

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>

Copyright: Metropolis-Verlag, Marburg 2025

Alle Rechte vorbehalten

ISBN 978-3-7316-1599-6

Contents

A. List of Abbreviations	9
B. List of Tables	11
C. List of Figures	13
1. Introduction	17
2. On the Neglection of Institutional DoL and Spec. in the Science Studies	21
2.1 Literature Review (Science Studies)	21
2.1.1 Epistemic Outcomes	21
2.1.2 Topics related to DoL and Spec.	25
2.2 Why DoL and Spec. are neglected determinants of epistemic outcomes	29
2.2.1 Adam Smith, Emile Durkheim and other thinkers revisited	29
2.2.2 Demarcating institutional DoL and Spec. from the topics reviewed in 2.1.2	34
2.2.3 Potential of institutional DoL and Spec. to explain pathologies in science	39
2.3 Research Question and Outline of the Empirical Analysis	45

3. Introducing a New Dataset to Measure Institutional DoL and Spec. (Dataset Construction and Descriptive Empirical Analysis)	49
3.1 Empirical operationalization of the institutional DoL and Spec. concept	49
3.1.1 The need to limit the empirical analysis to engineering and natural sciences	49
3.1.2 Available perspectives and data sets	53
3.1.2.1 Aggregate institutional level (input-based)	53
3.1.2.2 Bibliometric (output-based)	54
3.1.3 Justifying the need for a new data set on intra-institutional level (input-based)	57
3.1.4 The ‘Denomination Hypothesis’	61
3.2 Sources, Dataset Compilation and Construction of Variables	67
3.2.1 Sources	67
3.2.2 Compilation	73
3.2.3 Denomination harmonization and matching according to WoS scheme	79
3.2.4 Variables	81
3.3 Data	86
3.3.1 Descriptive statistics	86
3.3.1.1 Full sample (zooming out)	86
3.3.1.2 Institution-specific (zooming in)	104
3.3.2 Identifying university types (cluster analysis)	112
3.3.3 Identifying path dependencies of DoL and Spec. (correlation analysis)	123
4. Examining the Effect of Institutional DoL and Spec. on Universities’ Publication Productivity (Quantitative Empirical Analysis)	129
4.1 Methodology: Conditional efficiency framework	129
4.1.1 A brief introduction into non-parametric analysis of production efficiency	129
4.1.2 Efficiency measurement	132
4.1.2.1 FDH model	132
4.1.2.2 Directional distance functions	137

4.1.3	Accounting for extreme observations and external factors	138
4.1.3.1	Partial frontiers	138
4.1.3.2	Conditional frontiers	142
4.1.4	Interpreting the effect of DoL and Spec. on the efficiency measure	145
4.2	Data	149
4.2.1	Efficiency model	149
4.2.1.1	Input data	149
4.2.1.2	(Publication and citation) output data	153
4.2.1.3	Outliers	161
4.2.2	External factors	165
4.2.2.1	DoL – task division and task coord. (factor)	165
4.2.2.2	Spec. Concentration (factor) and gravity	170
4.2.2.3	Bandwidth selection	172
4.2.3	Descriptive statistics	175
4.3	Empirical Results	183
4.3.1	Pre-war frontier	183
4.3.1.1	Efficiency results	183
4.3.1.2	Effect of DoL and Spec. on pre-war efficiency	189
4.3.2	Post-war frontier	199
4.3.2.1	Efficiency results	199
4.3.2.2	Effect of DoL and Spec. on post-war efficiency	203
4.3.3	Full frontier according to university clusters	212
4.3.3.1	Efficiency results	212
4.3.3.2	Effect of DoL and Spec. on university types	215
5.	Discussion of Results	219
6.	Conclusions	225
C.	References	229
D.	Annex	243

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

Tab. 1:	Theory of DoL and Spec. applied to scientific production	38
Tab. 2:	Literature review on available perspectives and data sets – summary	59
Tab. 3:	Denominations documented according to institution and period	78
Tab. 4:	Web of Science categorization scheme limited to epistemic branches	80
Tab. 5:	Operationalization of the components of DoL and Spec. in data set variables	83
Tab. 6:	Nr. of denominations, ... and instantiations in sample according to disciplinarity	96
Tab. 7:	DEN, ... and INST. for 50th percentile of sample professors according to disciplinarity, discipline and research field	100
Tab. 8:	DEN, ... and INST. for 50th percentile of sample professors according to disciplinarity and (harmonized) denomination	102
Tab. 9:	(Unrotated) principal component analysis for task coordination variables	115
Tab. 10:	KMO-criterion, factor loadings, unique variances and predicted scores for the task coordination factor	115
Tab. 11:	(Unrotated) principal component analysis for specialization concentration variables	117
Tab. 12:	KMO-criterion, factor loadings, unique variances and predicted scores the specialization concentration factor	117
Tab. 13:	Correlation analysis results for DoL and Spec. variables for full sample and according to clusters	124
Tab. 14:	Correlation analysis results for disciplinary shares for full sample and according to clusters	125

Tab. 15:	Descriptive statistics of inputs, outputs and productivity ratios of the full sample	162
Tab. 16:	Replication of Tab 5 in section 3.2.4. with variables implemented as external factors printed in bold	167
Tab. 17:	(Unrotated) principal component analysis for the task coordination variables	168
Tab. 18:	KMO-criterion, factor loadings, unique variances and predicted scores for the task coordination factor	168
Tab. 19:	Descriptive statistics for external factors representing DoL in the full sample	169
Tab. 20:	(Unrotated) principal component analysis for the Spec. Conc. variables	171
Tab. 21:	KMO-criterion, factor loadings, unique variances and predicted scores for the Spec. Conc. factor	171
Tab. 22:	Descriptive statistics for external factors representing Spec. in the full sample	172
Tab. 23:	Summary of bandwidths selected according to external factor and model variant	175
Tab. 24:	Pre-war sample descriptive statistics	176
Tab. 25:	Pre-war values for inputs, outputs and external factors according to institutions	178
Tab. 26:	Post-war sample descriptive statistics	179
Tab. 27:	Post-war values for inputs, outputs and external factors according to institutions	181
Tab. 28:	Values for inputs, outputs and external factors according to university clusters	183
Tab. 29:	Pre-war period institutional mean of inputs, outputs and efficiency estimates	184
Tab. 30:	Post-war period institutional mean of inputs, outputs and efficiency estimates	200
Tab. 31:	Mean of inputs, outputs and efficiency estimates according to cluster	212

C. List of Figures

Fig. 1:	1960s-decade exemplary entries of MIT and LMU Munich in: Minerva. Jahrbuch der gelehrten Welt	68
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)’	70
Fig. 3:	Development of sum of professors, professorial types and denominations over time	87
Fig. 4:	Professorial staff, denominations and professors per denominations according to location for pre-war and post- war science	89
Fig. 5:	Development of task coordination and specialization gravity variables for full and sum of professors over time	91
Fig. 6:	Task coordination and specialization gravity according to location for pre-war and post-war science	94
Fig. 7:	Institution-specific absolute number and relative shares of sum of professors and professors per denomination	105
Fig. 8:	Institution-specific mean values for task coordination and specialization gravity	107
Fig. 9:	Institution-specific shares of disciplines (disciplinary) and shares of inter-, multi- and pluridisciplinary denominations ..	108
Fig. 10:	Institution-specific shares of research fields attributable to disciplines (disciplinary)	110
Fig. 11:	Dendrogram of hierarchical cluster analysis according to DoL and Spec. variables	118
Fig. 12:	Dendrogram for DoL and Spec. cluster analysis and box-whisker plots for DoL and Spec. components and factor variables	120
Fig. 13:	Toy Example – Banker, Charnes Cooper (1984) DEA Model	131

Fig. 14:	Toy Example – DEA versus FDH model	136
Fig. 15:	Toy Example – FDH versus Orderalpha model	140
Fig. 16:	Publications and citations per professor according to institution and period	163
Fig. 17:	Cross validation plots for PPM and CPM conditional model with task division as external factor	173
Fig. 18:	Densities for PP and CP, full, robust and partial efficiency estimates (pre-war period)	187
Fig. 19:	Ratios of conditional (task division) to unconditional for full, robust and partial PP efficiency estimates (pre-war period)	190
Fig. 20:	Ratios of conditional (task division) to unconditional for full, robust and partial CP efficiency estimates (pre-war period)	192
Fig. 21:	Plots for marginal and joint effects of (the components of) division of labor on pre-war period publication productivity model	194
Fig. 22:	Plots for marginal and joint effects of (the components of) division of labor on pre-war period citation productivity model	195
Fig. 23:	Plots for marginal and joint effects of (the components of) specialization on pre-war period publication productivity model	196
Fig. 24:	Plots for marginal and joint effects of (the components of) specialization on pre-war period citation productivity model	197
Fig. 25:	Densities for PP and CP, full, robust and partial efficiency estimates (post-war period)	201
Fig. 26:	Ratios of conditional (task division) to unconditional for full, robust and partial PP efficiency estimates (post-war period)	204
Fig. 27:	Ratios of conditional (task division) to unconditional for full, robust and partial CP efficiency estimates (post-war period)	205
Fig. 28:	Plots for marginal and joint effects of (the components of) division of labor on post-war period publication productivity model	207

Fig. 29:	Plots for marginal and joint effects of (the components of) division of labor on post-war period citation productivity model	208
Fig. 30:	Plots for marginal and joint effects of (the components of) specialization on post-war period publication productivity model	209
Fig. 31:	Plots for marginal and joint effects of (the Components of) specialization on post-war period citation productivity model	210
Fig. 32:	Densities of cluster 1, cluster 2 and cluster 3 for full, robust and partial PP model efficiency estimates	213
Fig. 33:	Densities of cluster 1, cluster 2 and cluster 3 for full, robust and partial CP model efficiency estimates	214
Fig. 34:	Marginal effects of DoL and Spec. on PP efficiency estimates for cluster 1, cluster 2 and cluster 3	216
Fig. 35:	Marginal effects of DoL and Spec. on CP efficiency estimates for cluster 1, cluster 2 and cluster 3	217

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.