

No. 33
Mai / 2015

Die Macht der Wissenschaftsstatistik und die Entwicklung der Ökonomie

Matthias Aistleitner, Marianne Fölker und Jakob Kapeller

**Institute for Comprehensive
Analysis of Economy**

**Institut für
die Gesamtanalyse der
Wirtschaft**

**Johannes Kepler
Universität Linz**

Altenbergerstraße 69
4040 Linz
Austria

Tel.: +49 732 2468 3402

csc@jku.at
www.icae.at

Die Macht der Wissenschaftsstatistik und die Entwicklung der Ökonomie

Matthias Aistleitner^{*}, Marianne Fölker^{*} und
Jakob Kapeller^{**}

Abstract

Zitationsmessungen und darauf basierende Indices und Ranglisten spielen eine immer wesentlichere Rolle im Wissenschaftsbetrieb, insbesondere im Rahmen der Bewertung und Evaluierung von Forschungsleistungen. Die Etablierung quantitativer Evaluationsroutinen im Wissenschaftsbetrieb steht dabei symptomatisch für einen umfassenderen sozialen Trend komplexe Sachverhalte durch ihre numerische Fassung scheinbar oder tatsächlich zu vereinfachen. In diesem Kontext versucht dieser Beitrag die Auswirkungen wissenschaftsstatistischer Evaluationsverfahren auf die weitere Wissenschaftsentwicklung zu analysieren und dabei insbesondere das Beispiel der ökonomischen Disziplin in den Blick zu nehmen. Dabei zeigt sich, dass die Macht der (evaluativen) Wissenschaftsstatistik vielfältige Auswirkungen hervorbringt: So werden akademische Reproduktionslogiken, das Verhalten einzelner AkteurInnen im Wissenschaftsbetrieb sowie die paradigmatische Entwicklung von Disziplinen durch wissenschaftsstatistische Verfahren beeinflusst.

^{*} Johannes Kepler Universität Linz, Institut für Philosophie und Wissenschaftstheorie

^{**} Johannes Kepler Universität Linz, Institut für Philosophie und Wissenschaftstheorie und Institut für die Gesamtanalyse der Wirtschaft (ICAE), jakob.kapeller@jku.at.

1. Einleitung

Die Frage nach der Macht der Wissenschaftsstatistik und ihrer Rolle für die Entwicklung einer gegebenen akademischen Disziplin ist eine aktuelle und zugleich komplexe Fragestellung. Sie ist aktuell, weil sie den nunmehr seit einigen Jahren relevanten Trend in den Blick nimmt, wissenschaftliche Leistungen und Standards in numerischer Form zu erfassen und mittels aggregierter Darstellung vergleichbar zu machen. Dies geschieht zumeist in Form von sogenannten Rankings, die angeben Hochschulen, ForscherInnen, Forschungsbeiträge und Fachzeitschriften, nach ihrer „Qualität“ zu reihen.¹ Diese Entwicklung markiert dabei auch eine Umkehrung der methodischen Orientierung innerhalb der Wissenschaftsstatistik. Letztere hat einen ihrer wesentlichen Ausgangspunkte in einer „verstehenden“ Analyse wissenschaftlicher Kommunikation mit dem Ziel die innere Logik akademischer Diskursprozesse besser nachvollziehen zu können (Solla-Price 1965, Rip und Courtial 1984). Derartige Fragestellungen wurden durch den Siegeszug einer evaluativen Szientometrie verdrängt, die zumeist versucht aus Nennungshäufigkeiten innerhalb der akademischen Literatur (im Bereich der Forschung) und Befragungen (im Bereich der Lehre) entsprechende Qualitätsmaße zu konstruieren (Mingers und Leydesdorff 2015, Adler et al. 2008).

Dieser Trend zu einer intensivierten szientometrischen Evaluation innerhalb der Wissenschaften ist dabei als Teil einer umfassenderen und historisch langfristigen Entwicklung hin zu einer evaluativ-numerischen Herangehensweise an Fragen sozialer Organisation zu sehen. Die dabei angewendete „Quantifizierung“ sozialer Sachverhalte ergab sich dabei historisch aus dem Versuch soziale Verhältnisse politisch plan- und steuerbar zu machen. Als archetypisches Beispiel gilt dabei die Entwicklung der Volkswirtschaftlichen Gesamtrechnung und der daraus abgeleiteten Größen wie das Bruttoinlandsprodukt, die auf dem Versuch beruhen die ökonomische Aktivitäten einer Nation zu „vermessen“ und dabei eine enorme soziale Wirkmächtigkeit erlangt haben. Neben politischen Verwaltungsbedürfnissen war für diese Tendenz freilich auch wissenschaftliches Erkenntnisinteresse verantwortlich. In jüngerer Vergangenheit sind auch Geschäftsinteressen maßgeblich involviert, den Trend zur Quantifizierung sozialer Sachverhalte zu verstärken.

¹ International bekannte Beispiele für Rankings im Hochschulbereich sind etwa das Times Higher Education World Reputation Ranking oder das QS World University Ranking. Im Bereich der Journal-Evaluation stechen besonders die Journal Citation Reports von Thomson Scientific heraus. Für die deutschsprachige Ökonomie hingegen ist das Ökonomen-Ranking des Handelsblatts von besonderem Interesse.

Dies gilt gerade in der Wissenschaftsevaluation, da hier internationale Konzerne wie Thomson Reuters (Web of Science), Elsevier (Scopus), Google (GoogleScholar) oder Bertelsmann (deutsches Hochschulranking) eine dominante Stellung einnehmen. Man kann dabei diese Quantifizierung sozialer Phänomene als „allgemeines soziologisches Phänomen“ verstehen, das die „Produktion und Kommunikation numerischer Information“ umfasst (vgl. Espeland und Stevens 2008, 401) und zwei Funktionen erfüllt: Numerische Information kann einzelne Dinge stellvertretend repräsentieren (z.B. „9/11“ und „Ground Zero“) oder unser Augenmerk auf die relativen Proportionen unterschiedlicher Dinge richten. Dabei hat numerische Information die doppelte Funktion unterschiedliche Objekte zuerst innerhalb einer gemeinsamen Skala zu vereinheitlichen, um sie in Folge besser unterscheiden und differenzieren zu können (vgl. Espeland und Stevens 2008)

Die Frage nach der Macht der Wissenschaftsstatistik in der Ökonomie ist aber nicht nur ein aktuelles Detail eines allgemeineren soziologischen Phänomens, sondern auch eine komplexe, eigenständige Fragestellung, deren Beantwortung die Berücksichtigung einer Reihe unterschiedlicher Perspektiven und Faktoren erfordert. Dabei sind zumindest drei Aspekte zu berücksichtigen, die im Folgenden genauer behandelt werden sollen: Erstens, sind szientometrische Routinen in der Wissenschaftsevaluation stets eingebettet in bestehende akademische Reproduktionslogiken; die Auswirkungen ihres verstärkten Einsatzes sind daher vornehmlich vor diesem speziellen Hintergrund zu verstehen, der in Kapitel 2 einleitend besprochen werden soll. Zweitens, handelt es sich bei zitationsbasierten Evaluationsverfahren um reaktive Messverfahren, da die Evaluation auf mechanischen Routinen beruht, die leicht zu antizipieren sind und damit Anreize zu strategischem Verhalten unter ForscherInnen setzen. Kapitel 3 widmet sich diesem Aspekt der Reaktivität mit speziellem Fokus auf die institutionellen Auswirkungen reaktiver Evaluationsverfahren. Drittens, ist die spezielle Situation der Ökonomie als „contested discipline“ (Lee und Elsner 2011) zu berücksichtigen: Die Ökonomie kennt als einzige Sozialwissenschaft ein dominantes theoretisches Paradigma – die Neoklassik, die im Kern der so genannten Mainstream-Ökonomie steht, weite Teile der ökonomischen Forschung zumindest heuristisch anleitet und deren Dominanz nur von einer kleinen Minorität zumeist heterodoxer ÖkonomInnen in Frage gestellt wird (Dobusch und Kapeller 2012a). Vor diesem Hintergrund versucht Kapitel 4 auf Basis der Methoden der verstehenden Wissenschaftsstatistik ein besseres Verständnis innerökonomischer Diskurse zu erlangen um die Wirkung der wissenschaftsstatistischer Verfahren in diesem vorläufig einzuschätzen .

2. Akademische Reproduktionslogiken: Von Matthäus-Effekten zur Pfadabhängigkeit

“The competition between paradigms is not the sort of battle that can be resolved by proofs.” (Kuhn 1962, 148)

Ein Standardresultat der szientometrischen Forschung ist, dass akademische Aufmerksamkeit – zumeist gemessen anhand von Zitationshäufigkeiten – hochgradig ungleich verteilt ist. Die Verteilung von Aufmerksamkeit, Einfluss und Prestige unter einer gegebenen Menge an einzelnen WissenschaftlerInnen oder wissenschaftlichen Beiträgen folgt dabei folgender Grundstruktur: Die allermeisten WissenschaftlerInnen und Beiträge erhalten keine oder nur sehr wenig Aufmerksamkeit, während einige wenige WissenschaftlerInnen und Beiträge besonders viel Aufmerksamkeit erhalten. Die Verteilung von Aufmerksamkeit, Einfluss und Prestige in der Wissenschaft hat daher formal gesehen ganz ähnliche Eigenschaften wie etwa die Vermögensverteilung oder die Attraktivität von Webseiten: sie folgt (zumindest an der Spitze) einer Potenzverteilung (engl.: „power law“), die durch die Zipfsche Verteilung oder die unter ÖkonomInnen beliebtere, äquivalente Pareto-Verteilung abgebildet werden kann. Derartige Verteilungen treten in sozialen Kontexten häufig auf; so folgt etwa die Bevölkerungszahl von Städten, die Anzahl der erhaltenen Telefonanrufe, die Zahl der in einem Text verwendeten Wörter oder auch die Verkaufszahlen von Buchtiteln grob einer solchen Potenzverteilung (vgl. Newman 2006). Eine zentrale Eigenschaft dieser Verteilungsklasse ist, dass die Elemente an der Spitze der jeweiligen Verteilung extrem selten auftreten aber einen überproportional hohen Anteil der Grundgesamtheit auf sich vereinen. Während dies etwa für die Bevölkerungsstruktur in Deutschland bedeutet, dass rund 10% der deutschen Bevölkerung in nur vier Städten wohnt, lässt sich dieser Umstand im wissenschaftlichen Diskurs so deuten, dass sich ein Großteil der innerakademischen Aufmerksamkeit auf einen kleinen Bruchteil der jeweiligen Literatur fokussiert.² Die nachstehende Abbildung 1 versucht dabei die Besonderheiten derartiger Verteilungen zu illustrieren und zeigt drei potenzverteilte Größen in absteigender Ordnung – die EinwohnerInnenzahl der 82 größten deutschen Städte, die österreichische Vermögensverteilung, sowie die Verteilung von Zitaten auf die Beiträge in der *American Economic Review* von 1981-1985³ – im Vergleich zu einer normalverteilten Größe (dem Gewicht von Neugeborenen). Hier zeigt sich nicht nur eine überraschende Ähnlichkeit der drei potenzverteilten Größen im Vergleich zum normalverteilten Gewicht der Neugeborenen,

² ÖkonomInnen sind derartige Verteilungen aus dem Kontext der Analyse von Einkommens- und Vermögensverteilungen lange bekannt und werden zumeist anhand der Pareto-Formulierung diskutiert.

³ Die diesen Artikeln zugeordneten Zitate stammen aus dem Zeitraum 1981-1990.

sondern auch, dass die Relation zwischen Median und Mittelwert und dem maximalen Wert der entsprechenden Reihe wesentlich geringer ist: Es zeigt sich also, dass jenen Beobachtungen an der Spitze besonders hohes Gewicht zukommt.

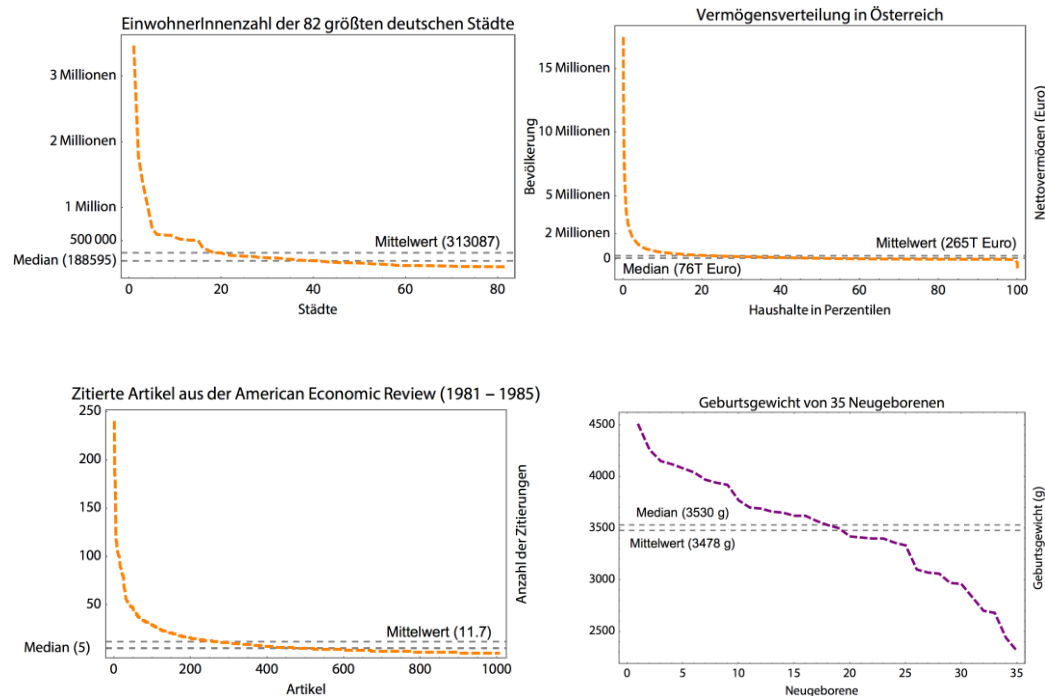


Abbildung 1: Vier Verteilungen im Vergleich (Eigene Darstellung)⁴

Das Verdienst die Aufmerksamkeitsverteilung in der wissenschaftlichen Diskurslandschaft früh erkannt und popularisiert zu haben gebührt dabei Derek J. Solla-Price (1965), der in seiner impulsgebenden Analyse von Zitationsdaten die grobe Hypothese postulierte, dass sich jeweils die nächste Generation an wissenschaftlicher Literatur schwerpunktmäßig nur auf einen kleinen Ausschnitt der vorangegangenen Literatur bezieht. Eine Schwäche dieser Hypothese ist allerdings, dass sie nur die erwartete, ungleiche Aufmerksamkeitsverteilung beschreibt, aber nicht weiter erklärt warum es überhaupt zu einem solchen Resultat kommen sollte. Es blieb damit Robert K. Merton vorbehalten diese Lücke mit seinem heute klassischen Beitrag zum „Matthäus-Effekt in den Wissenschaften“ (Merton 1968) zu füllen. Merton postulierte, dass der Zugewinn an Prestige und Aufmerksamkeit eng mit der bereits akquirierten Aufmerksamkeit korreliert und versuchte diese These mit dem Verweis auf Zitationsdaten ebenso wie mit Daten aus Interviews mit Nobelpreisträgern zu untermauern.

⁴ Datenquellen: Die Zahl der StadtbewohnerInnen wurde der Mathematica10-Datenbank entnommen. Die Schätzung für die Vermögenswerte basiert auf den Daten des Household Finance and Consumption Survey (HFCS) der österreichischen Nationalbank. Die Zahl der Zitierungen der American Economic Review wurde mittels Web of Science erhoben und die Gewichtsdaten zu Neugeborenen repräsentieren die gesammelten Geburten der ersten beiden Februarwochen in einem österreichischen Krankenhaus. Alle Daten sind auf Anfrage verfügbar.

Die in diesem Mechanismus implizite Logik des „wer hat, dem wird gegeben“ wird dabei mit Verweis auf eine entsprechende Stelle im Evangelium nach Matthäus auch als „Matthäus-Effekt“⁵ bezeichnet.

„[T]he Matthew effect consists in the accruing of greater increments of recognition for particular scientific contributions to scientists of considerable repute and the withholding of such recognition from scientists who have not yet made their mark.“ (Merton 1968, 58)

Derartige Matthäus-Effekte lassen sich dabei auch allgemeiner als selbstverstärkende Effekte charakterisieren, die im speziellen Fall der Matthäus-Effekte die folgende Struktur aufweisen: „Wird einE AutorIn/ein Beitrag x referenziert, dann wird x in der Zukunft als Referenzpunkt attraktiver.“

Selbstverstärkende Effekte in ähnlicher Form treten dabei in einer Reihe von sozialen Kontexten auf und sind in der wirtschaftswissenschaftlichen Literatur zur Pfadabhängigkeit ausführlich thematisiert worden (vgl. Sydow, Schreyögg und Koch 2009; Dobusch und Schüßler 2013). Die Kernthese dieses Theoriestrangs ist, dass die Etablierung technischer, organisationaler oder sozialer Standards zu selbstverstärkenden Effekten führen kann, die zu im Laufe der Zeit eine relative oder absolute Dominanz einzelner Standards verursachen. Dabei bildet sich der erstgenannte Fall relativer Dominanz zumeist in Form der bereits diskutierten Potenzverteilung ab.

Zu den klassischen Anwendungsfällen zählen dabei so unterschiedliche Beispiele wie die Konstanz einzelner technischer Standards (z.B. unser Tastaturdesign; siehe David 1985), die Monopolbildung auf den Softwaremärkten (Shapiro & Varian 1999), die relative Attraktivität von Internettauschbörsen oder die Cluster-Bildung im Bereich der High-Tech Industrie (Arthur 1994). Auch die Entwicklung der Pareto-Verteilung als „natürliche“ Form der Vermögensverteilung lässt sich durch die Integration selbstverstärkender Effekte einfach in einem theoretischen Modell nachvollziehen (Levy & Levy 2003). Die auf der empirischen Ebene identifizierten selbstverstärkenden Mechanismen sind zahlreich und reichen von direkten Netzwerkeffekten (je mehr ein Standard genutzt wird, desto attraktiver ist dieser), über Lerneffekte (je besser ein Standard erlernt ist, desto attraktiver ist dieser) bis hin zu indirekten Netzwerken-Effekten (ist ein Standard eine Voraussetzung zur Nutzung anderer

5 „Denn wer hat, dem wird gegeben, und er wird im Überfluss haben; wer aber nicht hat, dem wird auch noch weggenommen, was er hat“ (Mt 25,29)

Produkte, so wird der Standard attraktiver), wobei teils auch feingliedrigere Unterscheidungen zum Einsatz kommen.

Das bisher Gesagte erlaubt es nun die hier angesprochenen Themenfelder zu einem kohärenten Bild zusammenzufügen. Dabei bildet der Mertonsche Matthäuseffekt, als zentraler Mechanismus akademischer Reproduktion, einen geeigneten Ausgangspunkt, der sich mit der Theorie der Pfadabhängigkeit als „direkter Netzwerkeffekt“ verstehen lässt, da die Attraktivität eines Referenzpunkts mit der Zahl der vergangenen Referenzierungen korreliert. Dieser selbstverstärkende Effekt bildet die Grundlage für die Entstehung einer ungleich verteilten Aufmerksamkeit innerhalb des wissenschaftlichen Diskurses, in der einige, wenige Beiträge überproportional hohe Aufmerksamkeit erhalten. Dieser Umstand bildet sich in Folge in den wissenschaftsstatistischen Indikatoren zur Verteilung von Aufmerksamkeit, Prestige und Einfluss ab, die daher zumeist einer Potenzverteilung folgen.

Eine solche Diskussion akademischer Reproduktionslogik lässt sich nun nicht nur auf der Ebene individueller WissenschaftlerInnen und Beiträge führen, sondern auch mit einer Betrachtung der paradigmatischen Entwicklung der Wissenschaften im Sinne von Kuhn (1962) koppeln. Dies erfordert nach der Rolle selbstverstärkender Effekte in der Kuhnschen Argumentation bzw. der Wissenschaftsgeschichte im Allgemeinen zu fragen. Ein solches Unterfangen wurde dabei etwa von Sterman und Wittenberg (1999) im Rahmen einer simulationsbasierten Studie in Angriff genommen. Diese erzielten eine Reihe interessanter Resultate, die durchaus mit historischen Beobachtungen in Einklang zu bringen sind. So finden Sterman und Wittenberg, dass die intrinsische Qualität der Kernideen eines Paradigmas oft nur von geringer Bedeutung für dessen Erfolg ist und liefern so eine theoretische Rationalisierung für die historische Beobachtung nicht-linearer Entwicklungen im Bereich wissenschaftlicher Erkenntnis. Dieses Resultat steht dabei durchaus in Einklang mit Kuhns historischen Beschreibungen, etwa dem klassischen Beispiel der aufkommenden kopernikanischen Theorie, die in ihrer empirischen Erklärungskraft und Präzision der geozentrischen Auffassung hinterherhinkte, da letztere auf einen hochentwickelten theoretischen Apparat mit zahlreichen empirisch motivierten Korrekturthemen zurückgreifen konnte.

Für unsere Zwecke interessant ist ein Blick auf die Struktur des Modells von Sterman und Wittenberg, in dem sich drei wesentlich positive Feedback-Schleifen herausbilden, die die Konstanz von Paradigmen stabilisieren und somit eine explizite pfadtheoretische Begründung für die Kuhnsche Argumentation liefern. Die drei zentralen Feedback-Schleifen von Sterman

und Wittenberg haben die Form direkter Netzwerkeffekte und beziehen sich auf den akademischen Arbeitsmarkt, die wahrgenommene relative Erklärungskraft von Paradigmen und die Rolle sichtbarer Anomalien. In allen drei Fällen gilt, dass die größere Verbreitung eines Paradigmas zu selbstverstärkenden Effekten führen, da sich die Neueinstellungen, sowie die Zahl der erklärten Probleme und rationalisierten Anomalien proportional zur Größe des Paradigmas verhalten. Diese rein größenbedingten Faktoren wirken sich wiederum positiv auf die Attraktivität des Paradigmas aus – und stabilisieren so die Dominanz einmal etablierter Denkmuster im Sinne einer Pfadabhängigkeit der paradigmatischen Entwicklung.

“The prevalence of positive feedback processes in paradigm development means that the evolution of the system as a whole is strongly path-dependent.”
(Stermann und Wittenberg 1999, 333)

Vor dem Hintergrund dieser kurzen Einführung in die akademische Reproduktionslogik lässt sich nun die Rolle der evaluativen Wissenschaftsstatistik grob verorten. Zur Erinnerung: Die evaluative Wissenschaftsstatistik versucht durch die Ermittlung von normierten Nennungshäufigkeiten (zumeist erhaltene Zitationen eines/r AutorIn, eines Beitrags oder einer Zeitschrift) Einfluss innerhalb der Wissenschaft zu messen und interpretiert dieses Maß an „Impact“ im Umkehrschluss auch als Qualitätsmaß. Dabei sticht zuerst die selbstreferentielle Logik dieses Prozesses ins Auge: Es wird Einfluss gemessen und zur Grundlage einer institutionellen Bewertung gemacht, auf deren Basis der Einfluss in der Forschungslandschaft umverteilt wird. Damit trägt die Einführung von Rankingsystemen zur Bewertung von Forschungsleistungen zu einer weiteren Konzentration der akademischen Aufmerksamkeit bei, indem sie jene AkteurInnen, Beiträgen und Forschungsfeldern, denen bereits hohe Aufmerksamkeit zukommt, auch eine hohe Qualität bescheinigt (Dobusch und Kapeller 2009). Der Darstellung von Stermann und Wittenberg wird damit eine neue, vierte Feedbackschleife hinzugefügt: je größer ein Paradigma oder Forschungsfeld, desto höher ist auch die Zahl, der in diesem Bereich anfallenden Zitationen – es gilt also das Motto „big is beautiful“. Schließlich gelten die erhaltenen Zitationen ihrerseits als erstes und kaum hinterfragbares Signum wissenschaftlicher Qualität und erhöhen so die Attraktivität des betreffenden Paradigmas.

Diese zusätzliche Feedbackschleife hat dabei die Form eines indirekten Netzwerkeffekts: Die Etablierung von wissenschaftsstatistischen Indikatoren als Standard im Bereich der wissenschaftlichen Qualitätsbestimmung, macht größere, d.h. mit viel Aufmerksamkeit bedachte Forschungsstränge attraktiver, für all jene ForscherInnen, die die

Qualitätsstandards der Wissenschaftsevaluation internalisieren. Die sich daraus ergebende theoretische Diagnose, dass rankingbasierte Evaluationsverfahren der Wissenschaftsentwicklung inhärente Pfadabhängigkeiten weiter verstärken (siehe: Demange 2012), erlaubt auch einen ersten Befund zur Macht der Wissenschaftsstatistik zu erstellen: Eine Auswirkung der (evaluierenden) Wissenschaftsstatistik ist es bestehende akademische Reproduktionslogiken zu verstärken und damit zu einer noch ungleichmäßigeren Verteilung akademischer Aufmerksamkeit beizutragen.

3. Evaluationsroutinen und die Rolle reaktiver Messverfahren

Quantitative Evaluationsverfahren im Wissenschaftsbetrieb sind mannigfaltig. Sie basieren zumeist auf klassischen empirischen Forschungsmethoden wie Befragungen (etwa im Rahmen des peer-review Verfahrens zur Journalevaluation des Verbands der Hochschullehrer für Betriebswirtschaft⁶) oder Beobachtungen (z.B. von Zitierhäufigkeiten) und sind daher in der empirischen Sozialforschung grundsätzlich gut verankert. Allerdings wird von den Verfahren der empirischen Sozialforschung im Regelfall verlangt, dass sie gewissen methodologischen Anforderungen genügen. Zu diesen Anforderungen zählen typischerweise die Validität von Erhebungsmethoden (also dass tatsächlich gemessen wird, was gemessen werden sollen) sowie der Reliabilität (wiederholte Messungen führen zu ähnlichen Ergebnissen). Beide Aspekte verlangen dabei, dass empirische Mess-Verfahren nicht reaktiv sind, d.h. dass ihre Anwendung das untersuchte Verhalten nicht weiter beeinflusst.

Dieser idealisierten methodologischen Betrachtung steht dabei in der Praxis die Erfahrung gegenüber, dass das Phänomen Reaktivität im Rahmen innerwissenschaftlicher Evaluationsroutinen eine zentrale Stellung einnimmt. Reaktivität bedeutet, dass die Anwendung eines Untersuchungsinstruments oder –kontexts eine Verhaltensänderung der untersuchten Personen bewirken kann. Dabei wird Reaktivität im Forschungskontext primär als methodologisches Problem angesehen wird, das zu einer Verzerrung der erhobenen Daten führen kann. Für methodisch angeleitete Evaluierungen ergibt sich dabei das zusätzliche Problem, dass die bewerteten Personen oder Institutionen die Evaluationskriterien zu antizipieren versuchen und ihr Verhalten dementsprechend anpassen.

⁶ Siehe: <http://vhbonline.org/service/jourqual/vhb-jourqual-3/>

Eine wesentlichen Beitrag zum Verständnis der Auswirkungen reaktiver Messverfahren im Bereich der Wissenschaftsevaluation liefern Espeland und Sauder (2007), die sich in einer umfassenden Studie mit den Folgen der Einführung eines Rankings von *US-law schools* durch die Zeitschrift *U.S. News* befassen. Im Kern stellen Espeland und Sauder dabei fest, dass evaluative Messgrößen, insbesondere *rankings*, sowohl auf die soziale Umwelt als auch auf die AkteurInnen im Wissenschaftsbetrieb selbst einen starken Einfluss ausüben.

Konkret weisen die AutorInnen nach, dass das Verhalten und die Wahrnehmung maßgeblicher AkteurInnen im Kontext der *US-law schools* – Dekane und Fakultätspersonal, Studierende, öffentliche Institutionen, GeldgeberInnen, etc... – durch zwei zentrale reaktive Mechanismen beeinflusst wird, nämlich *self-fulfilling prophecies* und *commensuration*. Der erste Mechanismus bezieht sich dabei darauf, dass das *law school*-Ranking zu Verhaltensänderungen führt, die die Rankingergebnisse weiter verstärken und polarisieren. Damit erhält das Ranking auch einen selbstbestätigenden Charakter, da die Reihung der *law schools* nach Punktezahlen maßgebliche Unterschiede signalisiert, die ein geändertes Verhalten des sozialen Umfeldes (Studierende, Öffentlichkeit, Financiers etc..) induziert, welches die jeweilige Position bzw. den Trend (Auf- bzw. Abstieg) im Ranking festigt. So bestimmen etwa frühere Rankings über die Zuteilung finanzieller Mittel und im Rahmen von peer-review Verfahren werden vergangene Platzierungen direkt oder indirekt herangezogen um aktuelle Leistungen zu bewerten. Gerade im unteren Bereich einer Rangliste führt diese Macht der selbsterfüllenden Prophezeiung zu einer Abwärtsspirale: Je schlechter die Platzierung im Ranking, umso geringer die Ausstattung mit finanziellen Mitteln und Ressourcen durch externe Financiers, die im amerikanischen Hochschulsystem eine zentrale Rolle einnehmen, desto schwieriger wird es im Ranking wieder aufzusteigen. Auf Seiten der *law schools* führt die Internalisierung dieser Faktoren zu einem verstärkten Fokus auf jene Aspekte, die einen Einfluss auf die Position in der Rangliste haben und gleichzeitig zur Vernachlässigung von Aspekten, die für das Ranking irrelevant sind (Espeland und Sauder 2007, 11-14).

"Rankings create self-fulfilling prophecies by encouraging schools to become more like what rankings measure, which reinforces the validity of the measure."
(Espeland und Sauder 2007, 15)

Der zweite Mechanismus, *commensuration*, zielt auf eine Veränderung der menschlichen Wahrnehmung ab, die durch das numerisch-kompetitive, einfach gehaltene Framing eines komplexen sozialen Sachverhalts entsteht. Dabei spielt insbesondere die konzeptionelle Klarheit von Rankings eine zentrale Rolle, die suggeriert die relative Performance von

Ausbildungsinstitutionen anhand eines soliden gemeinsamen Maßstabs vergleichend abzubilden.

"[Commensuration] changes the locus and form of attention, both creating and obscuring relations among entities. Commensuration is characterized by the transformation of qualities into quantities that share a metric, a process that is fundamental to measurement." (Espeland und Sauder 2007, 16)

Konkret wird im Kontext des *law-school* Rankings der *commensuration*-Mechanismus vor allem durch das hierarchische Verhältnis ausgelöst, das durch eine explizite Reihung nach Punktezahlen entsteht. Diese Reduktion und gleichzeitige Normierung von Qualität und Reputation einer Fakultät auf einen einzelnen numerischen Wert führt zu einer signifikanten Simplifizierung, wobei komplexe qualitative Unterschiede zwischen einzelnen *law-schools* durch die quantitativ-präzise Differenzierung scheinbar vollständig erfasst werden. Darüber hinaus folgen die meisten Betrachter des Rankings dem Prinzip einer ordinalen Vereinfachung, in dem Positionsunterschiede zwischen *law-schools* – ungeachtet dessen, wie minimal das Ausmaß des numerischen Unterschiedes auch sein mag – zumeist simpel als "schlechter" oder "besser" rationalisiert werden (Espeland und Sauder 2007, 16-20). Dieselbe Wahrnehmungslogik gilt analog auch für Auf- bzw. Abstiege im Ranking, die große Auswirkungen bei der Beurteilung durch das relevante soziale Umfeld haben und somit auch in den Top-Segmenten des Rankings für hohen Wettbewerbsdruck sorgen. Der soziale Einfluss von Rankings beruht damit sowohl auf der Vereinheitlichung komplexer Sachverhalte als auch auf einer strikten Differenzierung zwischen unterschiedlichen Objekten vor dem Hintergrund eines gemeinsamen Maßstabs.

„By simultaneously unifying and distinguishing objects [...] rankings classify, reward and punish, and organize interventions.“ (Espeland und Stevens 2008, 416)

Dieser soziale Effekt von Rankings führt zu einer Reihe spezifischer Handlungsmustern auf Seiten der *law schools*: So geht etwa die bereits oben erwähnte verstärkte Orientierung hin zu Ranking-relevanten Faktoren mit einer Umverteilung von Ressourcen zugunsten dieser Faktoren einher. Darunter fallen verstärkte Ausgaben für Marketing, insbesondere das Drucken von Broschüren zur Steigerung der Bekanntheit mit dem Ziel die Chancen in künftigen peer-review Verfahren zu erhöhen⁷, oder die Änderungen interner Kriterien und Politiken. So erhöhen viele *law-schools* die Zahl der Leistungsstipendien um bessere

⁷ In Folge werden die entsprechenden Broschüren primär an andere *law-schools* versendet (vgl. Espeland und Sauder 2007, 26).

Studierende anzuwerben, deren Testergebnisse ebenso für die Bewertung relevant sind. Alternative Bewertungskriterien für KandidatInnen – etwa Bedürftigkeit oder soziales Engagement – verlieren hingegen an Bedeutung (Espeland und Sauder 2007, 24-26). Auch zentrale Serviceleistungen der *US-law schools* erfuhren im Zuge der Etablierung des *U.S. News-Rankings* Veränderungen: So wurden die fakultätseigenen *career services* angehalten den Erwerbsstatus von AbsolventInnen⁸, der ebenfalls zur Punktezahl im Ranking beiträgt, besonders im Auge zu haben und insbesondere auf eine breite arbeitsmarktpolitische Inklusion der AbsolventInnen zu achten – auch wenn dies etwa bedeuten sollte Job-Empfehlungen auszugeben, die eigentlich nicht der Qualifikation der AbsolventInnen entsprechen (Espeland und Sauder 2007, 28).

Die Vernachlässigung qualitativer und bedürfnisorientierter Aspekte bei der Vermittlung von Arbeitsstellen ist nur eine der Folgen dieses "Rating-Kalküls" – so wurden etwa durchwegs eigene Abteilungen zu arbeitsmarktpolitischen Observation der AbsolventInnen eingerichtet, um dem Rankingbetreiber entsprechende Daten zur Verfügung stellen zu können (Espeland und Sauder 2007, 27). Abschließend lässt sich auch das Aufkommen von sogenannten "Gaming-Strategien" beobachten, die einzelne Fakultäten mit dem Ziel verfolgen, die für das Ranking relevanten Faktoren zu erhöhen ohne jedoch die mit diesen Faktoren verknüpften Parameter tatsächlich zu verändern. Mögliche Gaming-Strategien sind beispielsweise die Umschichtung von Studierenden von Vollzeit-Programmen in Teilzeit-Modelle oder die Einführung von Frühlings-Ferien. Diese Vorgehensweise verbessert zwei zentrale Ranking-Kriterien, nämlich den LSAT-Mittelwert⁹ der Studierenden als auch das Betreuungsverhältnis (da Teilzeit-Studierende hier nicht mitgezählt werden). Ebenso kann durch die temporäre Anstellung von arbeitslosen AbsolventInnen die Beschäftigungszahlen erhöht werden, die ebenso einen für das Ranking-Verfahren relevanten Parameter darstellen (Espeland und Sauder 2007, 24-32).

Die Forcierung von Gaming Strategien als Resultat reaktiver Prozesse ist nicht nur im Kontext der oben dargestellten Fallstudie von Espeland und Sauder von Relevanz. Frost und Brockmann (2014) etwa identifizieren in einer weiteren Fallstudie, durchgeführt an deutschen Universitäten, bewusst strategisches Verhalten von individuellen ForscherInnen, um der immer stärker dominierenden Steuerungslogik der *Entrepreneurial University* und dem damit verbundenen quantitativen Verständnis von wissenschaftlicher Produktivität (z.B.

⁸ Im Falle der USN-Rankings sind hier zwei zentrale Indikatoren relevant: Berufstätigkeit bei Abschluss und Berufstätigkeit 9 Monate nach Abschluss (vgl. Espeland und Sauder 2007, 27).

⁹ LSAT die Abkürzung für Law School Admission Test.

die Anzahl von Publikationen in hochgerankten Journalen oder der Anteil akquirierter Drittmittel) zu entsprechen.

Dieser allgemeinen Logik folgend werden durch rankingbasierte Bewertungsverfahren systematische Anreize geschaffen. Hier spielen auch zitationsbasierte Evaluationsverfahren von Fachzeitschriften, die in die Erstellung von Journalrankings münden, eine zentrale Rolle. So hat etwa der Journal Impact Factor von *Thomson Reuters*¹⁰ eine zentrale Rolle bei der Beeinflussung der Wahrnehmung wissenschaftlicher Fachzeitschriften als auch einzelner Beiträge und AutorInnen – eine Problematik, die Willmott (2011) unter dem pointierten Titel des "*Journal list fetishism*" zusammenfasst. Für die AutorInnen wissenschaftlicher Beiträge gilt vor dem Hintergrund dieser dominanten Rolle von Journal-Rankings die Devise möglichst viele Impact-Faktor Punkte zu sammeln und daher (a) vorzugsweise Journal-Artikeln zu veröffentlichen (keine Bücher bzw. Buchkapitel), (b) so kurze Artikel wie möglich zu verfassen (c) an möglichst vielen kollaborativen Arbeiten mitzuwirken und (d) vorwiegend in englischer Sprache zu publizieren. Diese Strategien erleichtern es ForscherInnen, die mit wissenschaftlichen Evaluationsroutinen konfrontiert sind, zu reüssieren und sind daher von offensichtlicher Attraktivität (Kapeller 2010, 1386).

Diese empirischen Befunde zur Reaktivität quantitativer Evaluationsroutinen auf Ebene der AutorInnen lassen sich um weitere Befunde, speziell aus dem Umfeld der HerausgeberInnen wissenschaftlicher Fachzeitschriften ergänzen. Auch diese sehen sich mit bestimmten Anreizen konfrontiert, wobei hierbei meist die Zitierhäufigkeit der eigenen Beiträge im Fokus steht. Eine oft genutzte Möglichkeit die Zitationshäufigkeit zu steigern, stellt die Praxis dar, einreichende AutorInnen dazu zu ermutigen, journaleigene Artikel zu zitieren (Kapeller 2010, 1386; Wilhite und Fong 2012, 1). Auch intransparente, direkte Interventionen bei Datenbanktreibern zur Modifikation der Rankingergebnisse wurden dabei bereits dokumentiert (z.B. Rossner et al. 2007, Smith 2008). Zuweilen finden auch kuriosere – und leichter feststellbare – Gaming-Strategien Anwendung: Hierunter fällt etwa das Beispiel des *Journals of Gerontology*, das mittels Eigenziten in Editorials des Herausgebers seinen Impact Faktor von 2,9 auf 4,1 erhöhen konnte (Reedijk und Moed 2006, 188).

¹⁰ Der Impact Factor dient als zentrale Referenzgröße für die von *Thomson Reuters* veröffentlichten Journalrankings (*Journal Citation Report*) und ist eine Maßzahl, die die Häufigkeit angibt mit der ein Artikel in einer wissenschaftlichen Fachzeitschrift in einem bestimmten Zeitraum durchschnittlich zitiert wurde.

Auch auf der Ebene ökonomischer Fachzeitschriften finden sich derartige reaktive Trends, die Auswirkungen auf das Zitierverhalten innerhalb wissenschaftlicher Disziplinen nehmen. Eine Folge der dominanten Rolle von Journal-Rankings ist etwa, dass Journale als Bezugskriterium relativ zu einzelnen AutorInnen oder Beiträgen sichtbarer und damit attraktiver werden müssten und es daher zu einer Übertragung von Prestige in Richtung der Journale kommen sollte: In diesem Fall sollten alle Beiträge eines als hochqualitativ ausgewiesenen Journals gleichzeitig attraktiver werden und sich insgesamt eine gleichmäßigere Verteilung an Zitationen ergeben. Abbildung 2 zeigt, dass diese Erwartung für den Fall eines zeitraumbezogenen Vergleichs der Zitationen von Beiträgen aus drei als hochqualitativ angesehen Ökonomie-Journalen auch tatsächlich empirisch erfüllt ist. Konkret wird dabei die Anzahl jener Beiträge verglichen, die in den betrachteten Zeiträumen (1981-1985 und 2004-2008) publiziert wurden, und die innerhalb der genannten Zeiträume sowie in den fünf folgenden Jahren nie oder nur einmal zitiert werden.

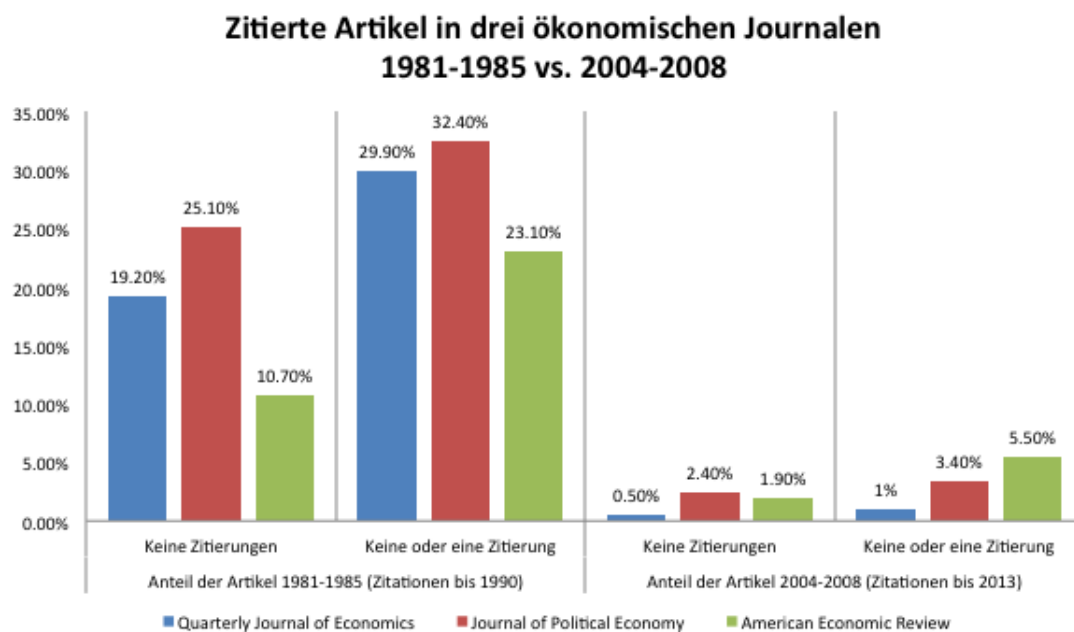


Abbildung 2: Vergleich der Zitationen publizierter Beiträge von drei einflussreichen Ökonomie-Journalen (Eigene Darstellung basierend auf Daten von Thomson Scientific).

Die von Solla-Price (1965) gewonnene Erkenntnis, nach der eine Mehrheit der verfügbaren Literatur – in diesem konkreten Fall eine Mehrheit der Beiträge in einem Journal – keine bzw. eine nur geringe Beachtung im Zitationsprozedere zukünftiger Beiträge findet, wurde bereits in Kapitel 2 erwähnt und auch empirisch mehrfach bestätigt (Garfield 2006, 91; Nature 2005, Seglen 1996). Die in Abbildung 2 deutlich erkennbare Verringerung des Anteils der nie bzw. nur einmal zitierten Beiträge ist daher hochgradig überraschend, kann aber durch die

postulierte Hypothese, dass sich der Referenzpunkt für akademische Reputation tendenziell weg von individuellen AutorInnen bzw. einzelnen Beiträgen hin zu hochgerankten Journalen verlagert, erklärt werden.

Ausgehend von der zu Beginn dieses Kapitels vorgestellten Fallstudie über die Auswirkungen der Reaktivität evaluativer Messverfahren und der Ergänzung dieser um weitere explorative Befunde aus dem Umfeld zitationsmetrischer Evaluationsverfahren – konkret Journalrankings – lässt sich folgender Befund zur Macht der Wissenschaftsstatistik ableiten: Die gegenwärtig praktizierte Wissenschaftsstatistik hat die Macht, das Verhalten von ForscherInnen und wissenschaftlichen Institutionen maßgeblich zu beeinflussen. In der Ökonomie kristallisiert sich diese Macht – wie am Schluss dieses Kapitels gezeigt wurde – anhand des Beispiels ökonomischer Fachzeitschriften in besonderem Maße heraus.

4. Ökonomische Diskurse und der Einfluss der Wissenschaftsstatistik auf die Zukunft der Ökonomie

Der fachwissenschaftliche Diskurs in der Ökonomie unterscheidet sich von anderen Sozialwissenschaften insofern, als dass die Ökonomie als einzige Sozialwissenschaft über ein dominantes Paradigma in Form der neoklassischen Theorie verfügt, das die akademische Lehre und Forschung im ökonomischen Bereich nachhaltig prägt. Die historischen Ursachen dieser paradigmatischen Dominanz sind dabei vielfältig (siehe z.B. Backhouse 2005 oder Dobusch und Kapeller 2009). Im Bereich des engeren Forschungsdiskurses stechen hierbei vor allem die 1970er und 1980er Jahre heraus. In diesem Zeitraum wurden nicht nur keynesianische Ansätze in der Makroökonomie von neoklassischen und monetaristischen Theorien verdrängt, sondern auch die ökonomische Journalkultur erfuhr eine signifikante Verengung. Diese fand ihren Ausdruck in dem nahezu vollständigen Ausschluss kritischer und auf alternativen theoretischen Ansätzen basierender Beiträge aus dem ökonomischen Fachdiskurs; derartige Beiträge wurden, mit wenigen Ausnahmen, in den entsprechenden Review-Verfahren aussortiert. Die Konfrontation mit diesem Ausschluss aus dem „offiziellen“ ökonomischen Diskurs führte in Folge zur Gründung von heute bedeutenden heterodox-ökonomischen Journalen, wie dem *Cambridge Journal of Economics*, dem *Journal of Post Keynesian Economics* oder dem *Journal of Economic Issues* (King 2003, 134-136). In diesem Sinne ist die Ökonomie zwar nach wie vor eine „umstrittene Disziplin“ (Lee und Elsner 2011), allerdings vor dem Hintergrund einer stetig ansteigenden paradigmatischen Homogenität. Aus demselben Grund ist es auch wenig überraschend, dass alternative bzw. heterodoxe Theorietraditionen in der Ökonomie quantitativ gesehen nur einen kleinen

Bruchteil der ökonomischen Disziplin ausmachen, da diese auf verschiedenen Ebenen mit derartigen Ausschlussroutinen konfrontiert sind.

„Die Konfrontation Heterodoxie kontra Mainstream bezieht ihre Existenz und ihre Berechtigung vielmehr aus dem gegenwärtigen Zustand des Wissenschaftsregimes im ökonomischen Bereich, das durch eine unübersehbare Bevorzugung und Förderung eines Mainstreams neoklassischer Prägung an Universitäten, Forschungsinstituten und staatlichen und internationalen Wirtschaftsorganisationen charakterisiert ist. Konzentration auf diese Richtung wird im Studium gefördert und spielt eine entscheidende Rolle für Beruf und Karriere. Von dieser Schiefelage sind alle nicht-neoklassischen und kritischen Richtungen betroffen, was die Herausbildung der Heterodoxie-Terminologie und der Heterodoxie-Bewegung erklärt.“ (Rothschild 2008, 25)

Freilich erlaubt die Beschreibung dieser Konstellation per se noch keine Schlussfolgerung hinsichtlich der Repräsentation alternativer ökonomischer Ansätze innerhalb des mainstream-ökonomischen Diskurses. Immerhin könnte es ja dennoch sein, dass alternative ökonomische Theorieansätze innerhalb der Mainstream-Ökonomie rezipiert und etwa als Anregung für eigene Forschung oder als Ansatzpunkt zu einer kritischen Auseinandersetzung mit den Argumenten Dritter verwendet werden. Die Frage inwiefern heterodoxe Ideen Eingang in den mainstreamökonomischen Diskurs finden ist also primär empirisch zu beantworten. Dies soll im Folgenden auch versucht werden: Zu diesem Zweck wurden alle wechselseitigen Zitationsbeziehungen zwischen 26 Ökonomie-Journalen – jeweils dreizehn hochrangige Journale aus den Bereichen der Mainstreamökonomie und der heterodoxen Ökonomie – in einem Zeitraum von zwanzig Jahren (1989-2008) erhoben¹¹ und hinsichtlich der Frage wechselseitiger Auseinandersetzung zwischen dem neoklassischem Mainstream und seinen heterodoxen Alternativen untersucht. Tabelle 2 zeigt ein zentrales Resultat dieser Untersuchung: Während die dreizehn untersuchten heterodoxen Journale ein ausgewogenes Zitierverhalten an den Tag legen und gleichermaßen heterodoxe wie mainstreamökonomische Quellen referieren, sind heterodoxe Argumente in der mainstreamökonomischen Literatur kaum repräsentiert. Der heterodoxe Diskurs ist demnach tatsächlich offen für die Diskussion standardökonomischer Argumente und in diesem Sinne pluralistisch, während die relative Geschlossenheit des orthodoxen bzw.

11 Als Grundlage für die Samplezusammenstellung diente dabei die Zitationsdatenbank Web of Science von Thomson Reuters. Die darin enthaltenen Journal Citation Report 2007 bilden auch die Grundlage für die Auswahl der relevanten Journale, wobei die dreizehn bestgerankten Journale als „Top 13 orthodox“ gedeutet wurden und die „Top 13 Heterodox“ unter Zuhilfenahme des Heterodox Economics Directory (www.heterodoxnews.com/hed) identifiziert wurden.

mainstreamökonomischen Diskurses auf eine tendenziell monistische Attitüde des Mainstreams verweist.

	<i>Anteil der Zitate aus den Top 13 der orthodoxen Journale</i>	<i>Anteil der Zitate aus den Top 13 der heterodoxen Journale</i>
<i>In den Top 13 "orthodox"</i>	97,15%	2,85%
<i>In den Top 13 "heterodox"</i>	47,58%	52,42%

Tabelle 2: Interaktionsmuster zwischen Mainstreamökonomie und Heterodoxie (1989-2008; entnommen aus Dobusch und Kapeller 2012b, 474)

Eine genaue Analyse der Daten aus Tabelle 2 zeigt zudem, dass der ausgewiesene Wert für den Export heterodoxer Argumente in die Mainstreamliteratur (2,85% der hier untersuchten Referenzen) massiv von statistischen Ausreißern getrieben ist. In absoluten Zahlen betrachtet handelt es sich bei den 2,85% um 753 Zitierungen, wobei 613 dieser Zitate von nur drei Journalen abhängig sind, die jeweils eine Sonderstellung im ökonomischen Diskurs innehaben.¹² Damit verbleibt für die restlichen 23 Journale nur ein minimaler Ideentransfer von 140 Zitationen innerhalb des Untersuchungszeitraums von 20 Jahren.

Auch wenn Tabelle 2 für den Nachweis reicht, dass alternative ökonomische Ansätze in der mainstreamökonomischen Literatur nur minimal repräsentiert sind, bleibt die Frage offen, ob das beobachtete Muster nun von paradigmatischen Trennlinien herrührt oder einfach das Resultat einer starken Fokussierung des ökonomischen Diskurses auf Beiträge in hochgerankten Journalen liegt: Immerhin bildet das orthodoxe Sample die Top 13 aus dem *Journal Citation Report 2007* ab, während die Rankings der heterodoxen Journalen von Rang 17 bis 130 streuen. Dieses Argument ist vor allem vor dem Hintergrund der Besonderheiten des fachwissenschaftlichen Diskurses in der Ökonomie von Bedeutung, der durch eine hohe Selbstbezüglichkeit sowie eine starke Konzentration und Elitenorientierung gekennzeichnet ist. Diese Aspekte wurden jüngst von Fourcade et al. (2015) ausführlich dokumentiert und zeigen, dass die ökonomische Literatur im interdisziplinären Vergleich viel weniger geneigt ist

¹² Auf der heterodoxen Seite des Samples generiert das sowohl für mainstreamökonomische als auch für heterodoxe Beiträge attraktive *Journal of Economic Behavior and Organization* 340 Export-Zitate, während auf der mainstreamökonomischen Seite des Samples zwei Journale aus dem Bereich der Wirtschaftsgeographie (*Journal of Economic Geography* und *Economic Geography*) inkludiert sind, die weitere 273 Zitate importieren. Dabei zeigt ist, dass die relative Offenheit für heterodoxe Ansätze im Bereich der Wirtschaftsgeographie ungleich höher ist als in der engeren Mainstreamökonomie.

andere Disziplinen zu referenzieren (Selbstbezüglichkeit) und einen viel stärkeren Fokus auf eine enge Gruppe an Top-Journalen aufweist, dessen AutorInnen primär von einer kleinen, homogenen Menge an Universitäten stammt (Konzentration; siehe hierzu bereits Hodgson und Rothman 1999). Eine solche Konstellation bedeutet dabei nicht nur eine wesentlich fokussiertere Aufmerksamkeitslogik innerhalb der ökonomischen Disziplin, sondern geht auch einher mit einer vergleichsweise straff hierarchischen internen Organisation (Elitenorientierung). Daher überrascht es auch nicht, dass sich eine kürzlich im *Journal of Economic Literature* publizierte ausführliche Analyse zur Entwicklung ökonomischer „Top-Journale“ auf die Betrachtung von genau fünf Fachzeitschriften beschränkte (Card und DellaVigna 2013).

Vor diesem Hintergrund scheint es naheliegend das Verfahren aus Tabelle 2 mit einer entsprechenden Kontrollgruppe zu replizieren um zu untersuchen, inwiefern die fehlende Rezeption alternativer Theorieansätze rein der Konzentration des ökonomischen Diskurses geschuldet ist. Aus diesem Grund zeigt Tabelle 3 eine Replikation von Tabelle 2, wobei das heterodoxe Sample durch eine Kontrollgruppe ersetzt wurde. Die Kontrollgruppe beinhaltet dabei jene 13 Journale, die im JCR 2007 jeweils einen Rang unterhalb der dreizehn zuvor untersuchten heterodoxen Journale platziert sind. Das Resultat zeigt, dass die zuvor beobachteten Diskursmuster nicht nur durch die relative Rangposition der Journale erklärt werden kann – im Gegenteil hat ist die Zitierhäufigkeit der Kontrollgruppe mehr als dreimal so hoch wie jene der heterodoxen Journale. Paradigmatische Faktoren scheinen daher in der Tat eine zentrale Rolle zu spielen und legen einen systematischen diskursiven Ausschluss alternativer Theorieansätze nahe.

	<i>Anteil der Zitate aus den Top 13 der orthodoxen Journale</i>	<i>Anteil der Zitate aus der Kontrollgruppe</i>
<i>In den Top 13 "orthodox"</i>	91,09%	8,91% ¹³
<i>In der Kontrollgruppe</i>	66,15%	33,85%

Tabelle 3: Interaktionsmuster zwischen den Top-Journalen der Mainstreamökonomie und einer Kontrollgruppe (1989-2008; Eigene Darstellung)

Freilich ist dieser Aspekt nicht die einzige Schwachstelle der hier referierten Untersuchung zur Frage interparadigmatischer Diskurse in der Ökonomie. So ist auch der abgebildete

¹³ Vergleicht man die relative Änderungen des Zitierverhaltens der orthodoxen Journale mit den Daten aus Tabelle 2 so ergibt sich ein statistisch signifikanter Unterschied auf dem 1%-Niveau.

Vorkrisenzeitraum ein möglicher Kritikpunkt. Sollte nicht die Finanz- und Wirtschaftskrise als eine zentrale Anomalie verstanden werden, die in scharfem Kontrast zu neoklassischen Basispostulaten – wie der Effizienzmarkthypothese und dem zugehörigen Arbitrage-Denken – steht? Und könnte es nicht sein, dass diese Krise auch als Türöffner fungiert, der alternativen Ansätzen im ökonomischen Denken vermehrt Gehör verschafft?

Die zweite Frage kann dabei mit Kuhn auch theoretisch beantwortet werden, wobei hier genau die umgekehrte Reaktion zu erwarten ist. Nach Kuhn würde ein dominantes Paradigma, das mit einer Anomalie von hoher Bedeutung konfrontiert wird, versuchen dieses Problem durch Adaption etablierter Thesen, Modelle und Methoden zu fassen und zugleich Grundlegendebatten mit Vehemenz vermeiden. Empirisch gesehen lässt sich etwa auf die Konstanz ökonomischer Argumente in Lehre und Politikberatung verweisen. Für den Bereich der Forschung hingegen, in dem unbestreitbar auf diese Anomalie reagiert wird (z.B. Young 2014), scheint eine differenziertere Analyse sinnvoll, die danach fragt, inwiefern diese kriseninduzierten Reaktion auch alternativen ökonomischen Theorieansätzen Rechnung tragen, oder ob diese eher dem Kuhnschen Muster einer „internen“ Adaption bestehender Modelle folgen. Aus diesem Grund liefert Tabelle 4 eine weitere Replikation von Tabelle 2, die anstelle des Vorkrisenzeitraums, die aktuelle Literatur (von 2009 bis 2013) ins Auge fasst.

	<i>Anteil der Zitate aus den Top 13 der orthodoxen Journale</i>	<i>Anteil der Zitate aus den Top 13 der heterodoxen Journale</i>
<i>In den Top 13 "orthodox"</i>	97,61% (97,15%)	2,39% (2,85%)
<i>In den Top 13 "heterodox"</i>	24,58% (47,58%)	74,42% (52,42%)

Tabelle 4: Interaktionsmuster zwischen Mainstreamökonomie und Heterodoxie (2009-2013; Eigene Darstellung; Werte aus Tabelle 2 in Klammern)

Hier zeigt sich – ganz im Sinne der Kuhnschen Auffassung – keine wesentliche Änderung im Zitierverhalten des dominanten Paradigmas; dieses bleibt auf Seiten des ökonomischen Mainstreams vielmehr weitgehend konstant. Die Anomalie der Finanz- und Wirtschaftskrise hat also die Rezeption alternativer Theorieansätze innerhalb des mainstreamökonomischen Diskurses nicht weiter intensiviert.

Vor diesem Hintergrund und der Einsichten aus Kapitel 3 lässt sich damit die These aufstellen, dass innerwissenschaftliche Kriterien – wie etwa Zitationsmetriken und die damit verbundenen institutionalisierten Evaluationsroutinen – einen stärkeren Einfluss auf die

Aufmerksamkeitsverteilung in der Ökonomie ausüben als die tatsächliche ökonomische Entwicklung in Form der Finanz- und Wirtschaftskrise. Ein wesentlicher Grund hierfür liegt im „size-bias“ von Zitationsmetriken, die quantitativ bedeutsame Forschungsfelder – wie die Mainstreamökonomie – per Definition als hochqualitativ ausweisen. Dass dies gerade in der Ökonomie als einer „umstrittenen Disziplin“ mit hochgradig ungleichen Anfangsausstattungen zu einer weiteren Verstetigung des dominanten Ansatzes führt scheint dabei naheliegend.

Erste Indizien für eine solche Entwicklung finden sich etwa in Frankreich: Hier spielen Zitationsmetriken schon seit einigen Jahren eine Rolle für die weitgehend zentralistische Berufungspolitik von ProfessorInnen der Ökonomie. Beginnend mit dem Jahr 2005 wurde dabei die Rolle von Zitationsmetriken durch die Einführung eines Punktesystems formalisiert. Bei gleicher Produktivität erzielen dabei heterodoxe ÖkonomInnen durch den Größenbias der Zitationsmetriken eine signifikant niedrigere „Punktezahl“, die in Folge eine Anstellung heterodoxer ÖkonomInnen weniger attraktiv erscheinen lässt. Dieser Umstand bildet sich auch in der Berufungspolitik ab: So sank parallel zu dieser Umstellung etwa der relative Anteil heterodoxer ÖkonomInnen unter den Neuberufungen von knapp 18% in den Jahren 2000-2004 auf 5% in den Jahren von 2005-2011 (FAPE 2014). Vergleichbare Trends finden sich auch in anderen Ländern, in den auf Zitationsmetriken beruhende akademische Evaluationsroutinen zum Einsatz kommen. So zeigt sich etwa in Großbritannien, wo im Rahmen einer *Research Assessment Exercise* jeweils fachlich vergleichbare Institute evaluiert werden, dass die sukzessive Evaluierung Einfluss auf die Berufungspolitik der jeweiligen Institute nimmt. Dabei werden die zu Gunsten des ökonomischen Mainstream verzerrten Rankingkriterien antizipiert und mainstreamökonomische Bewerbung in Folge positiv diskriminiert (siehe dazu Lee 2007, der die Situation zu Beginn der 2000er Jahre analysiert).

An dieser Stelle zeigt sich, dass die Macht einer evaluativen Wissenschaftsstatistik nicht zuletzt auch darin liegt vorherrschende wissenschaftliche Paradigmen in ihrer dominanten Rolle zu stabilisieren und so die Homogenisierung von Disziplinen zu fördern.

5. Fazit

Zweck dieser Untersuchung war es, einen Überblick über die Auswirkungen wissenschaftsstatistischer Evaluationsverfahren in den Wissenschaften zu gewinnen und dabei einen besonderen Fokus auf die Ökonomie zu legen. Dabei konnten drei zentrale Befunde identifiziert werden, die für eine maßgebliche Macht der (evaluativen)

Wissenschaftsstatistik sprechen: Erstens trägt die evaluative Wissenschaftsstatistik zu einer Verstärkung und Festigung der akademischen Reproduktionslogik im Sinn einer steigenden Ungleichverteilung von Aufmerksamkeit, Prestige und Ressourcen innerhalb einer Disziplin bei (Kapitel 2). Zweitens wird durch die mannigfaltigen Mechanismen der Reaktivität, die evaluativen wissenschaftsstatistischen Methoden grundsätzlich anhaften, das Verhalten von wissenschaftlichen AkteurInnen aktiv beeinflusst (Kapitel 3). Drittens tragen evaluative Zitationsmetriken maßgeblich dazu bei, die zukünftige Entwicklung einer wissenschaftlichen Disziplin – dargestellt am Beispiel des zitationsmetrischen Interaktionsmusters zwischen ökonomischen Mainstream und heterodoxer Ökonomie – im Sinne eines dominanten ökonomischen Paradigmas zu stabilisieren und somit den Einfluss alternativer oder kritischer Ansätze im wissenschaftlichen Diskurs zu begrenzen (Kapitel 4).

Literatur

- Adler, Robert, Ewing, John und Taylor, Peter (2008). Citation Statistics. *Report from the International Mathematical Union*. URL: <http://www.mathunion.org/fileadmin/IMU/Report/CitationStatistics.pdf> (dl. 10.03.2015).
- Arthur, Brian W. (1994): *Increasing returns and path-dependence in the economy*. University of Michigan Press.
- Backhouse, Roger E. (2005): The Rise of Free Market Economics: Economists and the Role of the State since 1970. *History of Political Economy*, 37: 355-392.
- Card, David und DellaVigna, Stefano (2013): Nine facts about top journals in economics. *Journal of Economic Literature*, 51(1): 144-161.
- David, Paul A. (1985): Clio and the economics of QWERTY. *American Economic Review Proceedings*, 75: 332-337.
- Demange, Gabrielle (2012): On the influence of a ranking system. *Social Choice and Welfare*, 39: 431-455.
- Dobusch, Leonhard und Kapeller, Jakob (2009): Why is Economics not an Evolutionary Science? New Answers to Veblen's old Question. *Journal of Economic Issues*, 43(4): 867-898.
- Dobusch, Leonhard und Kapeller, Jakob (2012a): Heterodox United vs. Mainstream City? Sketching a framework for interested pluralism in economics. *Journal of Economic Issues*, 46(4): 1035-1057.
- Dobusch, Leonhard und Kapeller, Jakob (2012b): A guide to paradigmatic Self-marginalization - Lessons for Post-Keynesian Economists. *Review of Political Economy*, 24(3): 469-487.
- Dobusch, Leonhard und Schüßler, Elke (2013): Theorizing Path Dependence: A Review of Positive Feedback Mechanisms in Technology Markets, Regional Clusters and Organizations. *Industrial and Corporate Change*, 22 (3), 617-647.
- Espeland, Wendy N. und Sauder, Michael (2007): Rankings and Reactivity: How Public Measures Recreate Social Worlds. *American Journal of Sociology*, 113(1), 1-40.
- Espeland, Wendy N. und Stevens, Mitchell L. (2008): A Sociology of Quantification. *European Journal of Sociology*, 49(3): 401-436.
- FAPE (2014): Evolution of Economics Professors' Recruitment since 2000 in France: The End of Pluralism. Working Paper. URL: http://www.heterodoxnews.com/HEN/attach/hen167/FAPE_pluralism_france.pdf (dl. 30.02.2014).

- Fourcade, Marion, Ollion, Etienne und Algan, Yann (2015): The superiority of economists. *Journal of Economic Perspectives*, 29(1): 89-114.
- Franke, Richard H. und Kaul, James D. (1978): The Hawthorne Experiments: First Statistical Interpretation. *American Sociological Review*, 43(5): 623-643.
- Frost, Jetta und Brockmann, Julia (2014): When qualitative productivity is equated with quantitative productivity: scholars caught in a performance paradox. *Zeitschrift für Erziehungswissenschaft*, 17: 25-45.
- Garfield, Eugene (2006): The History and Meaning of the Journal Impact Factor. *Journal of the American Medical Association*, 295(1): 90-93.
- Hodgson, Geoffrey M. und Rothman, Harry (1999): The editors and authors of economics journals: a case of institutional oligopoly? *Economic Journal*, 109: F165-F186.
- Kapeller, Jakob (2010): Citation Metrics: Serious drawbacks, perverse incentives and strategic options for heterodox economics. *American Journal of Economics and Sociology*, 69(5): 1376-1408.
- King, John E. (2003): *A History of Post-Keynesian Economics since 1936*. Cheltenham: Edward Elgar.
- Lee, Frederic S. (2007): The Research Assessment Exercise, the state and the dominance of mainstream economics in British universities. *Cambridge Journal of Economics*, 31(2): 309-325.
- Lee, Frederic S. und Elsner, Wolfram (2011): *Evaluating Economic Research in a Contested Discipline: Ranking, Pluralism, and the Future of Heterodox Economics*. John Wiley & Sons.
- Levy, Moshe and Levy, Heim (2003). Investment talent and the pareto wealth distribution: Theoretical and experimental analysis. *Review of Economics and Statistics*, 85(3): 709–725.
- Kuhn, Thomas (1962): *The Structure of Scientific Revolutions*. Chicago: University of Chicago Press.
- Merton, Robert K. (1968): The Matthew Effect in Science. *Science*, 159(3810): 56-63.
- Mingers, John und Leydesdorff, Loet (2015): A Review of Theory and Practice in Scientometrics. *Working Paper*. URL: <http://arxiv.org/abs/1501.05462> (dl. 23.02.2015).
- Nature (2005): Editorials. *Nature*, 435:1003-1004.
- Newman, M.E.J. (2006): Power laws, Pareto distributions and Zipf's law. URL: <http://arxiv.org/abs/cond-mat/0412004> (dl. 10.12.2014).
- Reedijk, Jan und Moed, Henk F. (2006): Is the Impact of Journal Impact Factors Decreasing? *Journal of Documentation*, 64: 183–192.
- Rip, A. und Courtial, Jean-Paul (1984): Co-Words Maps in Biotechnology: An Example of Cognitive Scientometrics. *Scientometrics*, 6(6): 381-400.
- Rossner, Mike, Van Epps, Heather und Hill, Emma (2007). Show Me the Data. *Journal of Cell Biology*, 179: 1091–1092.
- Rothschild, Kurt W. (2008): Apropos Keynesianer. In: Hagemann, Harald, Horn, Gustav und Krupp, Hans-Jürgen (Hrsg.): *Aus gesamtwirtschaftlicher Sicht: Festschrift für Jürgen Kromphardt*. Marburg: Metropolis, S. 19-29.
- Seglen, Per O. (1996): Why the impact factor of journals should not be used for evaluating research. *British Medical Journal*, 314:497
- Shapiro, Carl und Varian, Hal R. (1999): *Information Rules: A Strategic Guide to the Network Economy*, Boston: Harvard Business School Press.
- Smith, Richard (2008): Beware the tyranny of impact factors. *Journal of Bone and Joint Surgery*, 90-B(2): 125-126.
- Solla-Price, Derek J. (1965): Networks of Scientific Papers. *Science*, 149: 510-515.
- Sterman, John D. und Wittenberg, Jason (1999): Path Dependence, Comeptition, and Success in the Dynamics of Scientific Revolution. *Organization Science*, 10(3): 322-341.
- Sydow, Jörg, Schreyögg, Georg und Koch, Jochen (2009): Organizational Path Dependence: Opening the Black Box. *Academy of Management Review*, 34: 689–709.

- Thomson Reuters (2014): Essential Science Indicators & Journal Citation Reports. URL:
<http://thomsonreuters.com/content/dam/openweb/documents/pdf/scholarly-scientific-research/fact-sheet/esi-jcr-brochure.pdf> (dl: 31.03.2015)
- Wilhite, Allen W. und Fong, Eric A. (2012): Coercive Citation in Academic Publishing. *Science*, 335:542-543.
- Willmott, Hugh (2011): Journal list fetishism and the perversion of scholarship: reactivity and the ABS list. *Organization* 18(4), 429-442.
- Young, Brigitte (2014): Financial Crisis: Causes, policy responses, future challenges. Directorate-General for Research and Innovation, European Commission, Brussels.