

Neue Klassifikation von Fächergruppen zur verzerrungsfreien Berechnung universitärer Profilkarten.

Joel E. Fuchs, Thomas Heinze
Bergische Universität Wuppertal

Zusammenfassung

Der vorliegende Aufsatz stellt eine Klassifikation für Lehr- und Forschungsbereiche an Universitäten vor, die eine verzerrungsfreie Berechnung von Fächerprofilen in Form von Heat Maps erlaubt. Die neue Klassifikation beruht auf dem Activity Index (AI), der durch Transformation in einen Index mit symmetrischem Wertebereich überführt werden kann (RESP). Am Beispiel der staatlichen Universitäten in Deutschland zeigen wir, wie durch die neue Klassifikation eine verzerrungsfreie Berechnung von Fächerprofilen möglich ist.

Abstract

This paper presents a classification for teaching and research areas at universities that allows an unbiased calculation of disciplinary profiles using heat maps. This new classification is based upon the Activity Index (AI) which can be transformed into an index with symmetric value range (RESP). Using public universities in Germany as an example, we show how the new classification enables a bias-free calculation of disciplinary profiles.

Zitiervorschlag:

Fuchs, J.E., Heinze, T. (2022): Neue Klassifikation von Fächergruppen zur verzerrungsfreien Berechnung universitärer Profilkarten. Wuppertal: BUW. (DOI: 10.25926/992r-t504).

1. Einleitung

Der vorliegende Aufsatz entwickelt eine Klassifikation für Lehr- und Forschungsbereiche an Universitäten, die eine verzerrungsfreie Berechnung und Darstellung universitärer Fächerprofile erlaubt. Ausgangspunkt ist der in der Literatur bekannte *Activity Index* (AI, Narin, Carpenter and Woolf 1984), der auch *Revealed Technological Advantage* (Soete and Wyatt 1983) oder *Revealed Comparative Advantage* (Balassa 1965) genannt wird. Der AI kann durch entsprechende Transformation in einen Index mit symmetrischem Wertebereich $[-100, +100]$ überführt werden. Als solcher wird er dann *Scientific Specialization Index/SSI* (Costa 2014) oder *Relativer Spezialisierungsindex/ RESP* (Heinze et al. 2019) genannt. Auch die Transformation in einen Index mit Wertebereich $[-1, +1]$ ist in der Literatur zu finden, mit ähnlicher Bezeichnung: *Relative Specialization Index/RSI* (Piro et al. 2017).

Der AI und die von ihm abgeleiteten Spezialisierungsmaße können nur verzerrungsfrei berechnet (und dargestellt) werden, wenn ein vollständiger Datensatz zugrunde liegt. In diesem Fall hat der AI einen Erwartungswert von 1 und einen Wertebereich von $[0, \infty]$. Der vorliegende Aufsatz zeigt jedoch, dass bei Verteilungen mit fehlenden oder Nullwerten systematische Verzerrungen bei der Berechnung des AI auftreten, die zu fehlerhaften Interpretation führen können. Wir stellen eine neue Fächer-Klassifikation vor, die solche systematischen Verzerrungen ausschließt und eine zuverlässige Berechnung des AI ermöglicht. Als Beispiel dient uns das Feld der staatlichen Universitäten in Deutschland. Ausgehend von der Fächersystematik des Statistischen Bundesamtes entwickeln wir für die technischen und nicht-technischen Universitäten jeweils getrennte Klassifikationen aggregierter Fächergruppen, mit denen eine verzerrungsfreie Berechnung des AI möglich ist.

Der Aufsatz ist wie folgt aufgebaut. Zunächst erläutern wir anhand einschlägiger Studien methodische Probleme, die bei der Berechnung des AI mit unvollständigen Datensets entstehen können (Abschnitt 2). Auf dieser Basis erläutern wir eine Vorgehensweise zur Erstellung zwei neuer Fächerklassifikationen, mithilfe derer eine weitgehend verzerrungsfreie AI-Berechnung, zum einen für technische, zum anderen für nicht-technische Universitäten, möglich ist (Abschnitt 3). Der Aufsatz schließt mit überarbeiteten Profilkarten einer technischen und einer nicht-technischen Universität und einer Diskussion über die Vor- und Nachteile der beiden neuen Fächersystematiken (Abschnitt 4).

2. Berechnungsprobleme fachlicher Spezialisierung bei fehlenden Daten

Studien mit Informationen zu fachlicher Profilbildung an Universitäten sind vergleichsweise selten. Zu ihnen gehört Piro et al. (2017), die das bibliometrische Profil aller dänischen, finnischen, norwegischen und schwedischen Universitäten mithilfe des RSI im Zeitraum 1999-2014 entlang einer 16 Fächergruppen umfassenden Klassifikation ermitteln. Auch die Studie von Heinze et al. (2019) ist zu den neueren Studie zu zählen, die mithilfe des RESP für Deutschland bibliometrisch definierte Fachprofile staatlicher Universitäten und zudem vergleichbare Informationen für wissenschaftliches Personal, Grund- und Drittmittel sowie Studierende im Zeitraum 1992-2018 entlang einer 56 Fächergruppen umfassenden Klassifikation zur Verfügung stellt. Im europäischen Kontext wurden zudem Studien durchgeführt, die den positiven Einfluss universitärer Fachprofile auf die regionale wirtschaftliche Entwicklung nachweisen konnten, ohne dass jedoch die jeweils zugrunde liegenden Profilkarten der untersuchten Hochschulen dokumentiert wurden (z.B. Teixeira et al. 2012, Bonaccorsi et al. 2013).

In den Abbildungen 1 und 2 sind Ausschnitte der Heat Maps von Piro et al. (2017) und Heinze et al. (2019) dargestellt. In beiden Fällen sind Lücken und teilweise auch recht erhebliche Schwankungen sichtbar. Die jeweiligen Gründe hierfür sind verschieden. Im Fall von Piro et al. (2017) entstehen sie, weil die Autoren heterogene Hochschularten miteinander vergleichen: erstens hochspezialisierte Hochschulen mit nur wenigen Fächern und zugleich Volluniversitäten mit einem breiten Fächerspektrum; zweitens technische und nicht-technische Universitäten; und drittens Universitäten mit und ohne Medizin / Universitätsklinikum. Diese heterogene Fallauswahl führt dazu, dass es immer Einheiten gibt, die bestimmte Fächergruppen nicht aufweisen, was entweder zu fehlenden Werten führt (weiße Flecken) oder zu Nullwerten (maximaler Negativwert des Wertebereichs). Beispielsweise weist die Copenhagen Business School keine Publikationen im Bereich Biologie oder in der Biomedizin auf, was zu einem Profilverwert von jeweils -1 führt; der Bereich Chemie fehlt hier völlig, mit der Folge, dass ein weißer Fleck in der Profilkarte entsteht. Umgekehrt publiziert das Universitätsklinikum Aarhus nicht in den Business Studies, was zu einem Wert von -1 führt.

Hier Abbildung 1

Auch die Profilkarten der TU Dortmund und der Universität Bamberg, die wir beispielhaft für viele weitere Heatmaps heranziehen, weist Lücken auf (Abb. 2a, 2b). Diese entstehen jedoch im Gegensatz zu Piro et al. (2017) nicht aufgrund einer heterogenen Fallauswahl, denn in

Heinze et al. (2019) bzw. auf <https://fachprofile.uni-wuppertal.de> werden nur Volluniversitäten ohne das Fach Medizin / Universitätsklinikum betrachtet; zudem werden technische und nicht-technische Universitäten separat ausgewertet. Die Lücken entstehen vielmehr durch eine sehr differenzierte Fächerklassifikation, die n=56 Fächergruppen umfasst (während es bei Piro et al. (2017) gerade einmal n=16 Fächergruppen sind). Durch die hochdifferenzierte Klassifikation enthält die Profilkarte viele kleine Fächer, für die in einzelnen Jahren entweder Informationen fehlen oder bei denen bereits kleinere Schwankungen der Variablenwerte zu stark wechselnden Werten in der Heatmap führen.

Hier Abbildung 2a

Hier Abbildung 2b

Die fehlenden Werte in den Heatmaps sind Symptom eines tiefer liegenden methodischen Problems, das auch diejenigen Bereiche der Heatmap betrifft, die lückenlos dargestellt sind. Bevor dieses Problem näher erläutert wird, sollen zunächst die beiden Indizes AI und RESP eingeführt werden.

2.1. Activity Index (AI) und Relativer Spezialisierungsindex (RESP)

Wir gehen von einem Datenset D aus, bei dem sich jede Beobachtung in D eindeutig über zwei (oder mehrere) kategoriale Variablen identifizieren (Index-Variablen) lässt und zu jeder Kombination der identifizierbaren Variablen eine empirische Beobachtung existiert. Sei v eine numerische, nicht negative Variable des Datensets D , außerdem I und J zwei Index-Variablen. Dann bezeichnen wir mit v_{ij} den Wert der Variablen v zur Beobachtung (i,j) , $i \in I, j \in J$. Der AI ist folglich definiert als:

$$AI_{ij} := AI(v_{ij}) := \frac{v_{ij} / \sum_i v_{ij}}{\sum_j v_{ij} / \sum_{i,j} v_{ij}}$$

Im Falle eines vollständigen Datensets vergleicht der AI den Anteil, den ein Fach i bezüglich einer Variablen (zum Beispiel Professuren) an einer Universität j einnimmt, mit dem Anteil des Faches an allen Universitäten. Bei einem $AI > 1$ nimmt das Fach i an der Universität j einen überdurchschnittlich großen Anteil ein; bei einem $AI < 1$ nimmt das Fach i an der Universität j einen unterdurchschnittlichen Anteil ein.

Bei einer Normalverteilung der Ausgangswerte hat der AI einen Erwartungswert von 1 und einen Wertebereich von $[0, \infty]$. Außerdem sind seine Werte log-normalverteilt. Das legt nahe, den AI zu logarithmieren, aber auch daraus resultierende Wertebereich von $(-\infty, +\infty)$ ist für die Darstellung von Heatmaps eher ungeeignet. Eine Reduktion des Wertebereich erreicht

man durch Transformation der $\log(\text{AI})$ -Werte mit dem tangens hyperbolicus. Nach einer Skalierung um den Faktor 100 erhält man den *Relativen Spezialisierungsindex* (RESP, Heinze et al. 2019), der wie folgt definiert ist:

$$\text{RESP}_{ij} := \text{RESP}_{ij}(\text{AI}_{ij}) = 100 \frac{\text{AI}_{ij}^2 - 1}{\text{AI}_{ij}^2 + 1}.$$

Der RESP hat einen Wertebereich von $[-100, +100]$. Ein Wert $\text{RESP} > 0$ entspricht dabei einem $\text{AI} > 1$; ebenso entspricht ein Wert $\text{RESP} < 0$ einem $\text{AI} < 1$. Somit eignet sich der RESP ideal zur Darstellung von AI-Werten, zum Beispiel für Heatmaps. Siehe dazu auch die Website <https://fachprofile.uni-wuppertal.de>, auf der zahlreiche Heatmaps von Universitäten und Fächern dargestellt sind.

2.2. Konsequenzen unvollständiger Datensets für den AI und den RESP

Das in Abb. 1 und 2 angedeutete methodische Problem besteht darin, dass nicht immer ein vollständiges Datenset vorliegt. Zur Verdeutlichung können Abb. 3 und 4 verglichen werden. In Abb. 3 ist die Häufigkeit der Professor:innen in universitären Fächern dargestellt, so wie sie für unseren Datensatz staatlicher Universitäten in Deutschland vorliegt, allerdings ohne Nullwerte. Die Anzahl reicht von einer Professor:in pro Fach bis zu 98 Professor:innen (Wirtschaftswissenschaften, Universität zu Köln, Wintersemester 2017/18), wobei Abb. 3 aus Darstellungsgründen nur Kontingente bis maximal 50 Professor:innen pro Fach und Universität anzeigt. Die Verteilung der Variable folgt einer Log-normal Verteilung (Abb. 3).

Hier Abbildung 3

Stellt man Abb. 3 die Verteilung gegenüber, in der auch Nullwerte enthalten sind, erhält man Abb. 4, in der der Kerndichteschätzer keine Log-normal Verteilung mehr anzeigt. Dies liegt daran, dass bei der Berechnung des AI die Summe $\sum_{i \in I, j \in J} v_{ij}$ gebildet wird. Fächer an Universitäten mit einem Null-Kontingent an Professor:innen werden bei der Berechnung des AI folglich mitberücksichtigt. Da über alle Fächer $i \in I$ an allen Universitäten $j \in J$ summiert wird, fließen auch fehlende Werte in die Berechnung mit ein. Das Null-Kontingent scheint die Verteilung regelrecht zu sprengen (Abb. 4).

Hier Abbildung 4

Bei der in Abb. 4 gezeigten Verteilung kann man nicht mehr davon ausgehen, dass der AI-Erwartungswert noch 1 ist, was für seine Interpretation aber erhebliche Folgen hätte. Wie weiter oben erwähnt, bedeutet eine Fächer-Universität-Kombination mit einem Wert $\text{AI} > 1$,

dass ihr Fachanteil überdurchschnittlich ausgeprägt ist. Liegt der Erwartungswert des AI nun aufgrund einer ungünstigen Verteilung über 1, dann existieren überdurchschnittlich viele Fächer-Universität-Kombinationen mit einem Wert $AI > 1$. Dies ist gleichbedeutend mit einer positiven Verzerrung (*bias*), bei der eine Universität in einem Fach einen Wert $AI > 1$ nicht deswegen erreicht, weil ihr Fachanteil überdurchschnittlich ausgeprägt ist, sondern weil unter Umständen sehr viele Ausfälle (Nullwerte) bei anderen Universitäten in diesem Fach existieren.

2.3. Monte-Carlo Simulationen unvollständiger Datensets

Die AI-Verzerrung haben wir in einem Monte-Carlo Experiment untersucht (Carsey and Harden 2013). Hierfür wurden mit verschiedenen Datensets Simulationen durchgeführt. Die Fächervariable i wurde dabei mit 15, 25, 50 und 100 festgelegt. Die Institutionsvariable j wurde mit 10, 50, 100, 150 und 300 Universitäten angesetzt. Die untere Grenze für die Variable v („Professor:innen“) war stets 0, die obere Grenze wurde mit 35, 60, 85, 110, 135, 160, 185, 210 und 235 definiert. Die Werte von v wurden zufällig generiert auf Basis einer Log-normal Verteilung. Zudem wurde bei jedem Datenset eine bestimmte Anzahl von Werten, welche zufällig, aber gleichverteilt ausgesucht wurden, durch fehlende Werte ersetzt. Die Anzahl der Werte, die ersetzt wurden, entsprachen 0%, 5%, 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75% und 80% des simulierten Datensatzes. Jede dieser Konfigurationen wurde mit 1.000 unterschiedlichen Sets an Zufallszahlen und zufälligen Ausfällen getestet. Insgesamt wurden für 3,06 Millionen verschiedene zufällige Datensets die zugehörigen AI-Werte und der AI-Mittelwert (= AI-Erwartungswert) berechnet (Tab. 1).

Hier Tabelle 1

Das Ergebnis der Monte-Carlo Simulation (Abb. 5) zeigt, dass der durchschnittliche Erwartungswert von AI von fast 1 (genau: 1,004) bei keinen Datenausfällen auf 5,099 bei 80 Prozent Datenausfällen steigt. Je mehr Datenausfälle ein Datenset aufweist, umso höher fallen die AI-Erwartungswerte der universitären Fächer aus.

Hier Abbildung 5

Diese positive Verzerrung wirkt sich auch auf die RESP-Werte aus (Abb. 6). Während bei Log-normal verteilten Variablen (wie in unserem Fall) der RESP-Erwartungswert durchschnittlich bei -25 liegt, steigt er bei Datenausfällen (=fallende Abdeckungsquote) deutlich an. Bei einer Abdeckungsquote von 65% liegt der RESP-Erwartungswert bei 0, während er bei einer Abdeckungsquote von 20 Prozent bereits bei 50 liegt. Gemäß Abb. 6 gilt bei einer Abdeckungsquote von unter 60%, dass positive RESP-Werte nicht unbedingt auf

überdurchschnittliche Fächeranteile hindeuten, sondern durch hohe RESP-Werte aufgrund von Datenausfällen.

Hier Abbildung 6

2.4. Anwendung auf unseren Hochschuldatensatz

Wir übertragen die bisherigen Befunde nun auf einen realen Hochschuldatensatz. Dieser besteht aus staatlichen deutschen Universitäten im Zeitraum 1995-2018 (Heinze et al., 2019). Von den 82 Hochschulen in staatlicher Trägerschaft schließen wir hochspezialisierte Universitäten (WR 2010) von der weiteren Analyse aus, um die bei Piro et al. (2017) beobachteten Datenlücken zu vermeiden. Zudem bleiben auch Universitäten mit erheblichen Datenausfällen unberücksichtigt. Damit umfasst unsere Analyse 68 staatliche Universitäten, davon 51 nicht-technische (NTU) und 17 technische (TUs). Aus institutioneller Perspektive können als „technisch“ diejenigen Universitäten definiert werden, die entweder einen entsprechenden Namenszusatz führen oder Mitglied in der Allianz „TU9“ sind (TU9 2019). Heinze et al. (2019) zufolge sind bei den TUs die technischen Fächer personell und finanziell stärker als die Geistes- und Sozialwissenschaften vertreten. Genau diese Eigenschaft unterscheidet sie von den NTUs, bei denen diese Ungleichverteilung weniger stark ausgeprägt ist. Eine Auflistung befindet sich in den Anhängen 1 und 2.

Für diese Universitäten wurden Personal-, Finanz- und Studierendendaten direkt vom Statistischen Bundesamt im Rahmen einer Sonderauswertung bezogen. Sie entsprechen den Daten der Fachserie 11, Reihen 4.1, 4.4 und 4.5 und wurden auf der Ebene der Universitäten und deren Lehr- und Forschungsbereichen aufbereitet. Wir betrachten alle Fächer ohne die Medizin, weil eine Trennung der Klinikeinheiten von ihren angeschlossenen Universitäts-einrichtungen erst ab dem Jahr 2000 möglich ist. Bei der Aufarbeitung der Daten wurde auf institutionelle Kontinuität geachtet, besonders bei den ehemaligen Gesamthochschulen in Nordrhein-Westfalen und bei fusionierten Einrichtungen.

Wie Tab. 2 zu entnehmen ist, wären bei lückenloser Vollerhebung 91.392 Beobachtungen möglich. Tatsächlich fehlen in unserem Datensatz aber 46.058 Werte. Wir erreichen somit eine Abdeckungsquote (AQ0) von lediglich 49,6 Prozent, wobei hier Beobachtungen mit Professuren = 0 mitgezählt werden. Ignoriert man obendrein diese Nullwerte, so erhält man eine Abdeckungsquote (AQ1) von lediglich 45,1 Prozent. Wie erwähnt, können überproportional viele Beobachtungen mit dem Wert 0 die RESP-Werte verzerren. Daher ist AQ1 aussagekräftiger. Mit einer Abdeckungsquote von rund 45 Prozent sind wir aber in einem Bereich, in dem die RESP-Werte verzerrt werden. Für unseren Datensatz ist damit die Aussage,

dass ein $RESP > 0$ eine relative Spezialisierung indiziert, nicht ohne weiteres haltbar. Der durchschnittliche AI-Wert unseres Datensatzes liegt in keinem Jahr unter 1,74 und weicht damit deutlich vom Erwartungswert von 1,0 ab; bis 2018 steigt er sogar auf über 2,0 an (Abb. 7).

Hier Abbildung 7

Welche Konsequenzen ergeben sich aus dieser empirischen Verteilung der AI-Werte für die Verteilung der RESP-Werte? Idealerweise folgt die Verteilung der RESP Werte einer Normalverteilung mit Wertebereich $[-100,100]$, in Abb. 8 durch die rote Kurve dargestellt. Allerdings liegt bei unserem Hochschuldatensatz eine um den durchschnittlichen RESP-Wert von 15,67 nach rechts verschobene Normalverteilung (blaue Kurve) vor. Die überproportionale Häufung von hohen RESP Werten rührt nicht von besonderen universitären Fachprofilen her, sondern weil viele Beobachtungspunkte in unserem Datensatz als fehlende (NA) oder Nullwerte codiert sind.

Hier Abbildung 8

3. Konstruktion geeigneter Klassifikationen von Fächergruppen

Wie bereits erwähnt, entstehen die Lücken in unserem Hochschuldatensatz durch eine zu differenzierte Fächerklassifikation des Statistischen Bundesamtes, die $n=56$ Lehr- und Forschungsbereiche umfasst. Diese Klassifikation stellt die mittlere Aggregationsstufe dar, welche das Bundesamt verwendet. Die nächsthöhere Aggregationsstufe besteht aus $n=9$ Fächergruppen, einschließlich Medizin. Es verbleiben für unseren Datensatz somit 8 Fächergruppen, mit denen sich jedoch kein einigermaßen differenziertes Bild der universitären Fächerprofile mehr zeichnen lässt. Hierfür ist eine Klassifikation gefragt, die mehr als 8 und weniger als 56 Gruppen umfasst: $56 < K < 8$.

Bevor wir unsere eigenen Klassifikationen für die NTUs und TUs entwickeln, werfen wir noch einen Blick auf verfügbare internationale Klassifikationen. *Erstens* gibt es die „International Standard Classification of Education“ (ISCED), welche von der UNESCO und vom European Tertiary Education Register (ETER) verwendet wird. Sie umfasst *neun Fächergruppen*, die als „Fields of Education“ bezeichnet werden: General Programmes and Qualifications; Education; Humanities and Arts; Social Sciences; Business and Law; Natural Sciences, Mathematics and Statistics; Information and Communication Technologies; Engineering, Manufacturing and Construction; Agriculture, Forestry, Fisheries and Veterinary; Health and Welfare; and Services (Daraio et al. 2020; Lepori 2019). *Zweitens* gibt es die von der OECD im Frascati-Handbuch als

„Fields of Research and Development“ bezeichneten *sechs Fächergruppen*: Natur- und Ingenieurwissenschaften sowie Technologie; Medizinische und Gesundheitswissenschaften; Agrarwissenschaften und Veterinärmedizin; Sozialwissenschaften; Geisteswissenschaften; und Kunst (OECD 2018). *Drittens* verwendet das Times Higher Education (THE) für sein globales Hochschulranking *elf Fächergruppen*: Arts and Humanities; Clinical and Health; Engineering; Computer Science; Life Sciences; Physical Sciences; Business and Economics; Social Sciences; Psychology; Law; and Education (THE 2021). *Viertens* hat das Nordic Institute for Studies in Innovation, Research and Education (NIFU) die im Web of Science gelisteten 250 subject categories in *16 Fächergruppen* aggregiert: Agriculture, Fisheries & Forestry; Biology; Biomedicine & Molecular Biosciences; Business Studies & Economics; Chemistry; Clinical Medicine; Computer & Information Sciences; Engineering; Geosciences; Health Sciences; Humanities; Materials Science; Physics; Psychology; and Social Sciences (Piro et al. 2017, S. 9-10).

Den vier Klassifikationen ist gemein, dass der jeweilige Zuschnitt der Fächergruppen nicht weiter begründet, sondern offensichtlich pragmatisch und/ oder auf Plausibilitätserwägungen hin erstellt wurde. Wir möchten die jeweilige Fächergruppierung hier nicht kritisieren. Allerdings hielten wir es für eine willkürliche Entscheidung, einfach eine der vier Klassifikationen auszuwählen, ohne dass sichergestellt wäre, dass eine von ihnen das weiter oben erwähnte methodische Problem verzerrter AI-Werte durch unvollständige Datensets unmittelbar adressiert. Vielmehr entwickeln wir im Folgenden eine NTU-Fächerklassifikation und eine TU-Fächerklassifikation. Bei beiden steht das Ziel im Vordergrund, dass die bisherigen Abdeckungsquoten durch geeignete Aggregation von Fächern erhöht werden.

Bisher haben wir die Abdeckungsquote auf der Makroebene betrachtet, also auf der Ebene des gesamten Feldes der staatlichen Universitäten. Wir haben für die Variable „Professor:innen“ 41.243 Fälle > 0 bei 91.392 möglichen Kombinationen aus der Zahl der Jahre ($n=24$), der Universitäten ($n=68$) und der Fächer ($n=56$) beobachtet, was einer Abdeckungsquote von 45,1 Prozent ($=41.243/91.392$) entspricht. Eine solche Betrachtung kann man nicht nur auf der Makroebene, sondern auch auf der Mesoebene der Fächer durchführen.

Zum Beispiel zählen wir für die evangelische Theologie 1.018 Beobachtungen bei Professor:innen > 0 bei 1.632 möglichen Kombinationen aus den identifizierenden Variablen: Jahr ($n=24$) und Universität ($n=56$). Die evangelische Theologie hat also eine Abdeckungsquote von 62,4 Prozent ($=1.018/1.632$). Die katholische Theologie hat in der Variable „Professor:innen“ dagegen nur 754 Beobachtungen > 0 und damit eine Abdeckungsquote von 46,2 Prozent ($=754/1.632$). Aggregieren wir nun die evangelische und die katholische

Theologie zum neuen Fach „Theologie“, dann erhalten wir 1.095 Beobachtungen > 0 , wovon 341 Beobachtungen Professor:innen > 0 in evangelischer Theologie, 77 Beobachtungen Professor:innen > 0 in katholischer Theologie, und 677 Beobachtungen Professor:innen > 0 sowohl in evangelischer als auch katholischer Theologie sind. Da die Zahl der möglichen Kombinationen mit 1.632 gleich bleibt, erhalten wir eine Abdeckungsquote von 67,1 Prozent ($=1.095/1.632$).

Der Vorteil dieser Vorgehensweise ist offensichtlich. Wir haben die Abdeckungsquote der katholischen Theologie von 46,2 Prozent auf eine weitaus bessere Abdeckungsquote des neuen Faches Theologie von 67,1 Prozent verbessert. Der Nachteil ist aber auch offensichtlich: wir können nicht mehr zwischen evangelischer und katholischer Theologie unterscheiden. In der Erstellung der AI-Werte würde nur noch das neue Fach Theologie erscheinen. Wir erhalten ein neues Set von Fächern, in diesem Beispiel nun mit 55 statt 56 Fächern.

Trotz des Nachteils, dass eine feinere Differenzierung zwischen den Fächern verloren geht, haben wir uns dafür entschieden, diese Vorgehensweise systematisch auf die 56 Lehr- und Forschungsbereiche umfassende Klassifikation des Statistischen Bundesamtes anzuwenden, um auf diese Weise die erwähnte Verzerrung der AI-Werten zu beseitigen. Hierfür haben wir in einem ersten Schritt die Abdeckungsquote für alle 56 Fächer errechnet und dann in einem zweiten Schritt verwandte Fächer aggregiert, und zwar immer dann, wenn ihre Abdeckungsquoten einen niedrigen Wert aufwiesen und eine Aggregation zu einer deutlichen Verbesserung der Abdeckungsquoten führte. Dieses Vorgehen führen wir getrennt für die TUs und NTUs durch.

Für die bisherigen Lehr- und Forschungsgebiete des Statistischen Bundesamtes („Fächer“) haben wir die Abdeckungsquote in 5-Jahres-Zyklen berechnet, angefangen bei 2014-2018. Damit nehmen wir Rücksicht auf neuere Fächer, wie zum Beispiel Informatik, welches im ältesten 5-Jahreszyklus 1994-1999 im Durchschnitt nur von 37 NTUs und 15 TUs angeboten wurde, im jüngsten 5-Jahreszyklus 2014-2018 aber von 45 NTUs und allen 17 TUs. Dieses Vorgehen erzeugt keine Mehrarbeit oder gar verschiedene Fächersystematiken, denn wie wir sehen werden, lassen sich die Ergebnisse des jüngsten 5-Jahreszyklus auf die anderen Zyklen übertragen.

Bevor wir mit der eigentlichen Aggregation der Fächer beginnen, gehen wir auf eine Besonderheit der Fächersystematik des Statistischen Bundesamtes ein. Diese kennt – wie wir sie hier bezeichnen – „Residualfächer“, welche Studiengebiete subsumieren, die sich nicht eindeutig einem der anderen Fächer zuordnen lassen. Ein solcher Fall ist etwa die „Geschichte der Mathematik und Naturwissenschaften“, die offensichtlich weder ganz der Mathematik,

noch einer der anderen Naturwissenschaften zuzuordnen ist. Deswegen ordnet das Statistische Bundesamt diesen Fall dem Residualfach „Mathematik, Naturwissenschaften allgemein“ zu. Der Umgang mit den insgesamt sechs Residualfächern ist nicht-trivial, da sie eben nicht ein Fach, sondern stets eine ganze Fachgruppe („Geisteswissenschaften“, „Rechts-, Wirtschafts- und Sozialwissenschaften“, „Mathematik, Naturwissenschaften“, „Agrar-, Forst- und Ernährungswissenschaften“, „Ingenieurwesen“, und „Kunst, Kunstwissenschaft“) abdecken. Wir haben uns für zwei verschiedene Vorgehensweisen bezüglich dieser Fächer entschieden. Erstens: sind wir bei der Auswahl der Aggregationskandidaten (für das Prozedere siehe weiter unten) zu dem Schluss gekommen, dass eine komplette Fachgruppe aggregiert werden soll, dann haben wir das betreffende Residualfach zu dieser Fachgruppe hinzugefügt. Zweitens: in allen anderen Fällen kommen die Residualfächer in eine Sammelkategorie. Dadurch gehen zwar alle fachbezogenen Informationen dieser Residualfächer verloren, aber für die Berechnung der AI- und RESP-Werte werden diese Fächer weiterhin berücksichtigt. Das hat den Vorteil, dass die Grundgesamtheit, auf die sich die Berechnung der AI- und RESP-Werte beziehen, beziehungsweise die Variablensummen, welche je Universität gebildet werden, unverändert bleiben.

3.1. Neue Fächerklassifikation für die TUs

Tabelle 3 zeigt die Abdeckungsquoten für die TUs und den 5-Jahreszyklus 2014-2018. Man beachte, dass zwar 56 Fächer in unserem Datenset aufgeführt werden, aber im Zeitraum 2014-2018 von TUs nur 52 Fächer angeboten werden. Es fehlen „Bibliothekswissenschaft, Dokumentation, Publizistik“; „Außereuropäische Sprach- und Kulturwissenschaften“; „Verwaltungswissenschaften“; und „Darstellende Kunst, Film und Fernsehen, Theaterwissenschaft“. Es ist zudem in Tabelle 3 ersichtlich, dass zahlreiche Lehr- und Forschungsgebiete sehr niedrige Abdeckungsquoten aufweisen und folglich einer Aggregation bedürfen.

Hier Tabelle 3

Bei der Aggregation beginnen wir von unten in Tabelle 3, also mit dem Fach, welches die geringste Abdeckungsquote aufweist („Bottum-Up“), dies ist „Sonderpädagogik“. Dieses Fach gehört inhaltlich zu den Erziehungswissenschaften und beide Fächer befinden sich in derselben Fachgruppe der Systematik des Statistischen Bundesamtes (StBA 2013). Folglich werden beide Fächer zu einem neuen Fach „Erziehungswissenschaften und Sonderpädagogik“ aggregiert. Das neue Fach hat eine Abdeckungsquote von 72 Prozent. Wir überspringen das Residualfach „Agrar-, Forst- und Ernährungswissenschaften allgemein“ und ordnen in einem weiteren Schritt das Fach „Regionalwissenschaften“ dem inhaltlich passenden Fach „Rechts-

und Gesellschaftswissenschaften“ zu. Wir wiederholen die Zuordnung von Fächern mit geringen Abdeckungsquoten zu einem geeigneten Aggregationsfach solange, bis aller Fächer entweder eine Abdeckungsquote von 100 Prozent erreicht haben oder eine weitere Aggregation inhaltlich nicht sinnvoll oder möglich ist.

Tabelle 4 gibt die neue Fächerstruktur für TUs wieder und zeigt zugleich das von uns verwendete Aggregationsschema. Die 56 Fächer wurden für die TUs zu 14 neuen Fächergruppen zusammengefasst. Dabei sind zwei Fälle zu unterscheiden. Zum einen existieren aggregierte Fächer wie „Biologie, Chemie und Pharmazie“. Zum anderen gibt es aber auch nicht-aggregierte Fächer, wie „Architektur, Innenarchitektur“ oder „Bauingenieurwesen“, welche eine so hohe Abdeckung besessen haben, dass sie nicht weiter aggregiert werden mussten. In der Neubenennung von Fächer unterscheiden wir ebenfalls in zwei Fälle. Wo möglich, haben wir als Name für ein neues Fach einfach alle Aggregationskandidaten aufgezählt, wie bei „Biologie, Chemie und Pharmazie“. Es existieren aber auch neu konstruierte Fächer, bei denen nicht alle früheren Fächer im Namen erscheinen, wie zum Beispiel bei „Geistes- und Sprachwissenschaften“. In solchen Fällen haben wir stets versucht einen Namen zu wählen, der die aggregierten Fächer bestmöglich repräsentiert.

Hier Tabelle 4

Tabelle 5 zeigt die Abdeckungsquoten der neu konstruierten Fächer der TUs. Anders als Tabelle 4, welche die Abdeckungsquoten nur für den 5-Jahreszyklus 2014-2018 zeigt, werden hier alle 5-Jahreszyklen von 1994 bis 2018 wiedergegeben. Im jüngsten 5-Jahreszyklus ist keine Abdeckungsquote unter 70 Prozent (ausgenommen die Residualfächer-Kategorie, die nur zum Zwecke der AI-Berechnung weitergeführt wird). Wir haben also die Abdeckungsquote von 45 Fächern durch Aggregation verbessert. Nun ist es Ermessungssache, ob weitere Fächer aggregiert werden und man somit eine höhere, untere Grenze für die Abdeckungsquoten erreicht. Es sprechen aber zwei Punkte dafür, die neue TU-Fächersystematik so zu belassen.

Erstens reduziert die neue Fächersystematik die Verzerrung der AI- und RESP-Werte hinreichend, so dass eine weitere Aggregation als nicht nötig erscheint. Zweitens würden mit jeder weiteren Aggregation Informationen verloren gehen, so dass die Fachprofile der TUs immer allgemeiner und damit weniger differenziert ausfielen. So kennt die neue Fächersystematik nur das Fach „Geistes-, Sprach- und Kunstwissenschaften“ und differenziert nicht zwischen einzelnen Sprachdisziplinen. Da der Profilschwerpunkt der TUs auf ingenieurwissenschaftlichen Fächern liegt, nehmen wir den Informationsverlust bei den Sprachwissenschaften als akzeptables Defizit hin. Bei den ingenieurwissenschaftlichen

Fächern unterscheidet die neue Fächersystematik dagegen zwischen „Architektur, Innenarchitektur“; „Bauingenieurwesen“; „Elektrotechnik“; „Maschinenbau / Verfahrenstechnik“; und „Weitere Ingenieurwissenschaften“. Dieses Spektrum an ingenieurwissenschaftlichen Fächern ist breit genug, um ein ingenieurwissenschaftliches Fachprofil der einzelnen TUs erkennbar zu machen. Eine weitere Aggregation der ingenieurwissenschaftlichen Fächer würde zu großen Sammelfächern mit stark eingeschränkter Aussagekraft führen.

Hier Tabelle 5

Zugleich möchten wir darauf hinweisen, dass wir die Abdeckungsquoten in Tabelle 3 für den jüngsten 5-Jahreszyklus 2014-2018 erstellt und anhand dieser Tabelle die „Bottom-Up“-Aggregation durchgeführt haben. Daraus hat sich das in Tabelle 4 dokumentierte Aggregationsschema ergeben.

Nun ließe sich argumentieren, dass die neue Fächersystematik aus Tabelle 4 nicht bruchlos auf die anderen vier 5-Jahreszyklen angewendet werden kann. Diesem Argument folgend bestünde grundsätzlich die Möglichkeit, auch für die anderen 5-Jahreszyklen jeweils Tabellen mit Abdeckungsquoten zu erstellen und auf dieser Basis jeweils zyklenspezifische „Bottom-Up“-Aggregationen durchzuführen. Auf diese Weise würden vier weitere Aggregationsschemata und vier weitere Fächersystematiken entstehen, die unter Umständen nur bedingt vergleichbar wären. Wir haben uns daher dafür entschieden, die Fächersystematik aus Tabelle 4 auf die anderen 5-Jahreszyklen zu übertragen. Tabelle 5 zeigt, wie sich die Abdeckungsquoten der neuen Fächer bezüglich der anderen 5-Jahreszyklen verhalten. Abgesehen von der Residualfächer-Kategorie, gibt es mit „Human-, Erziehungs- und Sportwissenschaften“ nur eine Fächergruppe, deren Abdeckungsquote im Zeitraum 1994-2018 unter 70 Prozent sinkt. Diese Fächergruppe hat in den Zyklen 1994-1998 und 1999-2003 eine Abdeckung von jeweils 65 Prozent. Dieses Defizit halten wir für hinnehmbar, weil ihm als Vorteil eine für alle Jahre (1994-2018) gültige Fächersystematik gegenübersteht, in der für alle anderen Fächergruppen und Zyklen eine Abdeckung von mindestens 70 Prozent erreicht werden.

3.2. Neue Fächerklassifikation für die NTUs

Wir wiederholen das in Abschnitt 3.1 für die TUs beschriebene Vorgehen auch für die NTUs. Tabelle 6 zeigt die Abdeckungsquoten für die NTUs und den 5-Jahreszyklus 2014-2018. Im direkten Vergleich mit den TUs (Tabelle 3) fällt auf, dass Fächer aus dem sozial- und geisteswissenschaftlichen Bereich höhere Abdeckungsquoten als aus dem natur- und ingenieurwissenschaftlichen Bereich aufweisen. Nur die Mathematik hat sowohl bei den TUs

(100%) als auch den NTUs (96%) ähnlich hohe Abdeckungen. Man könnte an vielen weiteren Beispielen zeigen, dass die institutionelle Trennung zwischen TUs und NTUs in der Fächerstruktur der beiden Hochschularten begründet ist. Somit ist auch eine Trennung in zwei getrennte Fächersystematiken sinnvoll.

Hier Tabelle 6

Tabelle 6 ist wieder Ausgangspunkt für die „Bottom-Up“-Aggregation, das heißt zuerst wurde ein Aggregationskandidat für das Fach „Bergbau, Hüttenwesen“ gesucht und die Aggregation erstellt, die Abdeckungsquote der neuen Fächergruppe ermittelt und in die Tabelle eingefügt, um danach mit dem nächsten Fach am Ende der Tabelle fortzufahren. Das Prozedere wird fortgesetzt, bis alle verbliebenen alten oder neu eingefügten Fächergruppen eine möglichst hohe Abdeckung aufweisen. Tabelle 7 zeigt das so entstandene Schema für eine neue Fächersystematik (linke Spalte) und die Lehr- und Forschungsgebiete nach Systematik des Statistischen Bundesamtes (rechte Spalte), welche zur Erstellung der neuen Fächer aggregiert wurden.

Hier Tabelle 7

Die Frage ist sicher gerechtfertigt, ob wir nicht auf unterschiedlichen Wegen zwei Fächersystematiken erstellt haben, die im Prinzip identisch sind. Einer solchen Lesart widerspricht die Tatsache, dass die Unterschiede zwischen Tabelle 4 (Fächersystematik für TUs) und Tabelle 7 (Fächersystematik für NTUs) so offensichtlich sind, dass man von zwei unterschiedlichen Fächersystematiken sprechen kann. Beide Systematiken könnten zwar zu einer gemeinsamen Systematik zusammengefasst werden. Dann hätte man aber für beide Hochschularten das Problem, dass wichtige Differenzierungen in den Fächern zur Profilierung der einzelnen Hochschulen verloren gingen.

Tabelle 8 zeigt die Abdeckungsquoten der neuen Fächersystematik für die NTUs. Wie schon zuvor sind die Abdeckungsquoten des 5-Jahreszyklus 2014-2018 das Ergebnis der „Bottom-Up“-Aggregation. Das Ergebnis ist etwas besser als bei der Aggregation der TU-Fächersystematik, denn alle Fächer haben eine Abdeckung von über 80 Prozent. Eine Ausnahme bildet dabei das Fach „Weitere Ingenieurwissenschaften“ mit gerade einmal 62 Prozent. Eigentlich müsste dieses Fach weiter aggregiert werden. Allerdings findet sich unter den verbleibenden Fächern kein geeigneter Aggregationspartner mehr, daher kann die Abdeckung nicht weiter erhöht werden.

Hier Tabelle 8

Die Abdeckungsquote ist für uns bisher ein Indikator gewesen, um die Verzerrung auf AI- und RESP-Werte zu beschreiben. Die Verbesserung der Abdeckungsquoten sollte zu einer Reduzierung dieser Verzerrung führen, was im Folgenden überprüft wird. Dazu haben wir die AI- und RESP-Werte getrennt nach TUs und NTUs und bezüglich der beiden neuen Fächersystematiken berechnet. Die dabei erzielten Ergebnisse waren sich so ähnlich, so dass wir uns entschieden haben, die AI- und RESP-Werte beider Hochschularten zu poolen und gemeinsam darzustellen (Abb. 9).

Hier Abbildung 9

Die durchschnittlichen AI-Werte für jedes Jahr bewegen sich in Abbildung 9 zwischen 1,01 und 1,03 und damit sehr nahe am (theoretischen) Erwartungswert des AI von 1,0. Vor der Erstellung einer neuen Fächersystematik lag der durchschnittliche AI-Wert bekanntlich bei etwa 1,74. Gleiches sehen wir auch beim Histogramm der RESP-Werte. In Abbildung 5 war ein deutliches Übergewicht ab einem RESP Wert von 75 zu erkennen, welches klar von der Normalverteilung abwich. In Abbildung 10 existieren solche deutlichen Abweichungen dagegen nicht mehr.

Hier Abbildung 10

4. Diskussion

Als Fazit können wir festhalten, dass die beiden neuen Fächerklassifikationen weitgehend unverzerrte AI- und RESP-Werte erreichen. Im Einzelnen können zwar über- oder unterschätzte RESP-Werte noch vorkommen, aber eine systematische Verzerrung wie zuvor ist nicht mehr feststellbar. Dieser Fortschritt wird „erkauft“ durch eine Informationsreduktion von 56 Fächern auf 14 (neue TU-Systematik) bzw. 17 (neue NTU-Systematik). Die Reduktion der Komplexität der Fächersystematik bringt sicherlich einen Verlust an Differenzierungsfähigkeit hinsichtlich der Entwicklung einzelner Fächer mit sich. Sie ist aber für einen weitgehend verzerrungsfreien Vergleich der universitären Fächerprofile unumgänglich.

Als Beispiele für die neuen Profilkarten haben wir spiegelbildlich zu den Abbildungen 2a und 2b wieder die Darstellung des nicht-professoralen wissenschaftlichen Personals der TU Dortmund (TU) und der Universität Bamberg (NTU) ausgewählt. Wie in Abbildung 11a und 11b ersichtlich, enthalten die Profilkarten keine Lücken mehr. Die Heatmaps sind einheitlicher und

aufgrund der geringeren Anzahl von Fächern auch übersichtlicher und damit einfacher zu interpretieren.

Hier Abbildung 11a

Hier Abbildung 11b

Literaturverzeichnis

- Balassa, Bela A. 1965. "Trade liberalisation and "revealed" comparative advantage." *The Manchester School* 33(63):99-123.
- Carsey, Thomas M., and Jeffrey J. Harden. 2013. *Monte Carlo Simulation and Resampling Methods for Social Science*: SAGE Publications Inc
- Costa, Giovanni Abramo and Ciriaco Andrea D'Angelo and Flavia Di. 2014. "A new bibliometric approach to assess the scientific specialization of regions." *Research Evaluation* 23:183-94.
- Daraio, C., R. Bruni, G. Catalano, A. Daraio, G. Matteucci, M. Scannapieco, D. Wagner-Schuster, and B. Lepori. 2020. "A Tailor-made Data Quality Approach for Higher Educational Data." *Journal of Data and Information Science* 5(3):129-60.
- Heinze, Thomas, Dirk Tunger, Joel E. Fuchs, Arlette Jappe, and Paul Eberhardt. 2019. "Research and teaching profiles of public universities in Germany. A mapping of selected fields." Wuppertal: University of Wuppertal.
- Lepori, B. 2019. "How are European Higher Education Institutions funded? New evidence from ETER microdata." ETER.
- Narin, Francis, Mark P. Carpenter, and Patricia Woolf. 1984. "Technological performance assessments based on patents and patent citations." *IEEE Transactions on Engineering Management* EM-31(4):172-83.
- OECD (Ed.). 2018. *Frascati Manual 2015. Proposed Standard Practice for Surveys on Research and Experimental Development*: OECD.
- Piro, F.N. [Editor]. 2017. "Comparing Research at Nordic Higher Education Institutions Using Bibliometric Indicators : Covering the years 1999-2014." Pp. 207.
- Piro, Frederic Niklas, Henrik Aldberg, Dag W. Aksnes, Karlsson Staffan, Yrjö Leino, Anu Nuutinen, Mette Venas Overballe-Petersen, Sigurdir Oli Sigurdsson, and Gunnar Sivertsen. 2017. *Comparing research at nordic higher education institutions using bibliometric indicators covering the years 1999-2014. Policy Paper 4/2017*. Oslo: NIFU.
- Soete, L. G., and Sally M. E. Wyatt. 1983. "The use of foreign patenting as an internationally comparable science and technology output indicator." *Scientometrics* 5(1):31-54.
- THE. 2021. *Methodology for overall and subject rankings for the Times Higher Education World University Rankings 2022*. London.
- TU9. 2019. *Excellence in Engineering and Science. Made in Germany*. Berlin: TU9.
- WR. 2010. *Empfehlungen zur Differenzierung der Hochschulen*. Köln: Wissenschaftsrat.

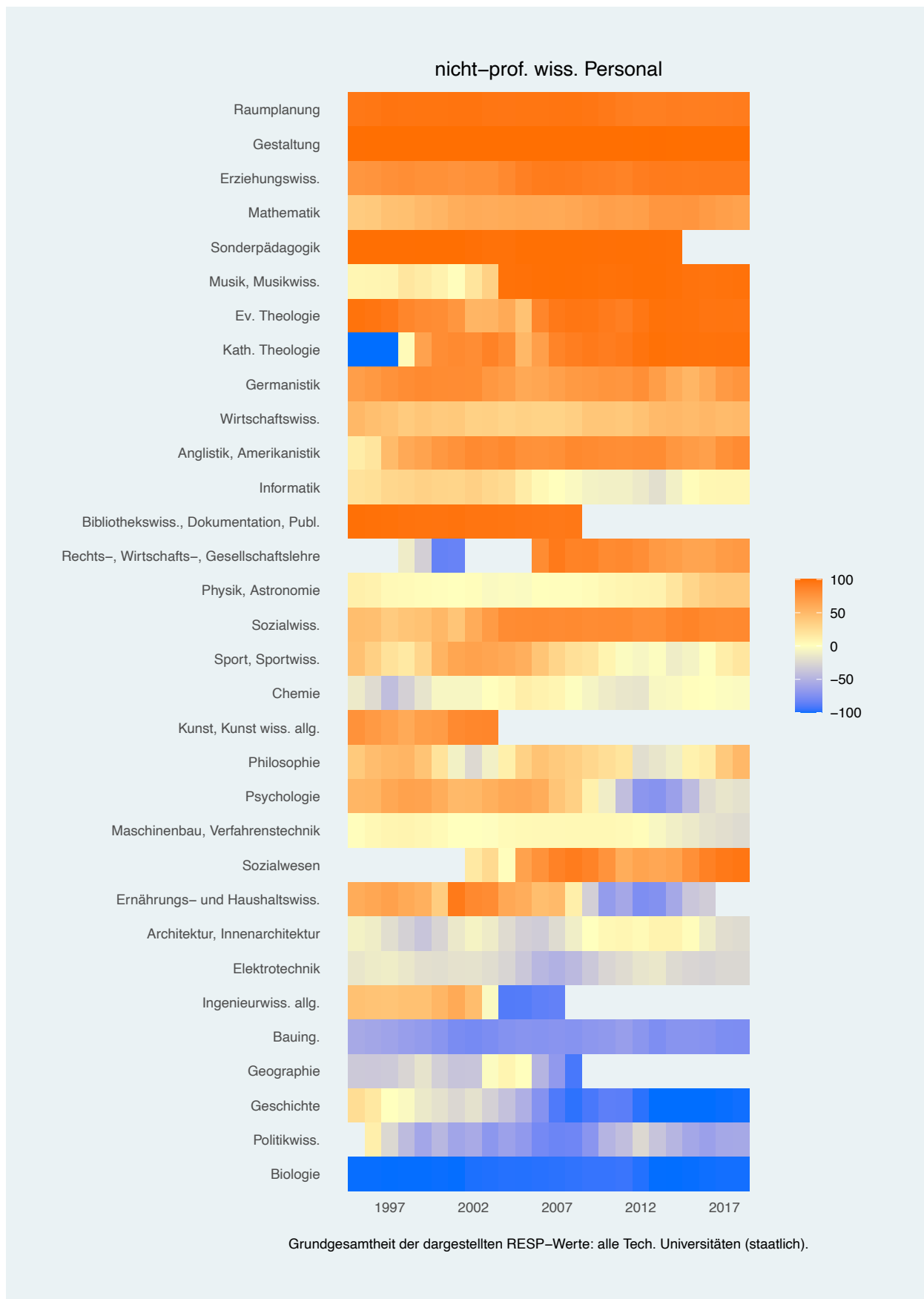
Abbildungen und Tabellen

Abbildung 1: Fächerprofile dänischer Universitäten

Country, university	Agriculture, Fisheries & Forestry	Biology	Biomedicine & Molecular Biosciences	Business Studies & Economics	Chemistry	Clinical Medicine
Aalborg University	-0.4	-0.5	-0.3	0.2	-0.5	-0.4
Aarhus University	0.4	0.3	0.0	0.4	-0.1	-0.2
Copenhagen Business School	-0.9	-1.0	-1.0	0.9		-0.9
Roskilde University	-0.3	0.3	-0.2	0.3	-0.2	-0.8
Technical University of Denmark	0.1	0.0	-0.1	-0.4	0.1	-0.8
University of Copenhagen	0.3	0.2	0.2	-0.1	-0.3	0.1
University of Southern Denmark	-0.6	-0.1	0.1	0.5	-0.2	0.0
Aalborg University Hospitals	-0.7	-0.8	0.0		-0.9	0.6
Aarhus University Hospitals	-0.8	-0.8	0.0	-1.0	-0.9	0.6
Copenhagen University Hospitals	-0.9	-0.9	0.1	-1.0	-0.9	0.6
University of Southern Denmark Hospitals	-1.0	-0.9	0.1		-0.9	0.6
HES Other	-0.4	0.2	-0.9	-0.2	-0.7	-0.7
Other	0.0	-0.3	0.1	-0.1	-0.4	0.2
HES Total	0.1	0.0	-0.1	0.2	-0.3	-0.1
Denmark	0.0	0.0	-0.1	0.1	-0.3	-0.1

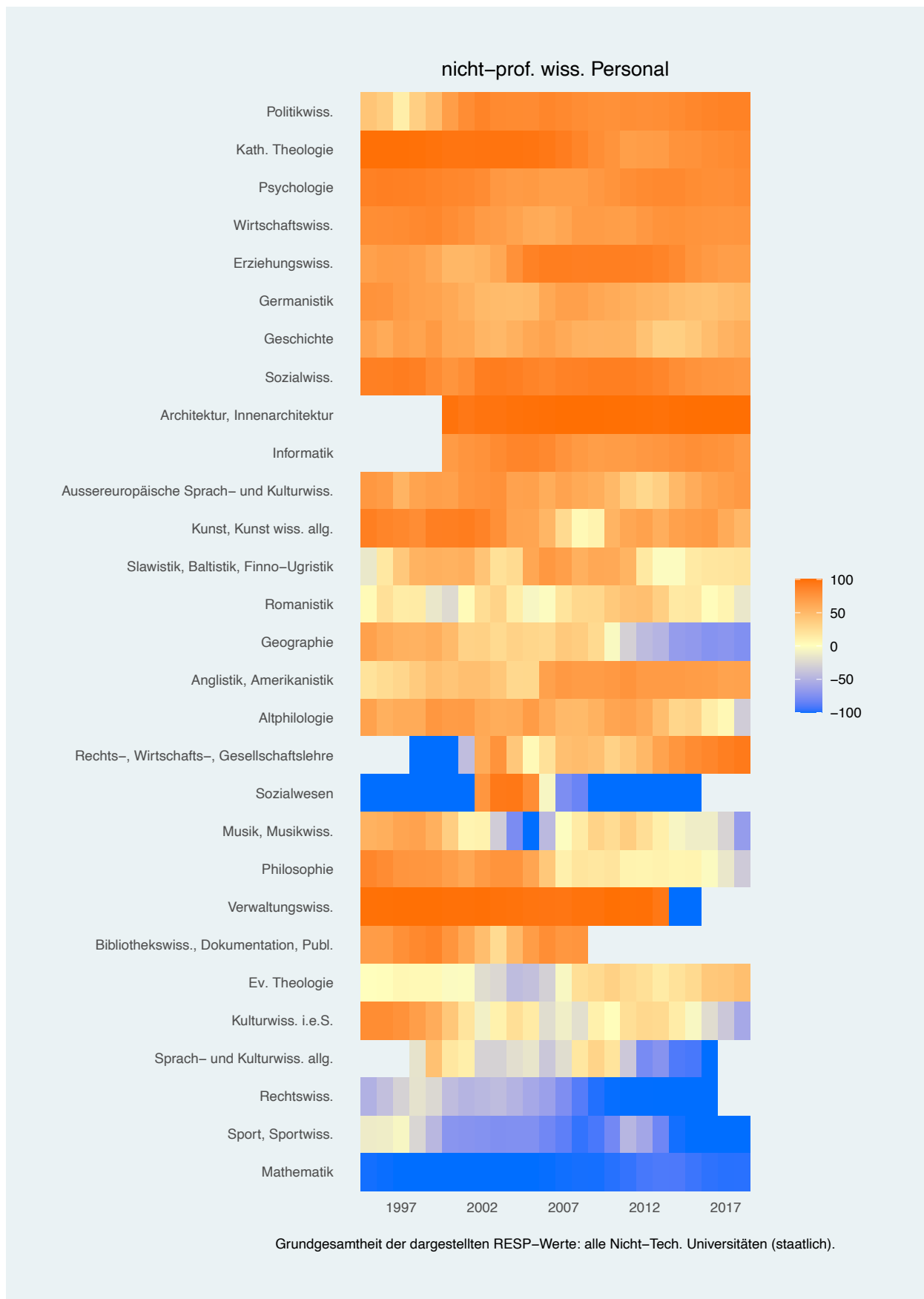
Quelle: Ausschnitt aus Piro et al. (2017, S. 162).

Abbildung 2a: Fächerprofil der Technischen Universität Dortmund



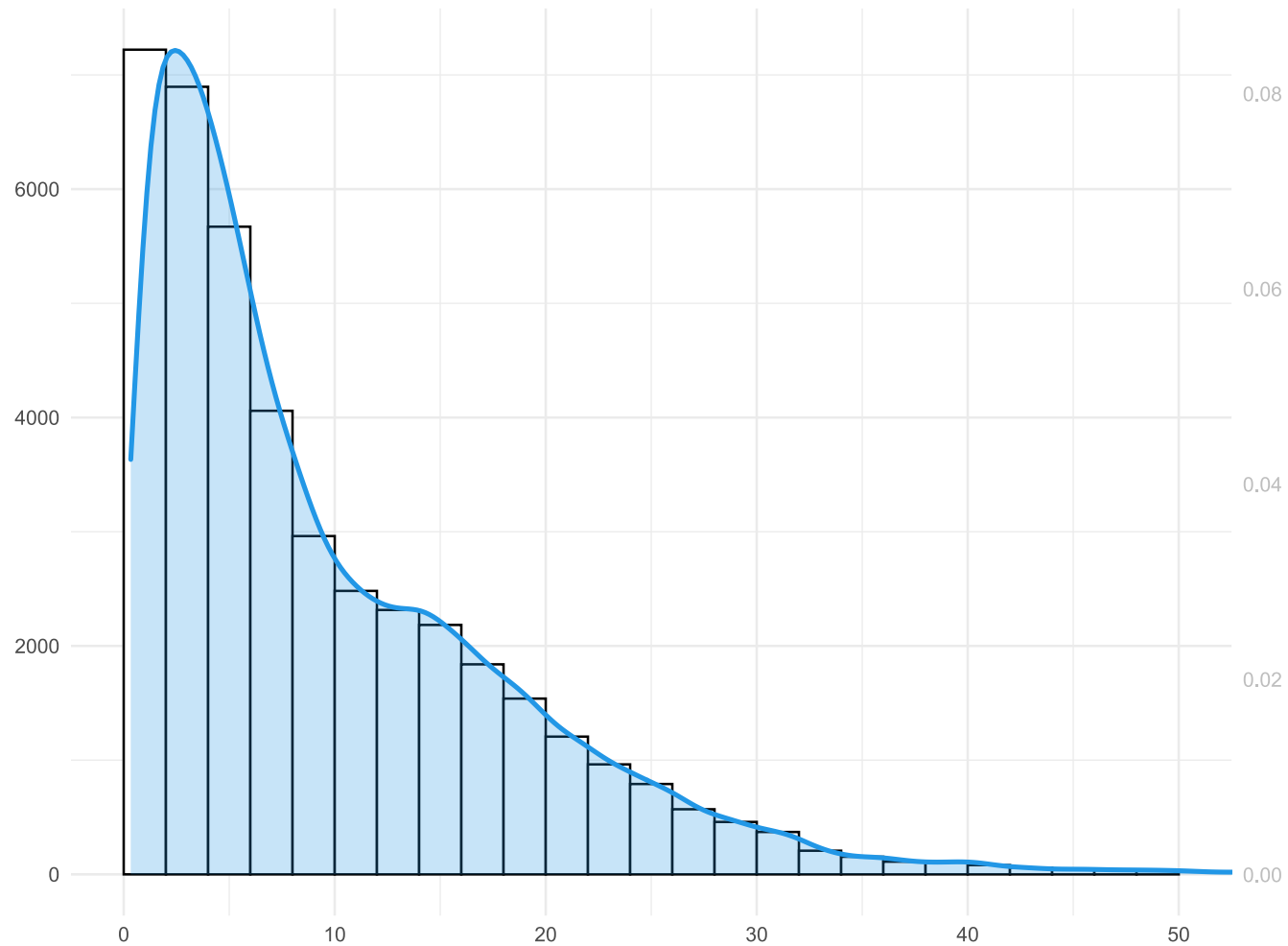
Quelle: <https://fachprofile.uni-wuppertal.de> (Heinze et al. 2019)

Abbildung 2b: Fächerprofil der Universität Bamberg



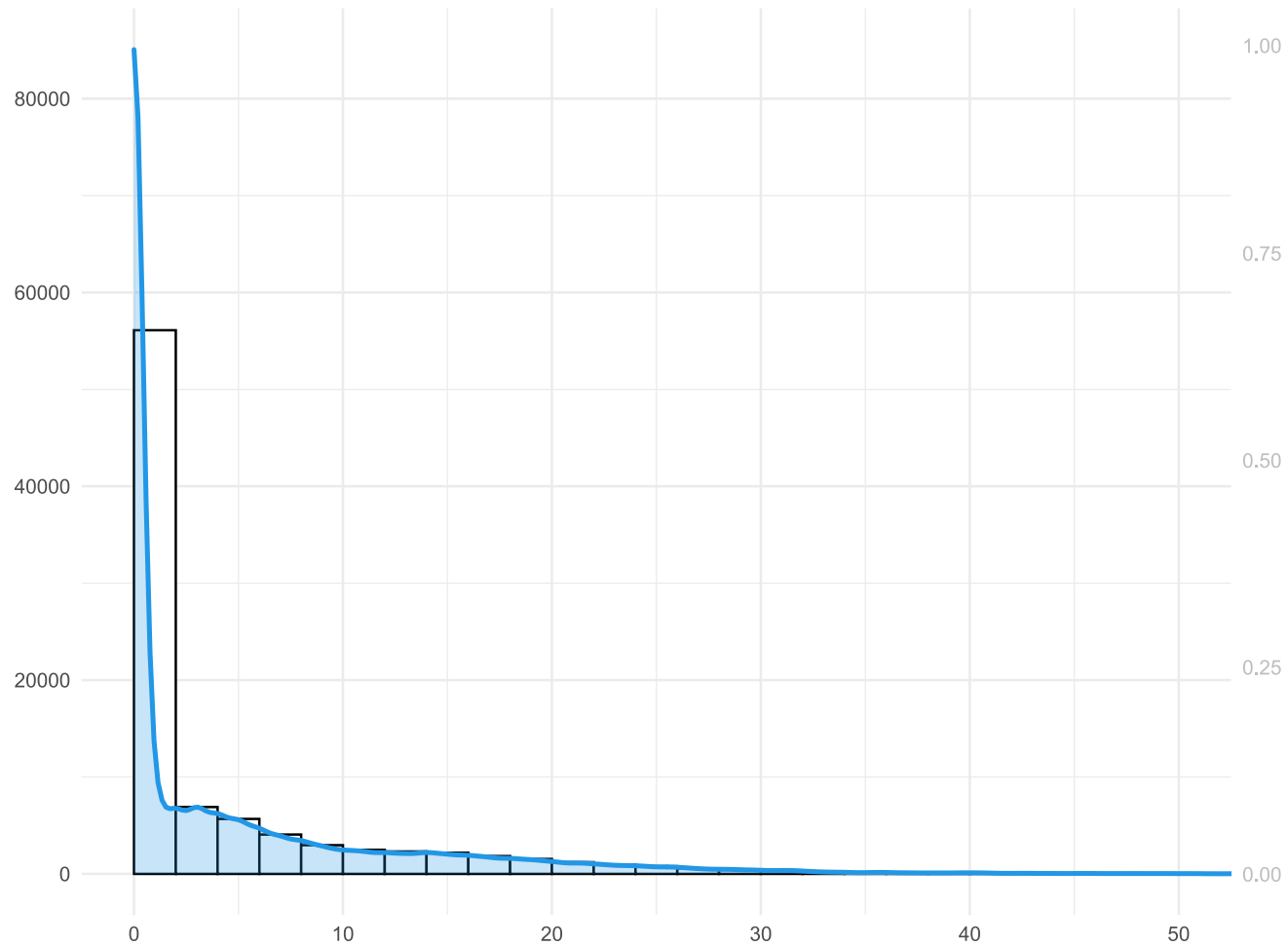
Quelle: <https://fachprofile.uni-wuppertal.de> (Heinze et al. 2019)

Abbildung 3: Histogramm und Dichte der Variable „Professor:innen“ (ohne Nullwerte)



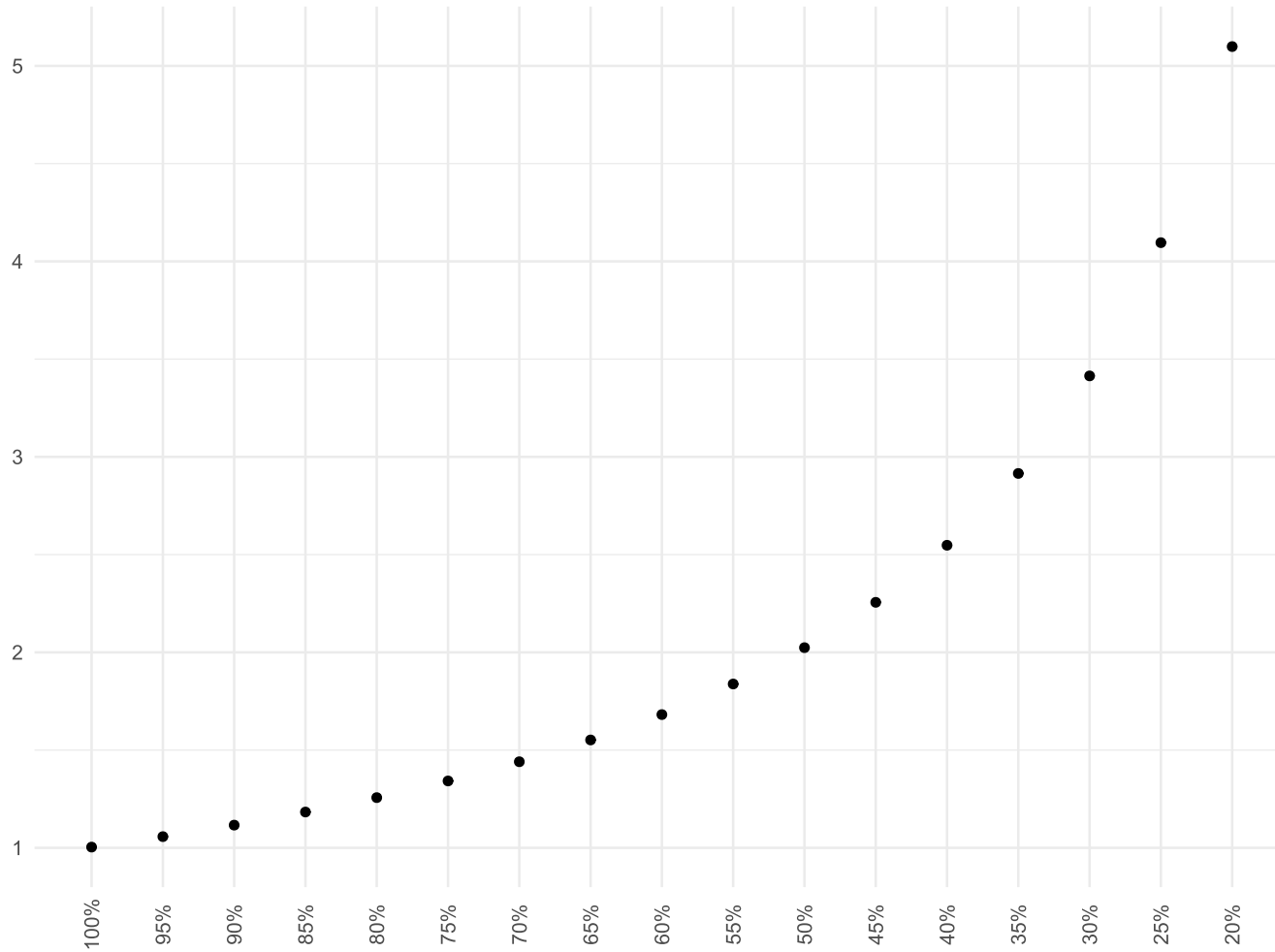
Die Abbildung umfasst keine Fächer mit Null Professor:innen oder fehlenden Werten. Die linke Y-Achse zeigt die Werte des Histogramms (inklusive rechter Seite, d.h. Profs ≤ 2 für die erste Säule). Die Darstellung reicht bis 50 Professor:innen pro Fach und Universität. In blau ist der zugehörige Kerndichteschätzer dargestellt (Werte des Dichteschätzers beziehen sich auf rechte Y-Achse in grau).

Abbildung 4: Histogramm und Dichte der Variable „Professor:innen“ (mit Nullwerten)



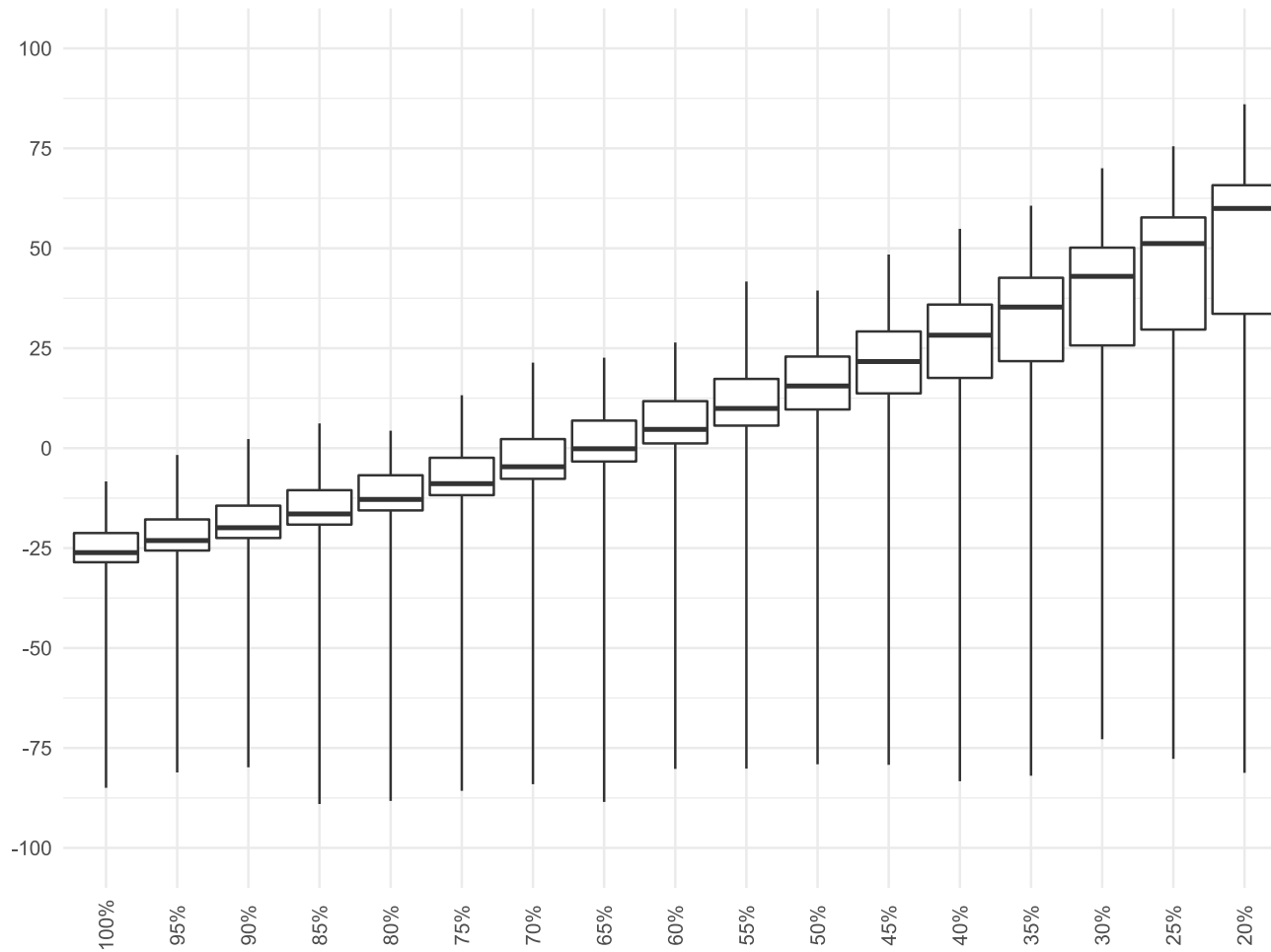
Die Abbildung umfasst auch Fächer mit Null Professor:innen oder fehlenden Werten, wobei letztere durch Null ersetzt wurden. Die linke Y-Achse zeigt die Werte des Histogramms (inklusive rechter Seite, d.h. Profs ≤ 2 für die erste Säule). Die Darstellung reicht bis 50 Professor:innen pro Fach und Universität. In blau ist der zugehörige Kerndichteschätzer dargestellt (Werte des Dichte-Schätzers beziehen sich auf rechte Y-Achse in grau).

Abbildung 5: AI-Erwartungswerte gemäß Monte-Carlo Simulationen



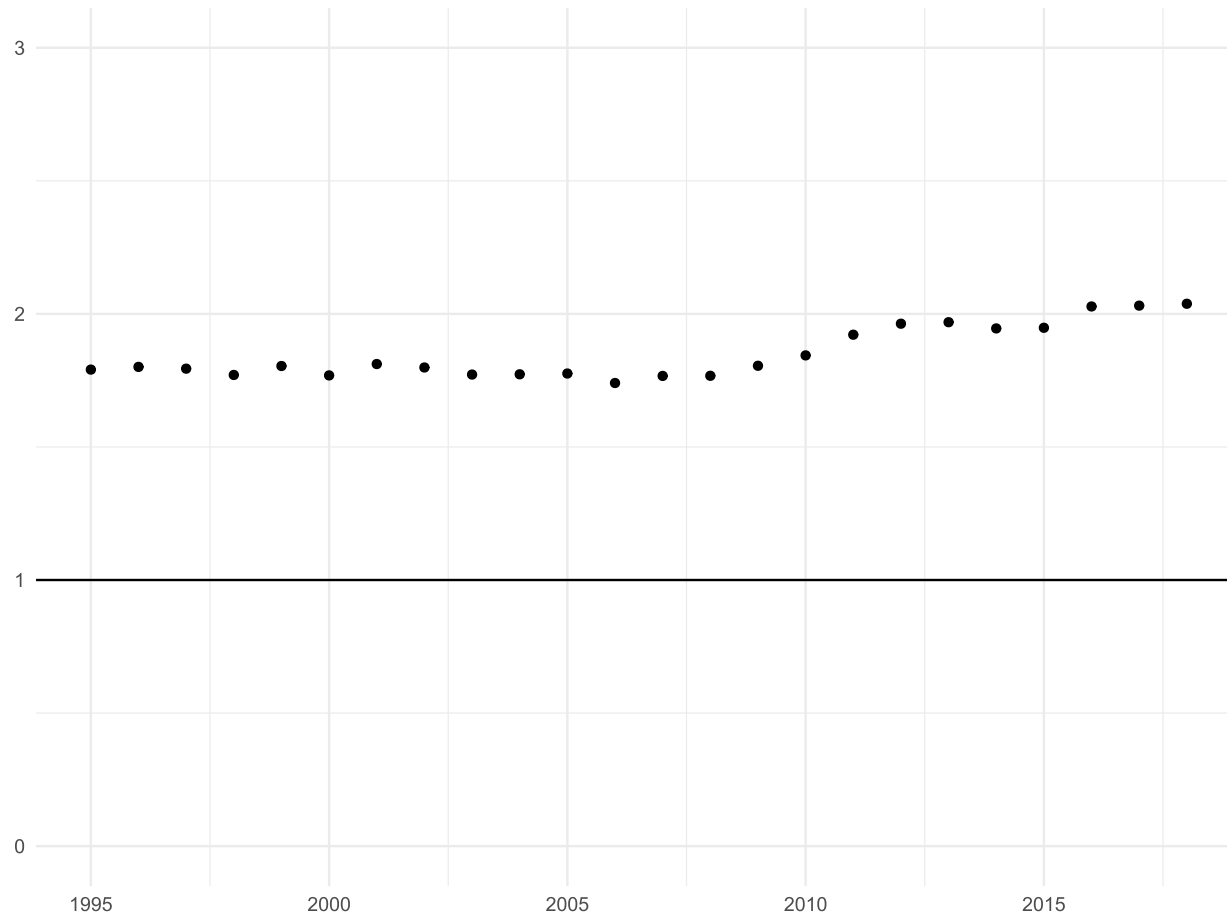
Die X-Achse zeigt die Abdeckungsquote der simulierten Daten, die Y-Achse den dazugehörigen Erwartungswert des AI. Es gilt, $\text{Abdeckungsquote} = 100\% - (\text{Beobachtungen mit fehlenden Werten in Prozent})$. Jeder Punkt entspricht dem Durchschnitt von AI Mittelwerten aus 180.000 simulierten Datensets.

Abbildung 6: RESP-Erwartungswerte gemäß Monte-Carlo Simulationen



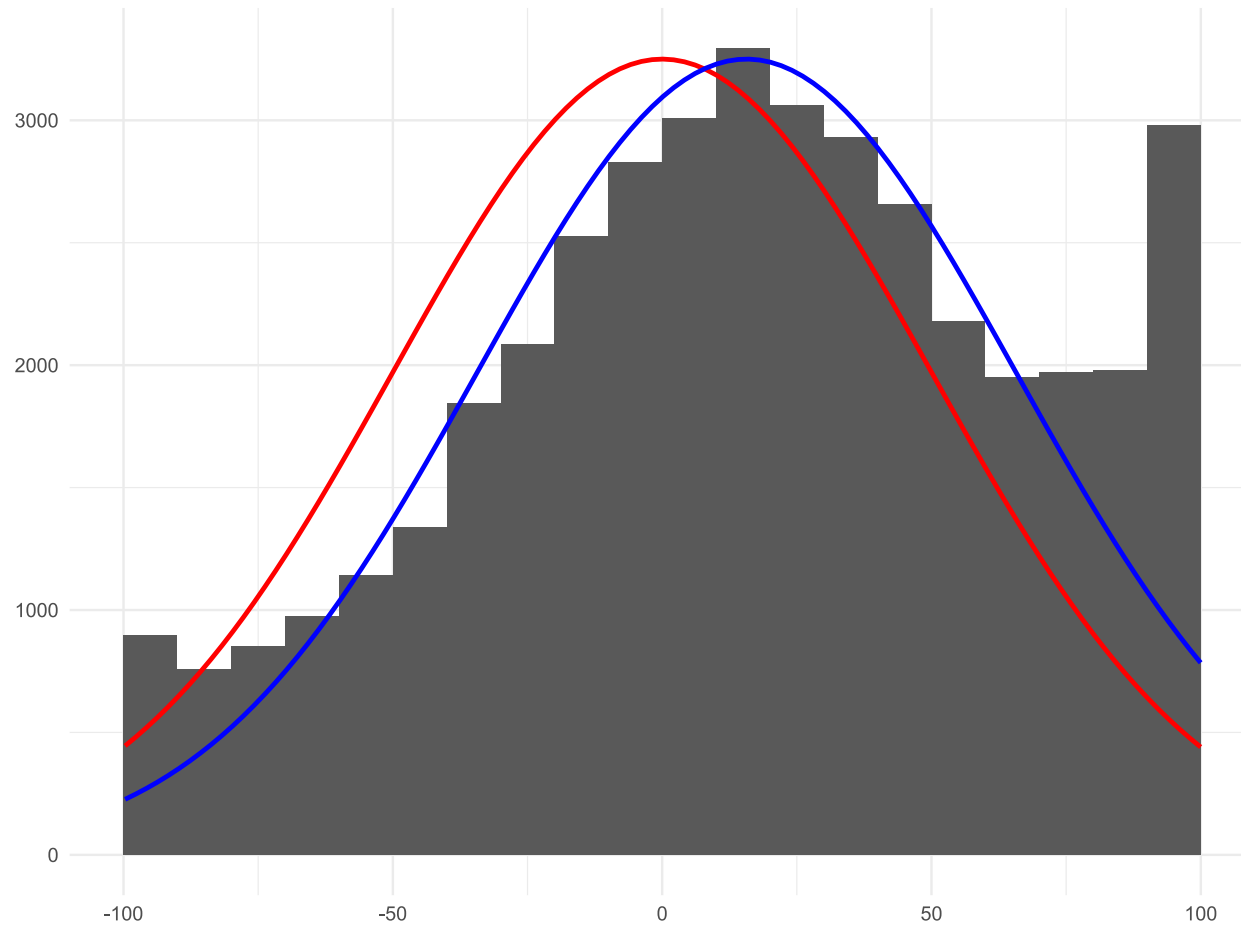
Die X-Achse zeigt die Abdeckungsquote der simulierten Daten, die Y-Achse die durchschnittlichen RESP Wertes eines Datensets. Es gilt, Abdeckungsquote = 100% - (Beobachtungen mit fehlenden Werten in Prozent). Jeder Boxplot entspricht den RESP-Durchschnittswerten aus 180.000 simulierten Datensets. Die horizontalen Balken stellen die Mediane da. Die Box um jeden Balken repräsentiert das 5% und das 95% Perzentil. Die Linien ober- und unterhalb der Balken zeigen den Wertebereich auf, also den maximalen und minimalen durchschnittlichen RESP-Wert.

Abbildung 7: Durchschnittliche AI-Erwartungswerte gemäß Hochschuldatensatz



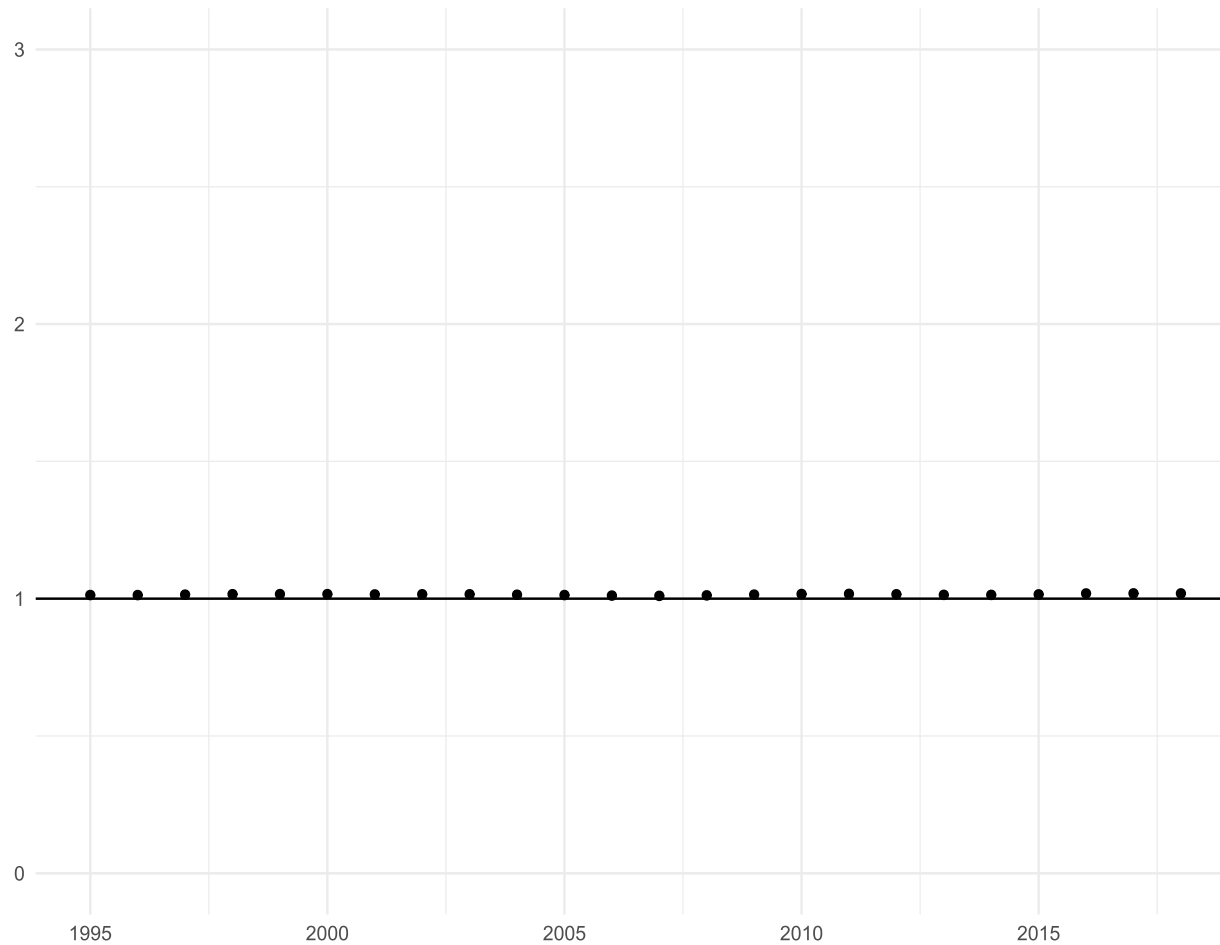
Verwendet wurde der Ursprungsdatensatz mit 68 Universitäten (51 NTUs und 17 TUs) und 56 Fächer nach der Systematik des StBA. Die X-Achse zeigt die Jahre 1995-2018, die Y-Achse den durchschnittlichen AI Wert. Der AI wurde für NTUs und TUs getrennt berechnet, dann aber gepoolt, bevor der durchschnittliche Jahreswert berechnet und als Scatterplot dargestellt wurde (eine getrennte Auswertung zeigte keine signifikante Abweichung der beiden Gruppen). Als schwarze Linie: der (theoretische) Erwartungswert des AI (= 1).

Abbildung 8: Histogramm der RESP-Werte gemäß Hochschuldatensatz



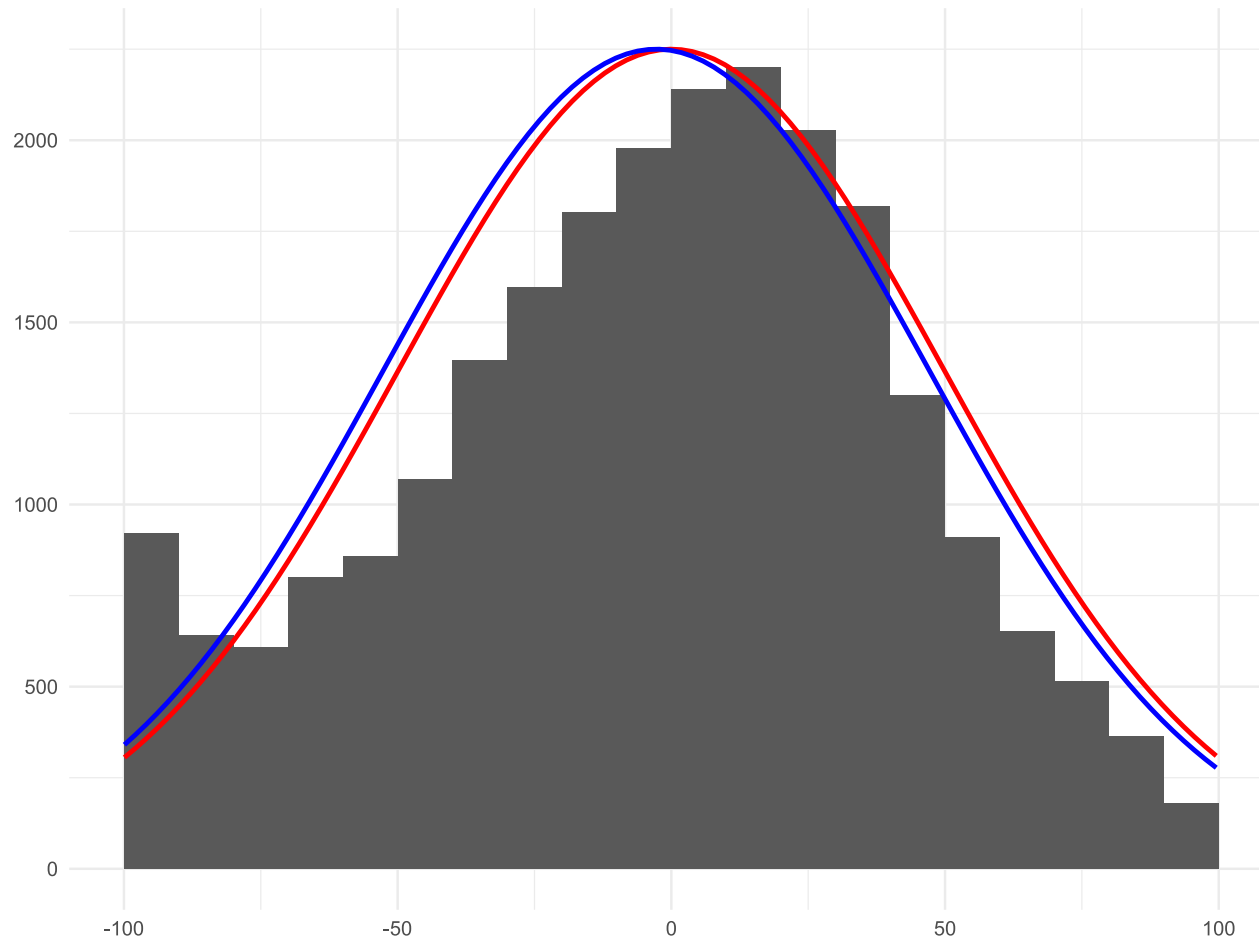
Verwendet wurde der Ursprungsdatensatz mit 68 Universitäten (51 NTUs und 17 TUs) und 56 Fächer nach der Systematik des StBA der Jahre 1995-2018. Die RESP-Werte sind für technische und nicht-technische Universitäten gepoolt dargestellt. Die rote Linie stellt die (zu erwartende) Normalverteilung $(407.327,1 \cdot N(0;2.500))$ dar, während die um den durchschnittlichen RESP-Wert nach rechts verschobene Normalverteilung $(407.327,1 \cdot N(15,67242;2.500))$ in blau aufgetragen ist.

Abbildung 9: Durchschnittlicher AI-Wert pro Jahr bezüglich der beiden neuen Fächersystematiken



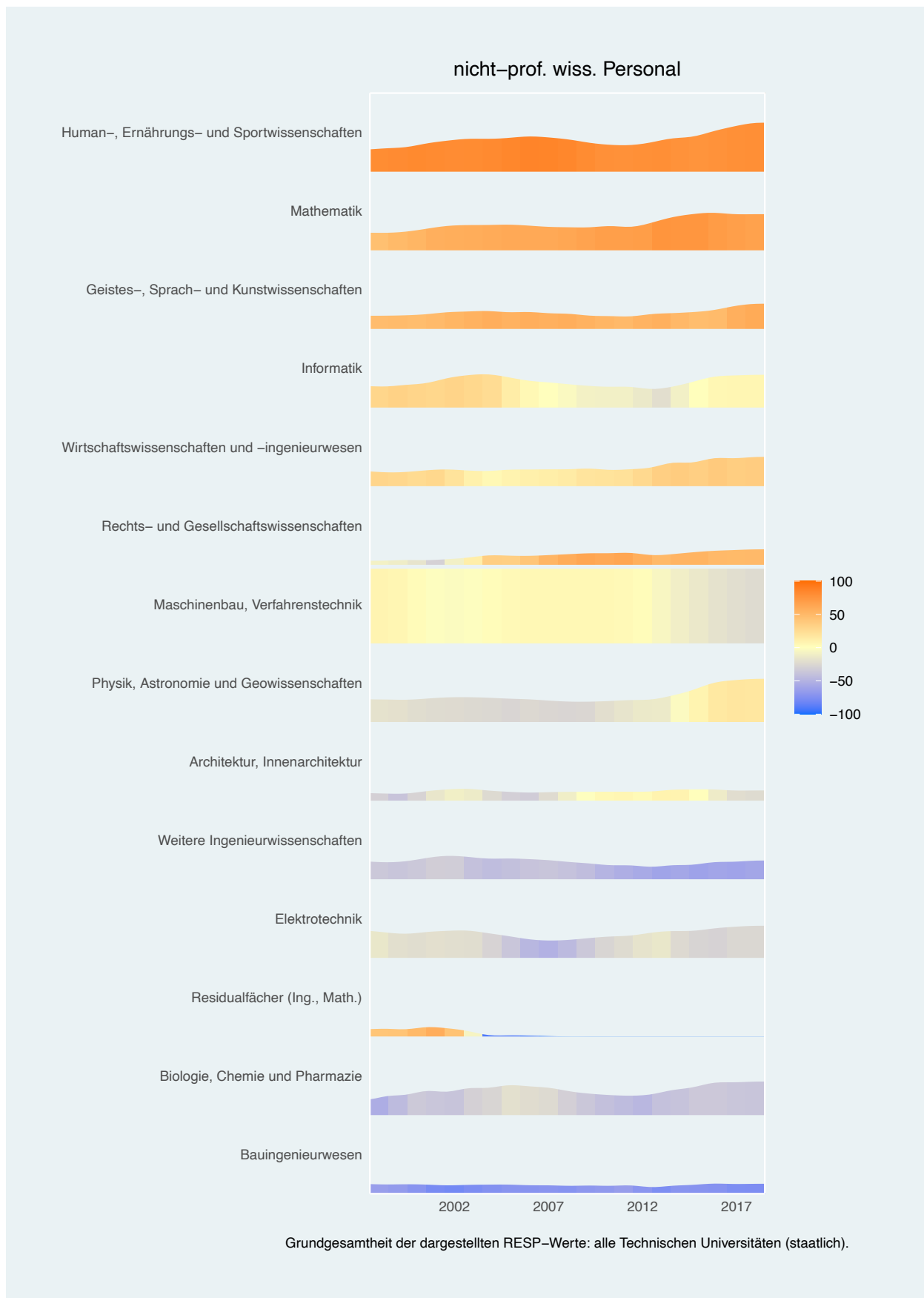
Verwendet wurde die neue Fächersystematik mit 68 Universitäten (51 NTUs und 17 TUs) und 17 Fächern (NTU) bzw. 14 Fächern (TU). Die X-Achse zeigt die Jahre 1995-2018, die Y-Achse den durchschnittlichen AI Wert. Der AI wurde für NTUs und TUs getrennt berechnet, dann aber gepoolt, bevor der durchschnittliche Jahreswert berechnet und als Scatterplot dargestellt wurde (eine getrennte Auswertung zeigte keine signifikante Abweichung der beiden Gruppen). Als schwarze Linie: der (theoretische) Erwartungswert des AI (= 1).

Abbildung 10: Histogramm der RESP-Werte gemäß Hochschuldatensatz mit neuer Fächersystematik



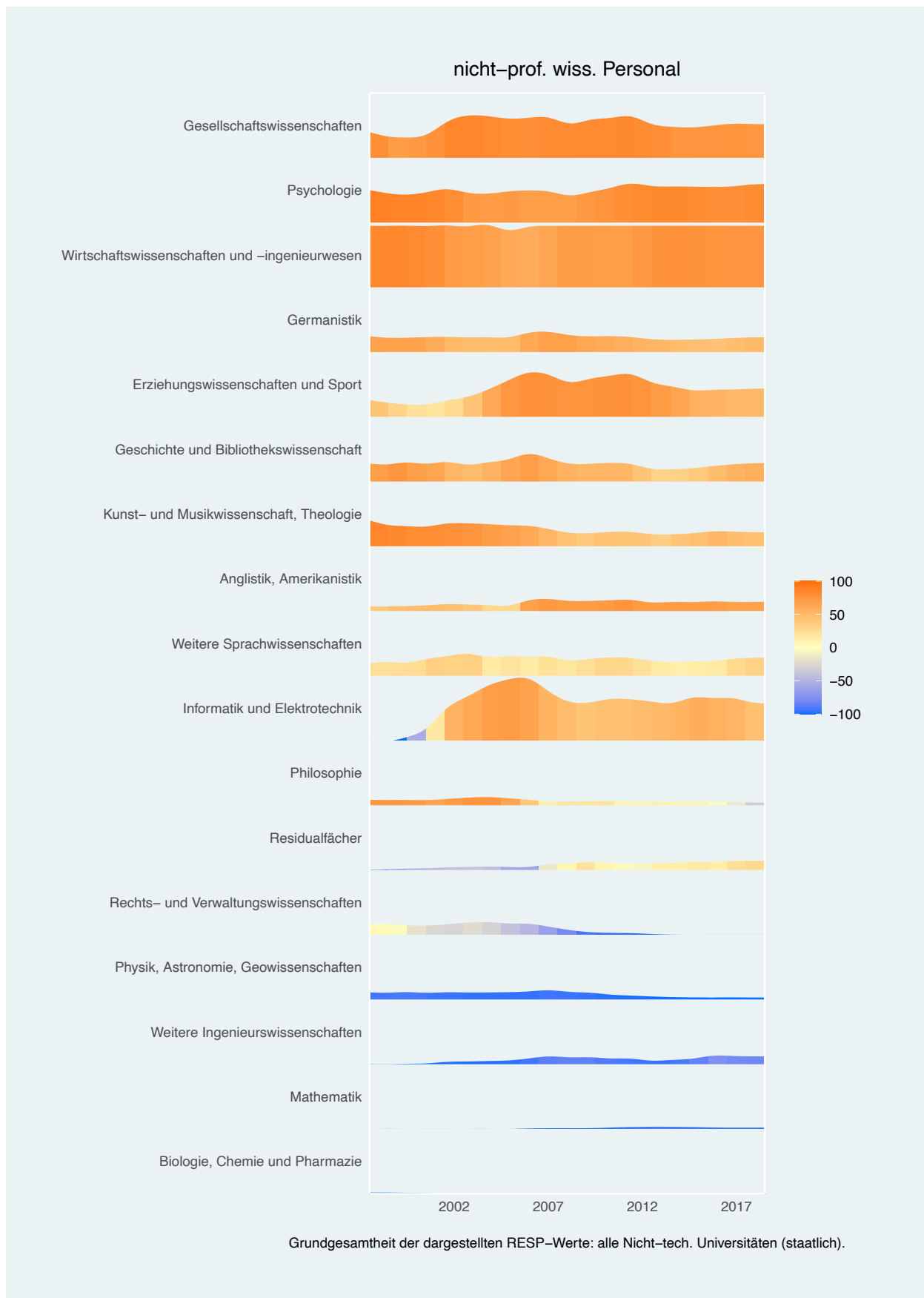
Verwendet wurde die neue Fächersystematik mit 68 Universitäten (51 NTUs und 17 TUs) und 17 Fächern (NTU) bzw. 14 Fächern (TU) für die Jahre 1995-2018. Die RESP-Werte sind für technische und nicht-technische Universitäten gepoolt dargestellt. Die rote Linie stellt die Normalverteilung $(281.995,7 \cdot N(0;2.500))$, während die um den durchschnittlichen RESP-Wert nach rechts verschobene Normalverteilung in $(281.995,7 \cdot N(-2,322014;2.500))$ in blau aufgetragen ist.

Abbildung 11a: Neues Fächerprofil der Technischen Universität Dortmund



Quelle: Eigene Erstellung

Abbildung 11b: Neues Fächerprofil der Universität Bamberg



Quelle: Eigene Erstellung

Tabelle 1: Kennzahlen zur Monte-Carlo Simulation einer Log-normal verteilten Variable mit Datenausfällen

Anzahl Fächer ($i \in I$)	15, 25, 50, 100
Anzahl Institutionen ($j \in J$)	10, 50, 100, 150, 300
Untere Grenze für Professuren ($\min(v)$)	0
Obere Grenze für Professuren ($\max(v)$)	35, 60, 85, 110, 135, 160, 185, 210, 235
Verteilung von v	Standard Log-normal Verteilung skaliert auf $[\min(v), \max(v)]$
Anteil an fehlenden Werten ($v = NA$)	0% bis 80% in 5% Schritten
Verteilung an fehlenden Werten	Gleichverteilung
Wiederholung einer jeden Konfiguration mit Zufallszahlen	1.000
Anzahl an simulierten Datensets	3.060.000

Tabelle 2: Kennzahlen des untersuchten Hochschuldatensatzes am Beispiel der Variable „Professor:innen“

Hochschulart	Universitäten	NTUs	TUs
Anzahl Lehr- und Forschungsbereiche (LuF)	56 (ohne Medizin)		
Anzahl Institutionen	68	51	17
Anzahl Jahre	24 (1995-2018)		
Mögliche Beobachtungen	91.392	68.544	22.848
0 Professoren in LuFs an Institutionen	4.091	3.076	1.015
Anzahl fehlende Werte für Professor:innen	46.058	32.778	13.280
Anzahl existierender Werte für Professor:innen	45.334	35.766	9.568
Anzahl existierender Werte für Professor:innen > 0	41.243	32.690	8.553
Abdeckungsquote (AQ1)	45,1%	47,7%	37,4%

Tabelle 3: Abdeckungsquoten für die TUs und den 5-Jahreszyklus 2014-2018

Lehr- und Forschungsgebiet	AQ	Lehr- und Forschungsgebiet	AQ
Mathematik	100%	Sprach- und Kulturwissenschaften allgemein	32%
Informatik	100%	Geographie	29%
Maschinenbau, Verfahrenstechnik	100%	Rechts-, Wirtschafts- und Gesellschaftswissenschaften allgemein	29%
Chemie	94%	Evangelische Theologie	25%
Elektrotechnik	94%	Romanistik	24%
Physik, Astronomie	91%	Musik, Musikwissenschaft	24%
Wirtschaftswissenschaften	87%	Allgemeine und vergleichende Literaturwissenschaft	24%
Bauingenieurwesen	76%	Sozialwesen	24%
Sozialwissenschaften	73%	Landespflege, Umweltgestaltung	21%
Erziehungswissenschaften	72%	Katholische Theologie	18%
Architektur, Innenarchitektur	71%	Agrarwissenschaften, Lebensmittel- und Getränketechnologie	15%
Politikwissenschaften	62%	Wirtschaftsingenieurwesen mit wirt.-wiss. Schwerpunkt	15%
Philosophie	60%	Wirtschaftsingenieurwesen	15%
Geschichte	60%	Mathematik, Naturwissenschaften allgemein	13%
Biologie	60%	Bergbau, Hüttenwesen	13%
Geowissenschaften ohne Geographie	55%	Bildende Kunst	12%
Raumplanung	54%	Forstwirtschaft, Holzwirtschaft	12%
Sport, Sportwissenschaften	53%	Ernährungs- und Haushaltswissenschaften	11%
Psychologie	51%	Slawistik, Baltistik, Finno-Ugristik	6%
Verkehrstechnik, Nautik	49%	Altphilologie	6%
Germanistik	47%	Kulturwissenschaften i.e.S.	6%
Ingenieurwissenschaften allgemein	47%	Pharmazie	6%
Rechtswissenschaften	46%	Gestaltung	6%
Anglistik, Amerikanistik	41%	Regionalwissenschaften	6%
Vermessungswesen	41%	Agrar-, Forst- und Ernährungswissenschaften allgemein	5%
Kunst, Kunstwissenschaft allgemein	39%	Sonderpädagogik	2%

Hinweis: Sortierung der Lehr- und Forschungsgebiete erfolgt absteigend nach Abdeckungsquote (AQ).

Tabelle 4: Neue Fächersystematik für TUs und zugehöriges Fächeraggregationsschema

Neues Fach/ neu konstruiertes Fach	Enthaltene Lehr- und Forschungsgebiete
Architektur, Innenarchitektur	Architektur, Innenarchitektur
Bauingenieurwesen	Bauingenieurwesen
Elektrotechnik	Elektrotechnik
Informatik	Informatik
Maschinenbau, Verfahrenstechnik	Maschinenbau, Verfahrenstechnik
Weitere Ingenieurwissenschaften	Agrar-, Forst- und Ernährungswissenschaften
	Agrarwissenschaften, Lebensmittel- und Getränketechnologie
	Bergbau, Hüttenwesen
	Forstwirtschaft, Holzwirtschaft
	Landespflege, Umweltgestaltung
	Raumplanung
	Verkehrstechnik, Nautik
	Vermessungswesen
	Wirtschaftsingenieurwesen mit ing.-wiss. Schwerpunkt
Biologie, Chemie und Pharmazie	Biologie
	Chemie
	Pharmazie
Mathematik	Mathematik
Physik, Astronomie und Geowissenschaften	Geographie
	Geowissenschaften ohne Geographie
	Physik, Astronomie
Wirtschaftswiss. und -ingenieurwesen	Wirtschaftsingenieurwesen mit wirt.-wiss. Schwerpunkt
Rechts- und Gesellschaftswissenschaften	Politikwissenschaften
	Rechts-, Wirtschafts- und Gesellschaftswissenschaften
	Rechtswissenschaften
	Regionalwissenschaften
	Sozialwissenschaften
	Verwaltungswissenschaften.

Tabelle 4: Neue Fächersystematik für TUs und zugehöriges Fächeraggregationsschema (Fortsetzung)

Neues Fach/ neu konstruiertes Fach	Enthaltene Lehr- und Forschungsgebiete
Geistes-, Sprach- und Kunstwissenschaften	Allgemeine und vergleichende Literaturwissenschaft
	Altphilologie
	Anglistik, Amerikanistik
	Außereuropäische Sprach- und
	Bibliothekswissenschaft, Dokumentation, Publizistik
	Bildende Kunst
	Darstellende Kunst, Film und Fernsehen
	Evangelische Theologie
	Germanistik
	Geschichte
	Gestaltung
	Katholische Theologie
	Kunst, Kunstwissenschaft allgemein
	Musik, Musikwissenschaft
	Philosophie
Romanistik	
Slawistik, Baltistik, Finno-Ugristik	
Sprach- und Kulturwissenschaften	
Human-, Erziehungs- und Sportwissenschaften	Ernährungs- und Haushaltswissenschaften
	Erziehungswissenschaften
	Kulturwissenschaften i.e.S.
	Psychologie
	Sonderpädagogik
	Sozialwesen
	Sport, Sportwissenschaften
Residualfächer (Ing.-Wiss., Math.)	Ingenieurwissenschaften allgemein
	Mathematik, Naturwissenschaften allgemein
	Wirtschaftswissenschaften

Tabelle 5: Abdeckungsquoten für die neue TU-Fächersystematik

	1994-1998	1999-2003	2004-2008	2009-2013	2014-2018
Architektur, Innenarchitektur	73%	76%	73%	71%	71%
Bauingenieurwesen	76%	76%	76%	76%	76%
Elektrotechnik	98%	99%	100%	94%	94%
Informatik	91%	96%	100%	100%	100%
Maschinenbau, Verfahrenstechnik	100%	100%	100%	100%	100%
Weitere Ingenieurwissenschaften	86%	89%	88%	91%	88%
Biologie, Chemie und Pharmazie	93%	94%	94%	94%	94%
Mathematik	92%	98%	99%	95%	100%
Physik, Astronomie und Geowissenschaften	93%	94%	94%	94%	91%
Wirtschaftswissenschaften und -ingenieurwesen	93%	89%	95%	96%	94%
Geistes-, Sprach- und Kunstwissenschaften	71%	74%	76%	76%	82%
Rechts- und Gesellschaftswissenschaften	80%	81%	82%	76%	80%
Human-, Erziehungs- und Sportwissenschaften	65%	65%	71%	76%	78%
Residualfächer (Ing.-Wiss., Math.)	46%	42%	51%	58%	58%

Tabelle 6: Abdeckungsquoten für die TUs und den 5-Jahreszyklus 2014-2018

Lehr- und Forschungsgebiet	AQ	Lehr- und Forschungsgebiet	AQ
Wirtschaftswissenschaften	99%	Geowissenschaften ohne Geographie	49%
Mathematik	96%	Pharmazie	40%
Psychologie	96%	Rechts-, Wirtschafts- und Gesellschaftswissenschaften allgemein	35%
Erziehungswissenschaften	95%	Maschinenbau, Verfahrenstechnik	31%
Germanistik	94%	Elektrotechnik	27%
Philosophie	94%	Mathematik, Naturwissenschaften	25%
Anglistik, Amerikanistik	94%	Sozialwesen	20%
Geschichte	92%	Darstellende Kunst, Film und Fernsehen	20%
Rechtswissenschaften	92%	Ingenieurwissenschaften allgemein	19%
Sozialwissenschaften	89%	Agrarwissenschaften, Lebensmittel- und Getränketechnologie	19%
Informatik	88%	Bauingenieurwesen	14%
Chemie	86%	Ernährungs- und Haushaltswissenschaften	14%
Physik, Astronomie	86%	Landespflege, Umweltgestaltung	11%
Romanistik	84%	Gestaltung	11%
Biologie	84%	Bildende Kunst	9%
Kunst, Kunstwissenschaft allgemein	76%	Verwaltungswissenschaften.	8%
Evangelische Theologie	73%	Architektur, Innenarchitektur	8%
Politikwissenschaften	72%	Sonderpädagogik	8%
Sport, Sportwissenschaften	71%	Bibliothekswissenschaft, Dokumentation, Publizistik	6%
Musik, Musikwissenschaft	71%	Forstwirtschaft, Holzwirtschaft	6%
Geographie	71%	Regionalwissenschaften	6%
Sprach- und Kulturwissenschaft	63%	Agrar-, Forst- und Ernährungswissenschaften	5%
Allgemeine und vergleichende Literaturwissenschaft	63%	Raumplanung	5%
Außereuropäische Sprach- und Kulturwissenschaften	62%	Verkehrstechnik, Nautik	4%
Slawistik, Baltistik, Finno-Ugristik	58%	Wirtschaftsingenieurwesen mit wirt.-wiss. Schwerpunkt	4%
Altphilologie	57%	Vermessungswesen	3%
Katholische Theologie	53%	Wirtschaftsingenieurwesen	2%
Kulturwissenschaften i.e.S.	51%	Bergbau, Hüttenwesen	2%

Hinweis: Sortierung der Lehr- und Forschungsgebiete erfolgt absteigend nach Abdeckungsquote (AQ).

Tabelle 7: Neue Fächersystematik für NTUs und zugehöriges Fächeraggregationsschema

Neues Fach/ neu konstruiertes Fach	Enthaltene Lehr- und Forschungsgebiete
Anglistik, Amerikanistik	Anglistik, Amerikanistik
Germanistik	Germanistik
Weitere Sprachwissenschaften	Allgemeine und vergleichende Literaturwissenschaft
	Altphilologie
	Außereuropäische Sprach- und Kulturwissenschaften
	Romanistik
	Slawistik, Baltistik, Finno-Ugristik
Geschichte und Bibliothekswissenschaft	Bibliothekswissenschaft, Dokumentation, Publizistik
Kunst- und Musikwissenschaft, Theologie	Bildende Kunst
	Darstellende Kunst, Film und Fernsehen
	Evangelische Theologie
	Gestaltung
	Katholische Theologie
	Kunst, Kunstwissenschaft allgemein
	Musik, Musikwissenschaft
	Geschichte
Philosophie	Philosophie
Gesellschaftswissenschaften	Kulturwissenschaften i.e.S.
	Politikwissenschaften
	Regionalwissenschaften
	Sozialwissenschaften
Psychologie	Psychologie
Rechts- und Verwaltungswissenschaften	Rechtswissenschaften
	Verwaltungswissenschaften
Erziehungswissenschaften und Sport	Erziehungswissenschaften
	Sonderpädagogik
	Sozialwesen
	Sport, Sportwissenschaften
Mathematik	Mathematik

Tabelle 7: Neue Fächersystematik für NTUs und zugehöriges Fächeraggregationsschema (Fortsetzung)

Neues Fach/ neu konstruiertes Fach	Enthaltene Lehr- und Forschungsgebiete
Biologie, Chemie und Pharmazie	Biologie
	Chemie
	Pharmazie
Physik, Astronomie, Geowissenschaften	Geographie
	Geowissenschaften ohne Geographie
	Physik, Astronomie
Informatik und Elektrotechnik	Elektrotechnik
	Informatik
Weitere Ingenieurwissenschaften	Agrar-, Forst- und Ernährungswissenschaften
	Agrarwissenschaften, Lebensmittel- und Getränketechnologie
	Architektur, Innenarchitektur
	Bauingenieurwesen
	Bergbau, Hüttenwesen
	Ernährungs- und Haushaltswissenschaften
	Forstwirtschaft, Holzwirtschaft
	Ingenieurwissenschaften allgemein
	Landespflege, Umweltgestaltung
	Maschinenbau, Verfahrenstechnik
	Raumplanung
	Verkehrstechnik, Nautik
	Vermessungswesen
Wirtschaftsingenieurwesen mit wirt.-wiss. Schwerpunkt	
Wirtschaftswissenschaften und -ingenieurwesen	Wirtschaftsingenieurwesen
	Wirtschaftswissenschaften
Residualfächer	Mathematik, Naturwissenschaften allgemein
	Rechts-, Wirtschafts- und Gesellschaftswissenschaften allgemein
	Sprach- und Kulturwissenschaft

Tabelle 8: Abdeckungsquoten für die neue NTU-Fächersystematik

	1994-1998	1999-2003	2004-2008	2009-2013	2014-2018
Anglistik, Amerikanistik	89%	93%	93%	94%	94%
Germanistik	90%	93%	93%	94%	94%
Weitere Sprachwissenschaften	94%	93%	93%	94%	94%
Geschichte und Bibliothekswissenschaft	95%	96%	97%	94%	92%
Kunst- und Musikwissenschaft, Theologie	93%	94%	94%	93%	90%
Philosophie	93%	93%	94%	95%	94%
Gesellschaftswissenschaften	96%	96%	98%	97%	96%
Psychologie	92%	93%	95%	95%	96%
Rechts- und Verwaltungswissenschaften	86%	91%	96%	94%	92%
Erziehungswissenschaften und Sport	96%	96%	96%	95%	95%
Mathematik	92%	92%	96%	96%	96%
Biologie, Chemie und Pharmazie	91%	86%	86%	88%	88%
Physik, Astronomie, Geowissenschaften	100%	98%	97%	93%	92%
Informatik und Elektrotechnik	76%	81%	89%	91%	90%
Weitere Ingenieurwissenschaften	59%	57%	57%	58%	62%
Wirtschaftswissenschaften und -ingenieurwesen	97%	100%	98%	98%	99%
Residualfächer	41%	42%	57%	82%	84%

Anhänge

Anhang 1: Liste der technischen Universitäten (alphabetisch)

Brandenburgische Technische Universität Cottbus-Senftenberg
Karlsruher Institut für Technologie
Leibniz Universität Hannover
Rheinisch-Westfälische Technische Hochschule Aachen
Technische Universität Bergakademie Freiberg
Technische Universität Berlin
Technische Universität Braunschweig
Technische Universität Chemnitz
Technische Universität Clausthal
Technische Universität Darmstadt
Technische Universität Dortmund
Technische Universität Dresden
Technische Universität Hamburg
Technische Universität Ilmenau
Technische Universität Kaiserslautern
Technische Universität München
Universität Stuttgart

Anhang 2: Liste der nicht-technischen Universitäten (alphabetisch)

Albert-Ludwigs-Universität Freiburg
Bergische Universität Wuppertal
Carl von Ossietzky Universität Oldenburg
Christian-Albrechts-Universität zu Kiel
Eberhard-Karls-Universität Tübingen
Europa-Universität Viadrina Frankfurt (Oder)
Freie Universität Berlin
Friedrich-Alexander-Universität Erlangen-Nürnberg
Friedrich-Schiller-Universität Jena
Georg-August-Universität Göttingen
Goethe-Universität Frankfurt am Main
Heinrich-Heine-Universität Düsseldorf
Humboldt-Universität zu Berlin
Johannes-Gutenberg-Universität Mainz
Julius-Maximilians-Universität Würzburg
Justus-Liebig-Universität Gießen
Leuphana Universität Lüneburg
Ludwig-Maximilians-Universität München
Martin-Luther-Universität Halle-Wittenberg
Otto-Friedrich-Universität Bamberg
Otto-von-Guericke-Universität Magdeburg
Philipps-Universität Marburg

Anhang 2: Liste der nicht-technischen Universitäten (Fortsetzung)

Rheinische Friedrich-Wilhelms-Universität Bonn
Ruhr-Universität Bochum
Ruprecht-Karls-Universität Heidelberg
Universität Augsburg
Universität Bayreuth
Universität Bielefeld
Universität Bremen
Universität des Saarlandes
Universität Duisburg-Essen
Universität Erfurt
Universität Greifswald
Universität Hamburg
Universität Hohenheim
Universität Kassel
Universität Koblenz-Landau
Universität Konstanz
Universität Leipzig
Universität Mannheim
Universität Osnabrück
Universität Paderborn
Universität Passau
Universität Potsdam
Universität Regensburg
Universität Rostock
Universität Siegen
Universität Trier
Universität Ulm
Universität zu Köln
Westfälische Wilhelms-Universität Münster