.15	0.05	0.05	-0.01	-0.04
1950_col	753	4337	94345	1.52	1.56	1.47	1.52	0.90	0.94
1950_eth	71	1223	18406	0.04	0.03	0.02	0.02	-0.09	-0.09
1950_gau	106	2194	13789	0.04	0.09	0.02	0.04	-0.05	-0.01
1950_har	394	5255	162627	0.62	0.62	0.58	0.58	0.28	0.30
1950_lee	64	1326	23901	0.03	0.03	0.00	0.00	-0.11	-0.11
1950_lmu	137	2281	15944	0.09	0.17	0.03	0.09	-0.03	0.00
1950_mit	310	3466	132883	0.41	0.41	0.37	0.37	0.24	0.24
1950_rfw	117	1782	8598	0.06	0.12	0.04	0.04	-0.03	0.00
1950_sta	409	1883	49175	0.66	0.66	0.62	0.62	0.43	0.40

1950_ucb	455	4675	123444	0.78	0.78	0.74	0.74	0.40	0.44
1950_ucd	186	719	10138	0.22	0.19	0.11	0.09	0.05	0.03
1950_ucl	239	2445	48473	0.22	0.22	0.22	0.21	0.12	0.12
1950_uos	245	1119	12324	0.24	0.24	0.24	0.24	0.13	0.13
1950_uow	389	1568	26630	0.61	0.61	0.57	0.57	0.44	0.41
1950_upp	37	925	19014	0.00	0.00	-0.04	-0.01	-0.18	-0.18
1950_uta	174	889	13345	0.18	0.19	0.10	0.09	0.05	0.01
1950_uzh	54	1153	8026	0.00	0.01	0.00	0.00	-0.14	-0.14
1960_auc	95	633	10010	0.08	0.06	0.08	0.04	-0.03	-0.03
1960_cal	195	4091	157707	0.11	0.11	0.11	0.11	0.00	0.00
1960_col	194	7068	245747	0.10	0.10	0.10	0.06	-0.01	0.00
1960_eth	125	1991	42784	0.06	0.14	0.05	0.06	-0.03	-0.02
1960_gau	190	3402	48516	0.13	0.16	0.09	0.09	0.00	0.01
1960_har	189	11481	518526	0.09	0.09	0.05	0.05	-0.08	-0.08
1960_lee	64	2594	53373	0.00	0.01	-0.01	0.00	-0.11	-0.11
1960_lmu	277	4545	68014	0.32	0.32	0.28	0.28	0.16	0.19
1960_mit	553	8442	361970	1.04	1.04	1.00	1.00	0.51	0.56
1960_rfw	195	3117	31599	0.14	0.17	0.11	0.11	0.01	0.02
1960_sta	212	7210	276782	0.15	0.15	0.11	0.11	0.00	0.01
1960_ucb	326	10423	414889	0.45	0.45	0.41	0.41	0.06	0.11
1960_ucd	204	2832	78266	0.16	0.13	0.13	0.13	0.03	0.03
1960_ucl	316	7963	224971	0.42	0.42	0.38	0.38	0.14	0.24
1960_ucs	71	3008	232926	0.00	0.00	-0.02	-0.10	-0.12	-0.18
1960_uos	291	2842	48352	0.36	0.36	0.32	0.32	0.20	0.21
1960_uow	257	5245	153986	0.27	0.27	0.23	0.23	0.11	0.12
1960_upp	54	2424	59360	0.00	0.00	-0.03	-0.03	-0.14	-0.14
1960_uta	124	2917	77479	0.05	0.13	0.00	0.03	-0.06	-0.05
1960_uzh	81	1664	22363	0.04	0.03	0.04	0.03	-0.07	-0.06
1970_auc	139	2182	44649	0.09	0.16	0.04	0.06	-0.02	-0.02
1970_cal	173	6931	370463	0.05	0.05	0.05	0.01	-0.05	-0.07
1970_gau	213	5671	79166	0.15	0.15	0.15	0.15	0.05	0.05
1970_har	301	24876	1305607	0.38	0.38	0.34	0.34	-0.19	-0.19
1970_lee	194	5765	125547	0.10	0.10	0.10	0.10	0.00	0.00
1970_lmu	359	10915	158768	0.53	0.53	0.49	0.49	0.15	0.31

1970_mit	448	13060	672376	0.76	0.76	0.73	0.73	0.28	0.28
1970_rfw	237	7608	102156	0.22	0.22	0.18	0.18	0.05	0.11
1970_sta	367	15316	772764	0.55	0.55	0.51	0.51	0.08	0.07
1970_ucb	515	15514	727541	0.90	0.94	0.85	0.90	0.29	0.32
1970_ucd	453	9238	241426	0.78	0.78	0.74	0.74	0.30	0.40
1970_ucs	198	11072	501336	0.11	0.11	0.08	0.08	-0.05	-0.05
1970_uos	293	5610	121761	0.36	0.36	0.32	0.32	0.19	0.20
1970_uow	768	14128	570435	1.35	1.60	1.29	1.51	0.79	0.85
1970_uzh	103	6184	116481	0.00	0.08	-0.13	0.00	-0.19	-0.09
1980_auc	228	4209	120159	0.19	0.19	0.19	0.15	0.09	0.09
1980_cal	181	11018	694919	0.07	0.07	0.03	0.03	-0.09	-0.15
1980_gau	377	8658	180611	0.58	0.58	0.54	0.54	0.20	0.29
1980_har	339	37431	2450483	0.26	0.02	0.21	0.00	-0.32	-0.36
1980_lmu	712	16024	338964	1.26	1.45	1.20	1.36	0.66	0.80
1980_mit	491	22506	1204153	0.84	0.88	0.72	0.79	0.15	0.15
1980_rfw	535	8698	154175	0.99	0.99	0.95	0.95	0.49	0.54
1980_sta	405	25900	1534993	0.61	0.65	0.56	0.56	-0.05	-0.08
1980_ucb	622	24015	1270877	0.89	1.02	0.83	1.00	0.41	0.41
1980_ucl	791	32039	1349629	1.08	1.08	0.96	1.04	0.34	0.78
1980_ucs	285	19119	1060414	0.34	0.34	0.30	0.30	-0.13	-0.14
1980_uos	353	9865	268186	0.52	0.52	0.48	0.48	0.14	0.21
1980_uow	1062	24626	1211541	1.79	1.79	1.61	1.67	0.77	1.05
1980_uzh	154	9220	291477	0.00	0.00	0.00	0.00	-0.11	-0.10
1990_auc	231	7182	314576	0.20	0.20	0.16	0.16	0.04	0.04
1990_cal	195	20198	1369484	0.11	0.11	0.07	0.07	-0.28	-0.37
1990_gau	441	13438	499917	0.75	0.75	0.71	0.71	0.26	0.27
1990_har	375	66556	6688846	0.00	0.00	-0.17	-1.23	-1.11	-2.97
1990_lmu	553	25880	1011578	0.80	1.04	0.74	0.95	0.23	0.31
1990_mit	468	32409	2258109	0.50	0.24	0.44	0.20	0.01	-0.02
1990_rfw	680	13705	444648	1.33	1.37	1.20	1.28	0.63	0.71
1990_sta	434	39340	3208990	0.17	0.15	0.15	0.04	-0.21	-0.65
1990_ucb	693	36029	2535642	0.83	0.83	0.71	0.71	0.09	-0.06
1990_ucd	927	28903	1479455	1.43	1.43	1.28	1.36	0.54	0.82
1990_ucl	537	46153	2826218	0.42	0.42	0.30	0.30	-0.32	-0.32

1990_ucs	468	32897	2584155	0.47	0.24	0.42	0.20	0.01	-0.18
1990_uos	414	18377	762773	0.68	0.68	0.64	0.64	0.10	0.19
1990_uow	392	42264	3067482	0.04	0.04	0.04	0.00	-0.35	-0.54
1990_upp	186	21734	1020180	0.08	0.08	0.04	0.04	-0.30	-0.28
1990_uta	633	20402	959783	1.06	1.25	1.00	1.15	0.46	0.50
1990_uzh	183	17035	846630	0.08	0.08	0.04	0.04	-0.27	-0.27
2000_auc	236	15653	789471	0.21	0.21	0.17	0.17	-0.17	-0.13
2000_cal	154	33846	2469619	0.00	0.00	-0.04	-0.09	-0.58	-0.75
2000_eth	189	37981	2439385	0.05	0.03	0.00	0.00	-0.61	-0.66
2000_gau	355	21487	1083274	0.52	0.52	0.48	0.48	-0.05	-0.02
2000_lee	216	23646	1242604	0.16	0.16	0.12	0.12	-0.31	-0.32
2000_lmu	645	40470	2125263	0.70	0.70	0.58	0.66	-0.05	0.27
2000_mit	463	52213	4078946	0.23	0.23	0.11	0.11	-0.60	-1.04
2000_rfw	720	24123	1153615	0.90	1.12	0.88	1.10	0.50	0.66
2000_sta	477	66201	5250278	0.09	0.27	0.02	0.00	-0.93	-1.74
2000_ucb	766	52128	4216520	0.84	1.02	0.67	0.75	-0.25	-0.86
2000_ucd	801	49320	2863607	0.93	1.11	0.80	0.99	-0.15	-0.03
2000_ucl	862	73770	5252023	0.53	1.23	0.38	0.00	-1.01	-1.74
2000_ucs	420	56368	4270251	0.12	0.12	0.00	0.00	-0.76	-1.15
2000_uos	218	41855	2005616	0.06	0.17	0.00	0.08	-0.57	-0.54
2000_uow	794	69504	5219396	0.58	1.09	0.35	0.03	-0.81	-1.72
2000_upp	336	30239	1706523	0.43	0.47	0.36	0.38	-0.25	-0.25
2000_uta	702	33457	1871792	0.85	0.85	0.73	0.81	0.19	0.43
2000_uzh	187	29876	1869815	0.09	0.09	0.05	0.03	-0.45	-0.54
2010_auc	436	34809	1065309	0.39	0.73	0.33	0.69	-0.07	0.11
2010_cal	169	43085	1906296	0.00	0.04	-0.13	0.00	-0.69	-0.58
2010_mit	442	81850	4147014	0.00	0.17	-0.57	0.06	-1.38	-1.09
2010_sta	660	111186	5242204	0.00	0.74	-0.67	0.01	-2.75	-1.74
2010_ucb	830	70570	3380966	0.53	1.18	0.44	0.92	-0.86	-0.23
2010_ucl	918	102713	4107384	0.39	1.41	0.00	0.98	-2.36	-0.77
2010_uow	2575	107990	4492350	0.15	1.87	0.00	0.65	-2.60	-0.95
2010_uta	713	54691	1894394	0.70	0.88	0.55	0.83	-0.39	0.46

S10. Full frontier values of inputs, outputs and efficiency estimates

i	X	Yp	Yc	FDHp	FDHc	Oa99p	Oa99c	Oa90p	Oa90c
1890_col	38	173	405	0.02	0.01	0.01	0.00	-0.01	0.00
1890_har	66	361	570	0.04	0.04	0.04	0.02	-0.01	0.00
1890_lmu	52	78	62	0.05	0.02	0.03	0.01	0.00	0.00
1890_mit	30	36	1425	0.02	0.01	0.01	0.00	0.00	0.00
1890_rfw	32	201	334	0.01	0.01	0.01	0.00	-0.01	0.00
1890_ucb	25	7	9	0.02	0.01	0.01	0.00	-0.02	-0.02
1890_uos	21	12	9	0.01	0.01	0.00	0.00	-0.04	-0.04
1890_uta	8	3	2	0.00	0.00	0.00	0.00	-0.08	-0.08
1890_uzh	40	64	454	0.03	0.01	0.02	0.01	0.00	0.00
1900_col	88	517	1962	0.12	0.07	0.09	0.07	-0.01	0.00
1900_eth	43	6	4	0.03	0.02	0.02	0.01	0.01	0.00
1900_gau	40	472	1386	0.01	0.01	0.00	0.01	-0.02	0.00
1900_har	124	587	1887	0.17	0.19	0.14	0.08	0.00	0.00
1900_lee	32	65	207	0.02	0.01	0.01	0.00	0.00	0.00
1900_lmu	61	133	143	0.05	0.03	0.03	0.02	0.00	0.00
1900_mit	52	125	565	0.05	0.02	0.02	0.01	0.00	0.00
1900_rfw	31	343	465	0.00	0.01	0.00	0.00	-0.02	0.00
1900_sta	36	143	407	0.01	0.01	0.01	0.00	-0.01	0.00
1900_ucb	68	43	54	0.06	0.05	0.05	0.02	0.01	0.00
1900_uos	44	20	3	0.03	0.02	0.03	0.01	0.01	0.00
1900_uow	14	18	11	0.00	0.00	0.00	0.00	-0.06	-0.06
1900_upp	22	24	32	0.01	0.01	0.00	0.00	-0.03	-0.03
1900_uta	19	3	12	0.00	0.00	0.00	0.00	-0.04	-0.04
1900_uzh	26	79	206	0.02	0.01	0.01	0.00	-0.02	-0.02
1910_col	137	862	3675	0.15	0.24	0.12	0.12	0.00	0.00
1910_eth	42	137	2427	0.02	0.02	0.02	0.01	0.00	0.00
1910_gau	42	243	845	0.02	0.02	0.02	0.01	-0.01	0.00
1910_har	116	753	3264	0.16	0.16	0.13	0.07	-0.02	0.00
1910_lee	32	133	482	0.01	0.01	0.01	0.00	-0.01	0.00
1910_lmu	80	196	436	0.09	0.08	0.07	0.06	0.00	0.00
1910_mit	58	148	503	0.05	0.02	0.02	0.01	0.00	0.00
1910_rfw	36	218	267	0.01	0.01	0.01	0.00	-0.01	0.00

1910_sta	59	315	684	0.04	0.02	0.02	0.02	-0.01	0.00
1910_ucb	99	191	2242	0.15	0.10	0.13	0.07	0.01	0.00
1910_uos	57	23	86	0.06	0.02	0.03	0.01	0.01	0.00
1910_uow	34	97	150	0.01	0.01	0.01	0.00	-0.01	0.00
1910_upp	28	36	184	0.02	0.01	0.01	0.00	-0.01	-0.01
1910_uzh	36	180	829	0.01	0.01	0.01	0.00	-0.01	0.00
1920_auc	3	5	15	0.00	0.00	-0.02	-0.02	-0.10	-0.10
1920_cal	19	308	7441	0.00	0.00	-0.01	-0.01	-0.04	-0.04
1920_col	247	1258	5135	0.34	0.34	0.34	0.34	0.15	0.14
1920_eth	43	183	2498	0.02	0.02	0.02	0.01	0.00	0.00
1920_gau	60	803	5154	0.02	0.02	0.02	0.02	-0.04	-0.01
1920_har	163	1319	13193	0.22	0.28	0.12	0.13	-0.01	0.00
1920_lee	30	298	1746	0.00	0.01	0.00	0.00	-0.02	0.00
1920_lmu	149	648	1827	0.17	0.28	0.16	0.15	0.01	0.01
1920_mit	113	533	4699	0.15	0.15	0.14	0.07	0.00	0.00
1920_rfw	84	624	888	0.11	0.07	0.07	0.07	-0.02	0.00
1920_sta	56	719	2703	0.01	0.02	0.01	0.01	-0.04	0.00
1920_ucb	105	528	4807	0.14	0.12	0.13	0.07	0.00	0.00
1920_uos	19	81	184	0.00	0.00	0.00	0.00	-0.04	-0.04
1920_uow	21	161	370	0.01	0.01	0.00	0.00	-0.04	-0.04
1920_upp	31	98	355	0.01	0.01	0.00	0.00	-0.01	0.00
1920_uta	24	46	213	0.02	0.01	0.01	0.00	-0.03	-0.03
1920_uzh	33	465	4104	0.00	0.00	-0.01	0.00	-0.03	-0.01
1950_auc	22	160	2042	0.01	0.01	0.00	0.00	-0.03	-0.03
1950_cal	134	1857	60129	0.11	0.22	0.08	0.07	-0.07	-0.07
1950_col	753	4337	94345	2.12	2.17	2.05	2.12	1.15	1.18
1950_eth	71	1223	18406	0.06	0.05	0.03	0.03	-0.06	-0.02
1950_gau	106	2194	13789	0.06	0.13	0.03	0.06	-0.10	-0.01
1950_har	394	5255	162627	0.87	0.87	0.82	0.82	0.25	0.25
1950_lee	64	1326	23901	0.04	0.04	0.00	0.00	-0.07	-0.03
1950_lmu	137	2281	15944	0.12	0.24	0.05	0.12	-0.08	-0.01
1950_mit	310	3466	132883	0.57	0.57	0.51	0.51	0.26	0.18
1950_rfw	117	1782	8598	0.08	0.17	0.06	0.06	-0.07	-0.01
1950_sta	409	1883	49175	0.92	0.92	0.87	0.87	0.39	0.34

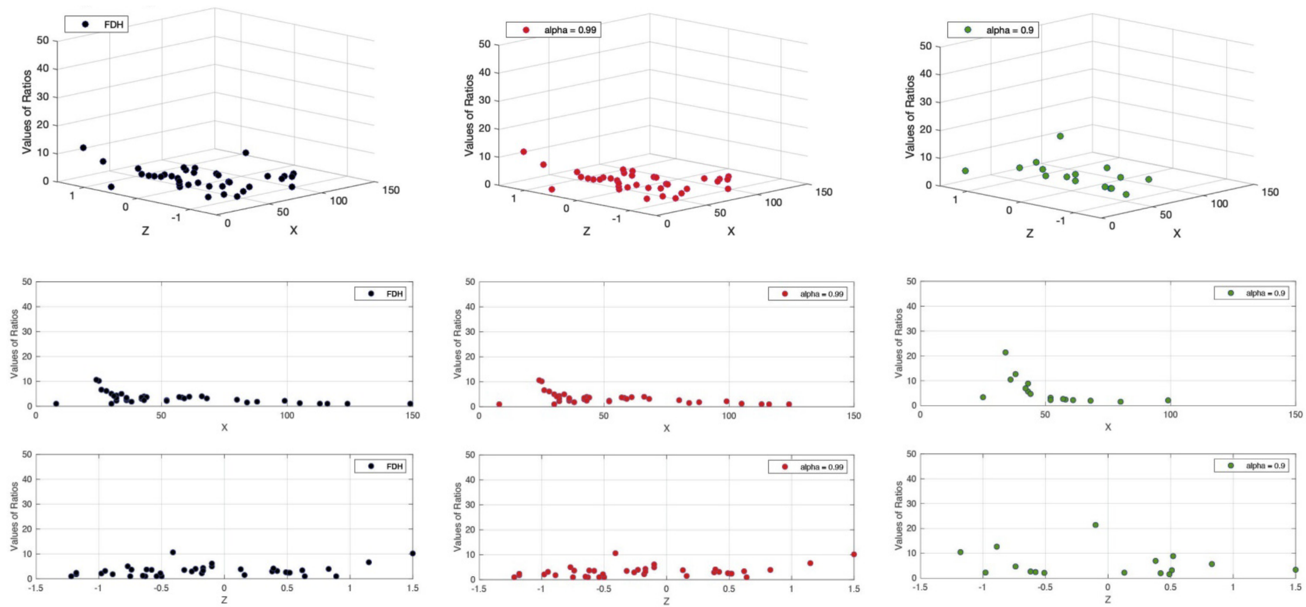
1950_ucb	455	4675	123444	1.09	1.09	1.04	1.04	0.42	0.42
1950_ucd	186	719	10138	0.30	0.28	0.16	0.14	0.04	0.01
1950_ucl	239	2445	48473	0.31	0.31	0.31	0.31	0.10	0.10
1950_uos	245	1119	12324	0.35	0.33	0.33	0.33	0.16	0.13
1950_uow	389	1568	26630	0.85	0.85	0.80	0.80	0.39	0.34
1950_upp	37	925	19014	0.00	0.00	-0.03	-0.01	-0.05	-0.02
1950_uta	174	889	13345	0.26	0.28	0.15	0.13	0.02	0.00
1950_uzh	54	1153	8026	0.00	0.01	0.00	0.00	-0.06	-0.01
1960_auc	95	633	10010	0.12	0.09	0.11	0.06	-0.02	-0.01
1960_cal	195	4091	157707	0.15	0.15	0.15	0.15	-0.05	-0.05
1960_col	194	7068	245747	0.15	0.15	0.15	0.09	-0.09	-0.09
1960_eth	125	1991	42784	0.08	0.20	0.07	0.08	-0.08	-0.05
1960_gau	190	3402	48516	0.19	0.24	0.13	0.13	-0.02	0.00
1960_har	189	11481	518526	0.13	0.13	0.07	0.07	-0.30	-0.31
1960_lee	64	2594	53373	0.00	0.01	-0.01	0.00	-0.15	-0.07
1960_lmu	277	4545	68014	0.45	0.45	0.39	0.39	0.17	0.21
1960_mit	553	8442	361970	1.45	1.45	1.39	1.39	0.54	0.55
1960_rfw	195	3117	31599	0.21	0.26	0.15	0.15	0.00	0.02
1960_sta	212	7210	276782	0.21	0.21	0.16	0.16	-0.07	-0.06
1960_ucb	326	10423	414889	0.62	0.62	0.57	0.57	-0.04	-0.04
1960_ucd	204	2832	78266	0.23	0.20	0.18	0.18	0.02	0.02
1960_ucl	316	7963	224971	0.59	0.59	0.53	0.53	0.00	0.08
1960_ucs	71	3008	232926	0.00	0.00	-0.03	-0.15	-0.17	-0.29
1960_uos	291	2842	48352	0.50	0.50	0.44	0.44	0.23	0.24
1960_uow	257	5245	153986	0.37	0.37	0.32	0.32	0.09	0.10
1960_upp	54	2424	59360	0.00	0.00	-0.04	-0.04	-0.15	-0.07
1960_uta	124	2917	77479	0.08	0.19	0.01	0.05	-0.12	-0.09
1960_uzh	81	1664	22363	0.06	0.05	0.05	0.04	-0.09	-0.03
1970_auc	139	2182	44649	0.13	0.24	0.06	0.09	-0.07	-0.05
1970_cal	173	6931	370463	0.07	0.07	0.07	0.01	-0.15	-0.18
1970_gau	213	5671	79166	0.21	0.21	0.21	0.21	0.00	0.05
1970_har	301	24876	1305607	0.53	0.53	0.48	0.48	-0.43	-0.43
1970_lee	194	5765	125547	0.15	0.15	0.15	0.15	-0.07	-0.01
1970_lmu	359	10915	158768	0.74	0.74	0.69	0.69	0.01	0.20

1970_mit	448	13060	672376	1.07	1.07	1.01	1.01	0.18	0.16
1970_rfw	237	7608	102156	0.30	0.30	0.25	0.25	-0.03	0.07
1970_sta	367	15316	772764	0.77	0.77	0.72	0.72	-0.03	-0.03
1970_ucb	515	15514	727541	1.27	1.31	1.25	1.25	0.29	0.34
1970_ucd	453	9238	241426	1.08	1.08	1.03	1.03	0.28	0.35
1970_ucs	198	11072	501336	0.16	0.16	0.11	0.11	-0.27	-0.29
1970_uos	293	5610	121761	0.50	0.50	0.45	0.45	0.11	0.20
1970_uow	768	14128	570435	2.01	2.23	1.92	2.10	1.06	1.06
1970_uzh	103	6184	116481	0.00	0.12	-0.18	0.00	-0.30	-0.14
1980_auc	228	4209	120159	0.27	0.27	0.27	0.22	0.05	0.04
1980_cal	181	11018	694919	0.10	0.10	0.04	0.04	-0.27	-0.49
1980_gau	377	8658	180611	0.81	0.81	0.75	0.75	0.14	0.18
1980_har	339	37431	2450483	0.39	0.02	0.31	0.00	-0.52	-0.95
1980_lmu	712	16024	338964	1.87	2.02	1.79	1.97	0.85	0.95
1980_mit	491	22506	1204153	1.17	1.22	1.07	1.10	0.08	0.08
1980_rfw	535	8698	154175	1.38	1.38	1.33	1.33	0.48	0.61
1980_sta	405	25900	1534993	0.86	0.91	0.78	0.78	-0.21	-0.23
1980_ucb	622	24015	1270877	1.32	1.53	1.24	1.49	0.41	0.34
1980_ucl	791	32039	1349629	1.51	1.51	1.35	1.45	0.19	0.53
1980_ucs	285	19119	1060414	0.48	0.48	0.42	0.42	-0.33	-0.37
1980_uos	353	9865	268186	0.72	0.72	0.67	0.67	0.05	0.10
1980_uow	1062	24626	1211541	2.49	2.49	2.25	2.37	0.92	1.07
1980_uzh	154	9220	291477	0.00	0.00	0.00	0.00	-0.23	-0.21
1990_auc	231	7182	314576	0.28	0.28	0.22	0.22	-0.02	-0.05
1990_cal	195	20198	1369484	0.15	0.15	0.09	0.09	-0.62	-0.74
1990_gau	441	13438	499917	1.04	1.04	0.99	0.99	0.15	0.25
1990_har	375	66556	6688846	0.00	0.00	-0.24	-1.83	-1.81	-5.30
1990_lmu	553	25880	1011578	1.19	1.45	1.11	1.32	0.28	0.31
1990_mit	468	32409	2258109	0.74	0.34	0.65	0.28	-0.18	-0.64
1990_rfw	680	13705	444648	1.85	1.91	1.68	1.85	0.77	0.81
1990_sta	434	39340	3208990	0.26	0.21	0.21	0.05	-0.63	-1.31
1990_ucb	693	36029	2535642	1.15	1.15	0.99	0.99	-0.08	-0.37
1990_ucd	927	28903	1479455	2.00	2.00	1.84	1.94	0.49	0.67
1990_ucl	537	46153	2826218	0.59	0.59	0.42	0.42	-0.79	-0.89

1990_ucs	468	32897	2584155	0.71	0.34	0.62	0.28	-0.21	-0.82
1990_uos	414	18377	762773	0.94	0.94	0.89	0.89	0.00	0.03
1990_uow	392	42264	3067482	0.06	0.06	0.06	0.00	-0.83	-1.35
1990_upp	186	21734	1020180	0.12	0.12	0.06	0.06	-0.69	-0.64
1990_uta	633	20402	959783	1.57	1.74	1.49	1.61	0.57	0.57
1990_uzh	183	17035	846630	0.11	0.11	0.05	0.05	-0.54	-0.55
2000_auc	236	15653	789471	0.30	0.30	0.24	0.24	-0.36	-0.37
2000_cal	154	33846	2469619	0.00	0.00	-0.05	-0.13	-1.02	-1.19
2000_eth	189	37981	2439385	0.07	0.04	0.00	0.00	-0.99	-1.15
2000_gau	355	21487	1083274	0.73	0.73	0.67	0.67	-0.22	-0.24
2000_lee	216	23646	1242604	0.22	0.22	0.17	0.17	-0.58	-0.60
2000_lmu	645	40470	2125263	0.98	0.98	0.82	0.92	-0.39	-0.17
2000_mit	463	52213	4078946	0.32	0.32	0.16	0.16	-1.12	-1.97
2000_rfw	720	24123	1153615	1.31	1.68	1.25	1.64	0.42	0.66
2000_sta	477	66201	5250278	0.13	0.37	0.02	0.00	-1.69	-3.46
2000_ucb	766	52128	4216520	1.18	1.42	1.00	1.05	-0.71	-2.14
2000_ucd	801	49320	2863607	1.30	1.55	1.19	1.38	-0.52	-0.42
2000_ucl	862	73770	5252023	0.73	1.77	0.56	0.00	-2.21	-3.47
2000_ucs	420	56368	4270251	0.16	0.16	0.00	0.00	-1.36	-2.21
2000_uos	218	41855	2005616	0.09	0.23	0.00	0.11	-0.94	-0.91
2000_uow	794	69504	5219396	0.85	1.52	0.49	0.04	-1.91	-3.42
2000_upp	336	30239	1706523	0.61	0.66	0.53	0.53	-0.46	-0.48
2000_uta	702	33457	1871792	1.19	1.19	1.02	1.12	0.03	0.04
2000_uzh	187	29876	1869815	0.12	0.12	0.07	0.05	-0.82	-1.00
2010_auc	436	34809	1065309	0.57	1.02	0.49	0.97	-0.34	0.00
2010_cal	169	43085	1906296	0.00	0.05	-0.18	0.00	-1.12	-1.05
2010_mit	442	81850	4147014	0.00	0.24	-0.79	0.08	-2.77	-2.06
2010_sta	660	111186	5242204	0.00	1.03	-0.94	0.01	-4.80	-3.45
2010_ucb	830	70570	3380966	0.78	1.65	0.62	1.28	-1.99	-1.08
2010_ucl	918	102713	4107384	0.59	1.97	0.00	1.46	-4.21	-2.01
2010_uow	2575	107990	4492350	0.22	2.80	0.00	0.97	-4.55	-2.43
2010_uta	713	54691	1894394	0.98	1.23	0.82	1.16	-0.89	0.07

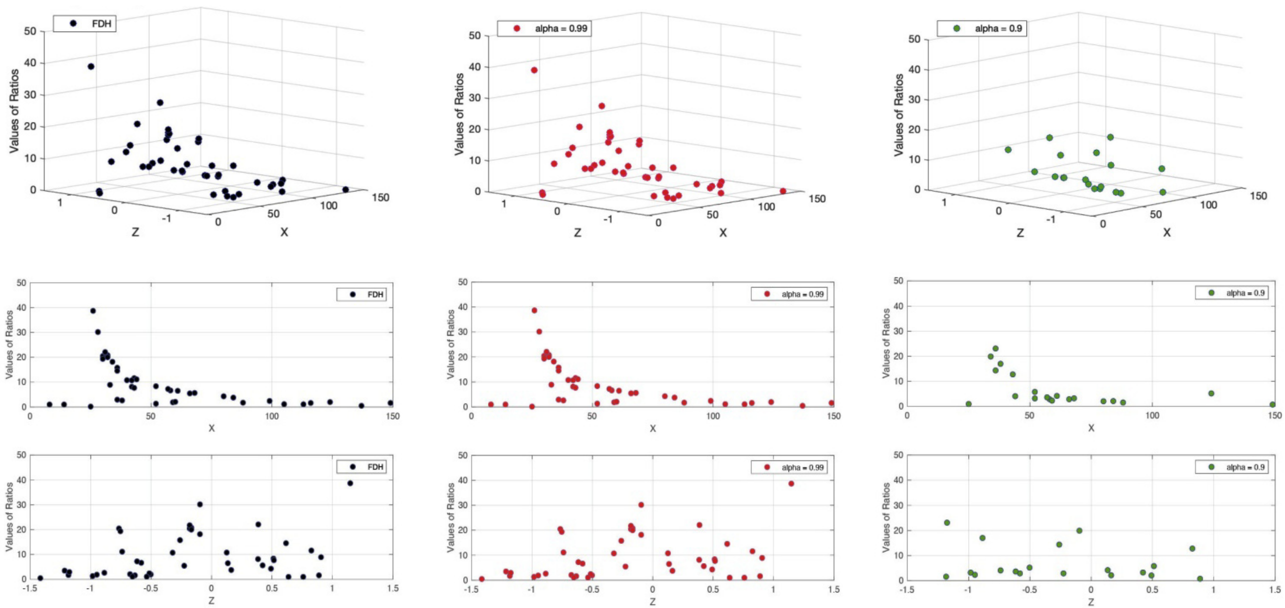
S11. Ratios of (task coordination) conditional to unconditional model for full, robust and partial PP efficiency estimates (pre-war period)

β_α (X,Y|Z) / β_α (X,Y) vs. X vs. Z (task coordination)



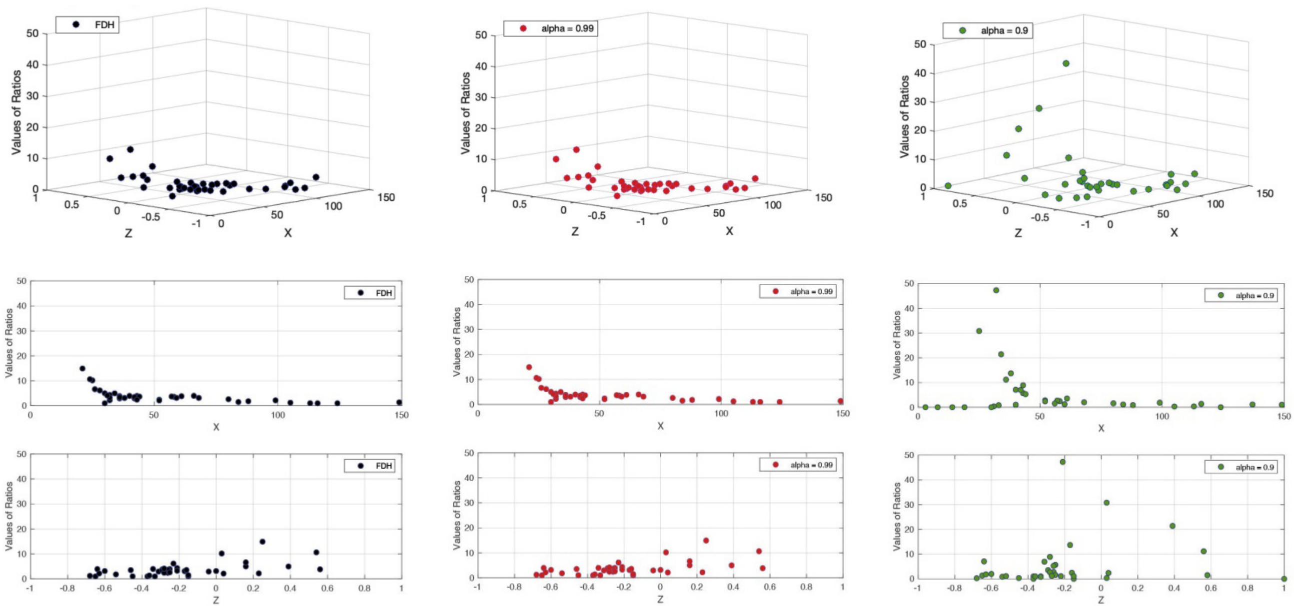
S12. Ratios of (task coordination) conditional to unconditional model for full, robust and partial CP efficiency estimates (pre-war period)

$\beta_{\alpha}(X,Y|Z) / \beta_{\alpha}(X,Y)$ vs. X vs. Z (task coordination)



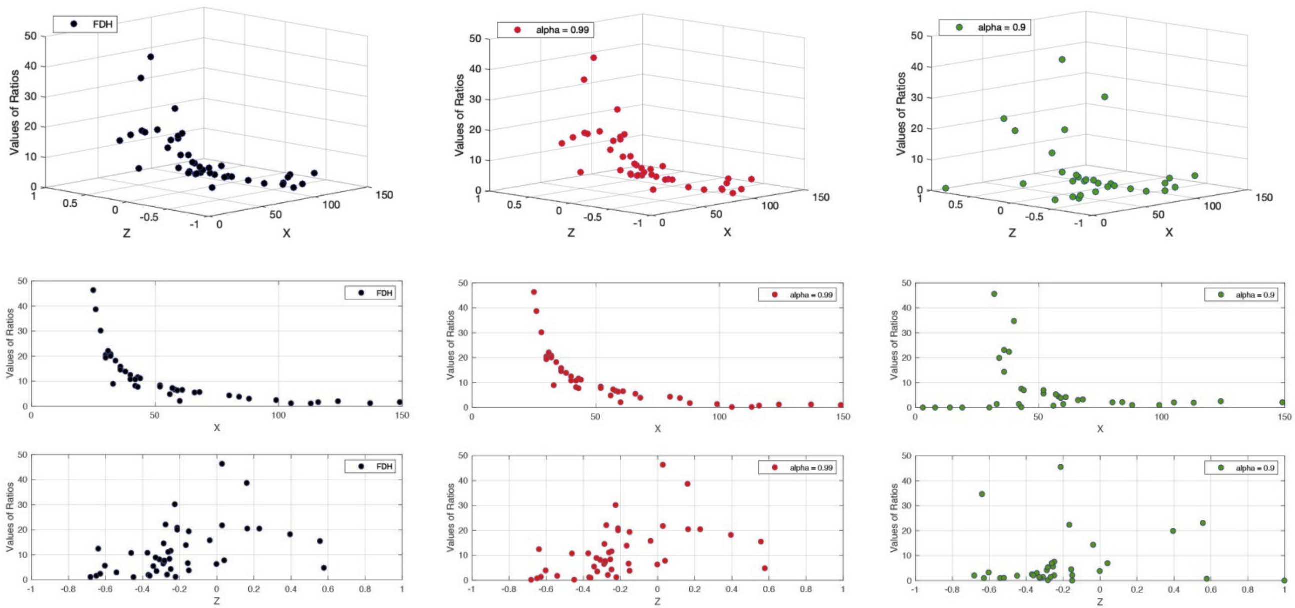
S13. Ratios of (specialization concentration) conditional to unconditional model for full, robust and partial PP efficiency estimates (pre-war period)

β_{α}(X,Y|Z) / β_{α}(X,Y) vs. X vs. Z (specialization concentration)



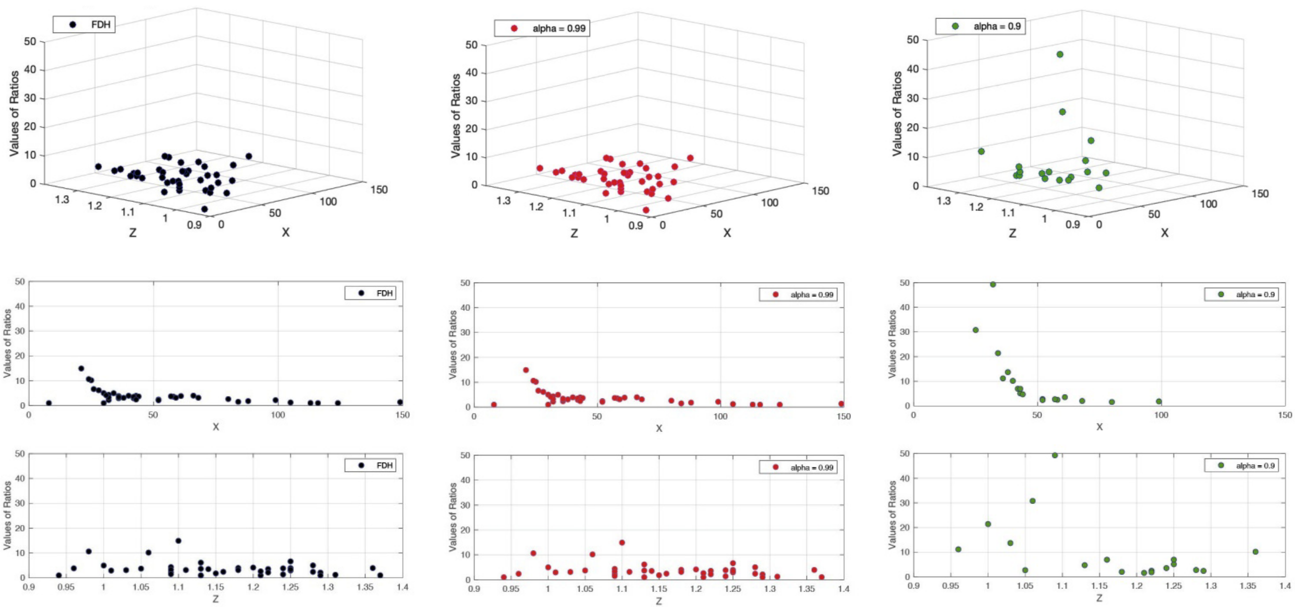
S14. Ratios of (specialization concentration) conditional to unconditional model for full, robust and partial CP efficiency estimates (pre-war period)

β_α (X,Y|Z) / β_α (X,Y) vs. X vs. Z (specialization concentration)



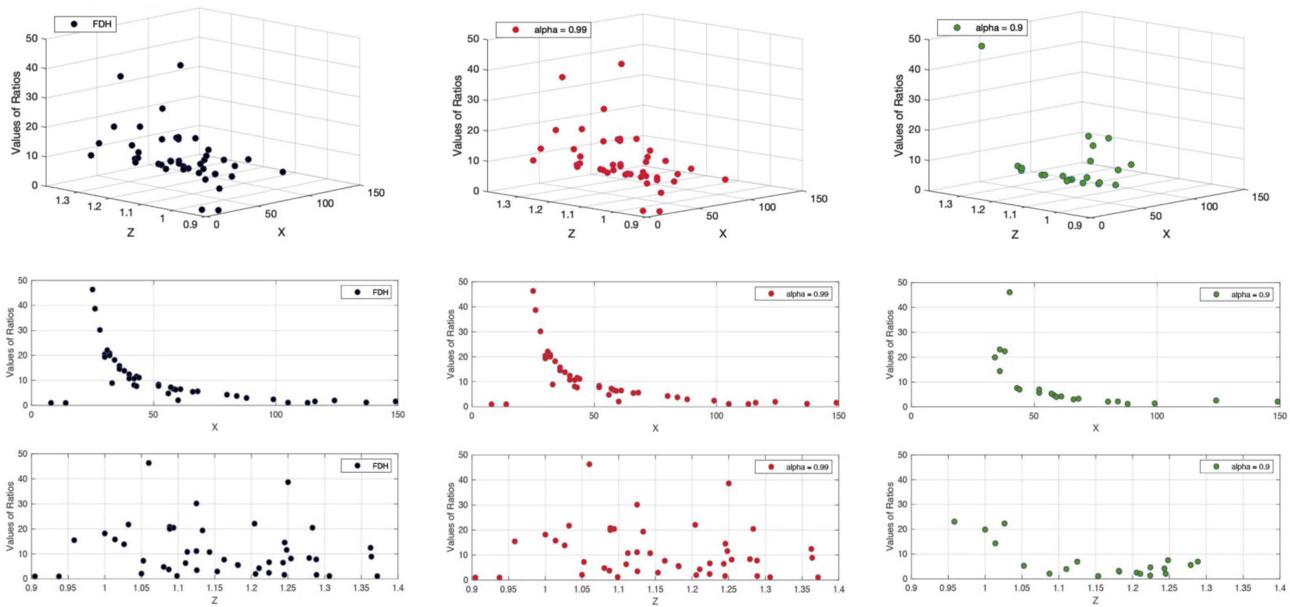
S15. Ratios of (specialization gravity) conditional to unconditional model for full, robust and partial PP efficiency estimates (pre-war period)

$\beta_\alpha(X,Y|Z) / \beta_\alpha(X,Y)$ vs. X vs. Z (specialization gravity)



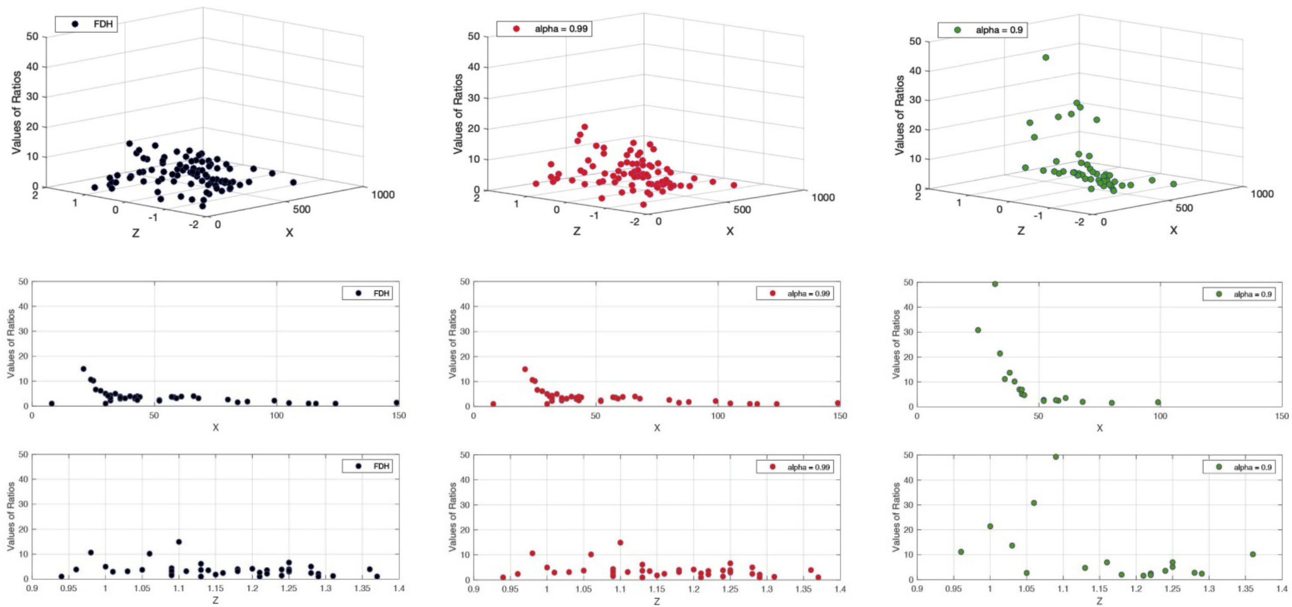
S16. Ratios of (specialization gravity) conditional to unconditional model for full, robust and partial CP efficiency estimates (pre-war period)

β_α(X,Y|Z) / β_α(X,Y) vs. X vs. Z (specialization gravity)



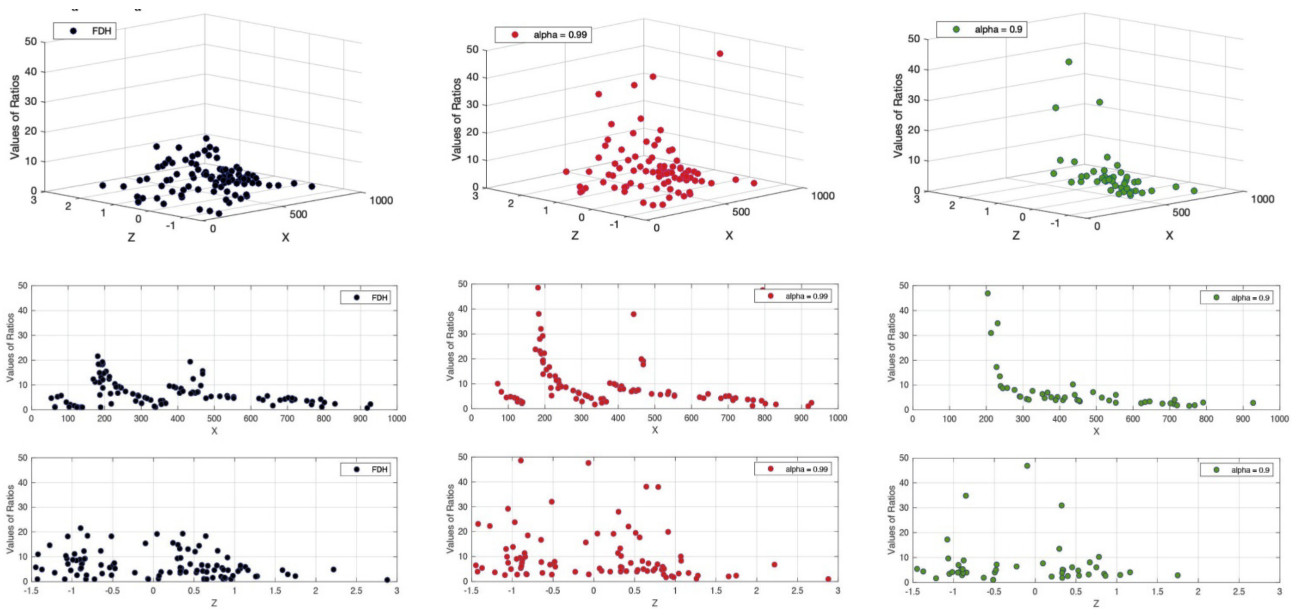
S17. Ratios of (task coordination) conditional to unconditional model for full, robust and partial PP efficiency estimates (post-war period)

$\beta_{\alpha}(X,Y|Z) / \beta_{\alpha}(X,Y)$ vs. X vs. Z (task coordination)



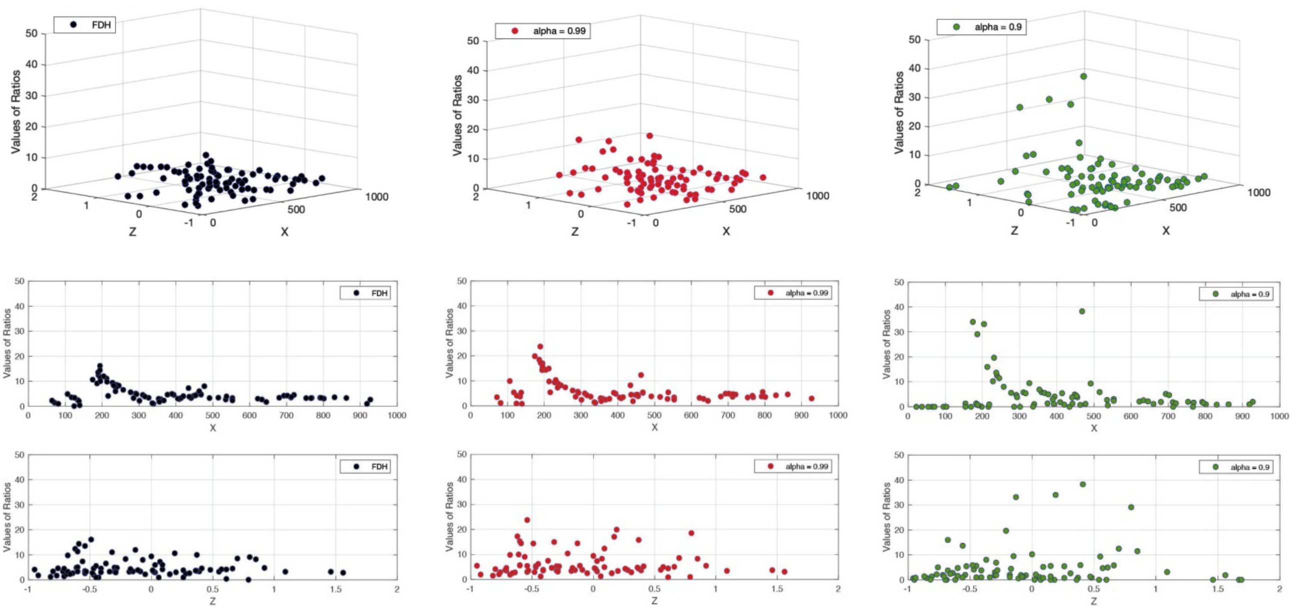
S18. Ratios of (task coordination) conditional to unconditional model for full, robust and partial CP efficiency estimates (post-war period)

$\beta_\alpha(X,Y|Z) / \beta_\alpha(X,Y)$ vs. X vs. Z (task coordination)



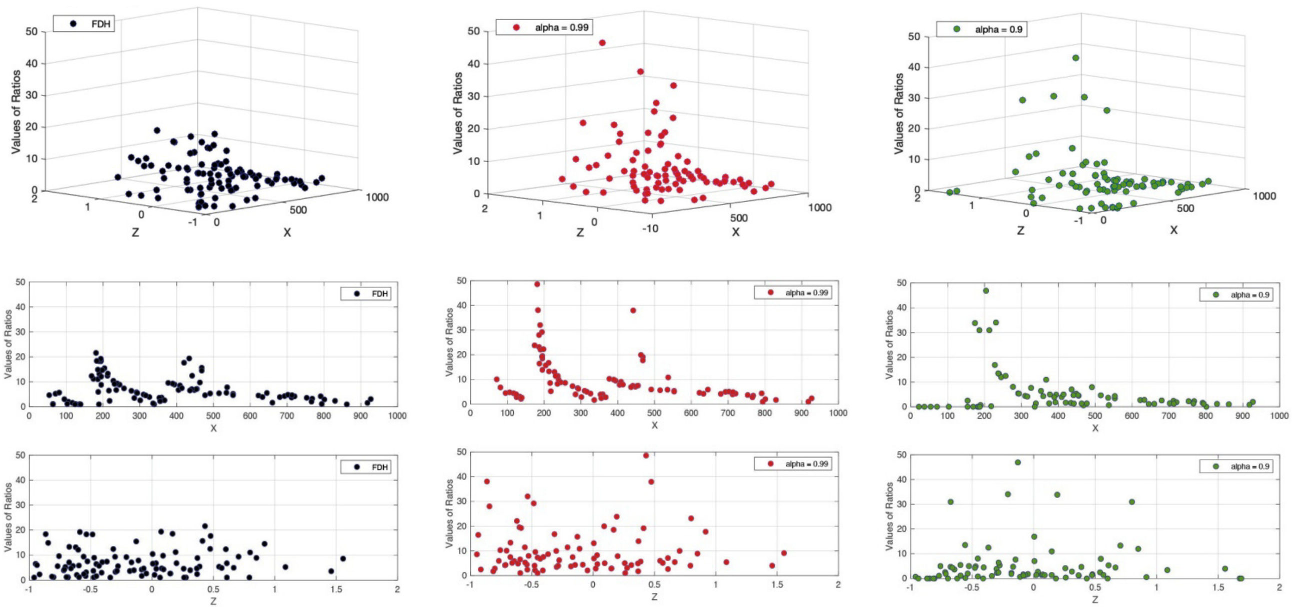
S19. Ratios of (specialization concentration) conditional to unconditional model for full, robust and partial PP efficiency estimates (post-war period)

β_α (X,Y|Z) / β_α (X,Y) vs. X vs. Z (specialization concentration)



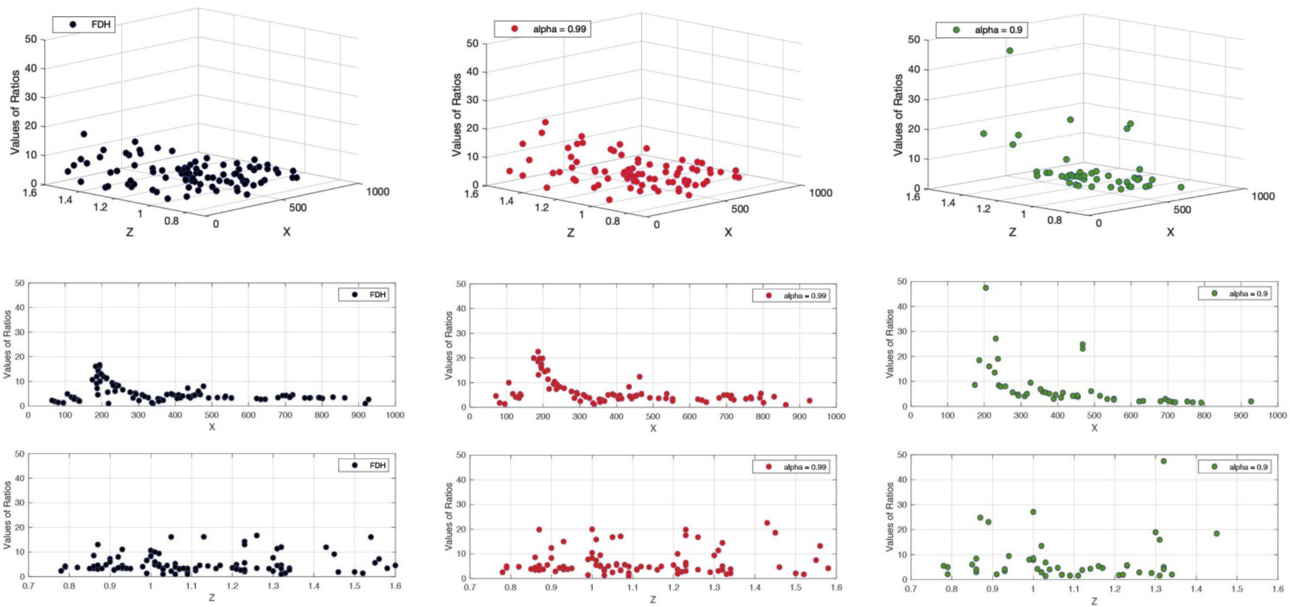
S20. Ratios of (specialization concentration) conditional to unconditional model for full, robust and partial CP efficiency estimates (post-war period)

β_α (X,Y|Z) / β_α (X,Y) vs. X vs. Z (specialization concentration)



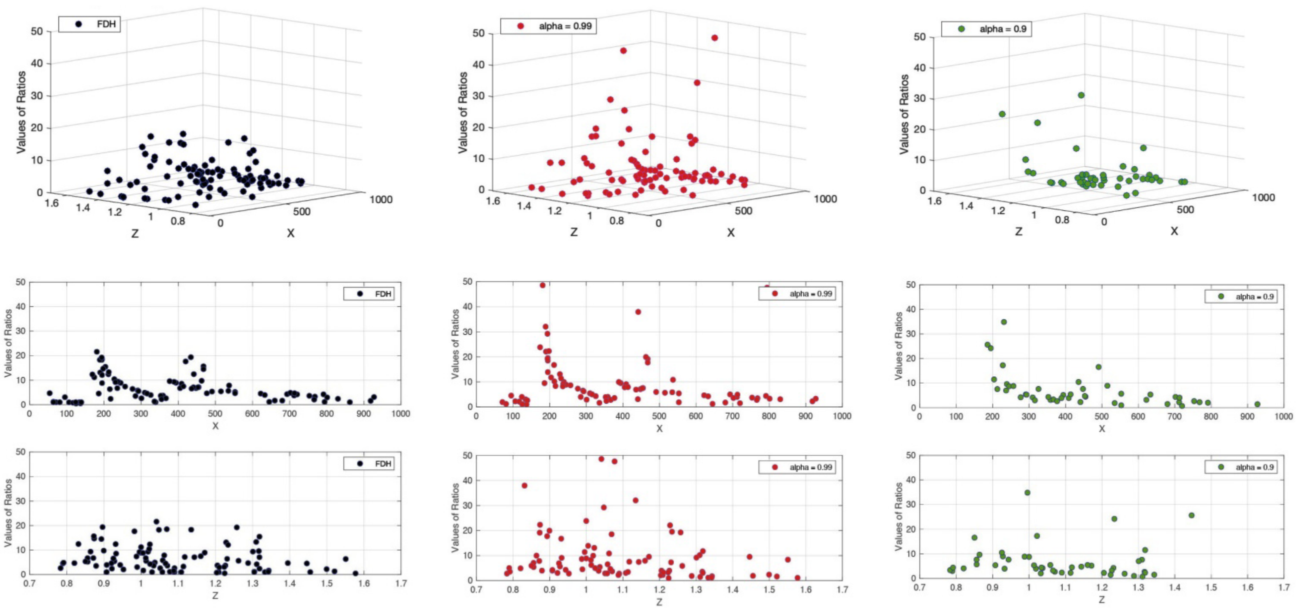
S21. Ratios of (specialization gravity) conditional to unconditional model for full, robust and partial PP efficiency estimates (post-war period)

β_α (X,Y|Z) / β_α (X,Y) vs. X vs. Z (specialization gravity)



S22. Ratios of (specialization gravity) conditional to unconditional model for full, robust and partial CP efficiency estimates (post-war period)

$\beta_\alpha(X,Y|Z) / \beta_\alpha(X,Y)$ vs. X vs. Z (specialization gravity)



Recently, evidence is accumulating that despite the ever-growing scientific community, the innovative power of original research has stagnated. In addition, a social divide between unquestioning belief and irrational skepticism, as evidenced during the Covid pandemic, reinforces the impression of a science in crisis.

A central starting point of this work is the idea that excessive division of labor and specialization could be a fundamental cause of this crisis. Despite the importance that thinkers such as Adam Smith and Emile Durkheim ascribed to the division of labor for explaining scientific progress, modern economic findings on coordination costs have rarely been applied in the context of scientific institutions. If addressed at all, the topics are explored in highly specialized studies on topics such as cognitive diversity or bibliometric specialization. Given the prevalence of economic thought within science studies, this is surprising and may itself be evidence of a lack of coordination among increasingly isolated branches of science. Thus, a thorough discussion of division of labor and specialization as factors influencing scientific inquiry has been largely missing until now.

This work takes a first step toward addressing this gap by introducing a new dataset that documents the development of professorial denominations at twenty globally highly ranked universities over the last century. The data reveals a continuous increase in specialization, as well as thematic path dependencies, especially in European institutions. The results of a non-parametric conditional efficiency framework indicate a functional relationship between the degree of division of labor and specialization with the productivity of the universities studied. Findings across university clusters point to a beneficial effect of the US department system on productivity when compared to its European professorial chair model counterpart.

Part of the observed crisis could thus be mitigated by monitoring coordination costs to avoid overspecialization and promote organizational forms that enable enhanced collaboration within research institutions. This work further contributes to the literature by proposing an alternative to the vague concept of interdisciplinarity, offering an institutional-based and more granular perspective that allows for monitoring and measurement.

ISBN 978-3-7316-1599-6