

ÖKONOMENSTIMME

Gut publizieren = gute Publikation?

<http://oekonomenstimme.org/a/532/>



Bruno S. Frey



Margit Osterloh

Die Qualitätsbeurteilung von Ökonomen im wissenschaftlichen Bereich geschieht vorwiegend über deren Publikationsliste. Je mehr Beiträge jemand in Zeitschriften mit hohem Impact Factor hat, umso besser für die Karriere. Doch sagt ein hoher Impact Factor wirklich etwas über die Qualität der Arbeiten aus?

Wenn sich heutzutage Ökonomen treffen, dreht sich nach kurzer Zeit das Gespräch unweigerlich um das Publizieren in wissenschaftlichen Fachzeitschriften. Dabei wird als selbstverständlich unterstellt, dass ein in einer „guten“ Zeitschrift veröffentlichter Artikel auch eine gute Publikation und einen wertvollen Beitrag zur ökonomischen Forschung darstellt. Wer in einer Top-Zeitschrift, also in einer A+ Zeitschrift wie dem Journal of Political Economy oder dem American Economic Review publiziert hat, findet unter Fachkollegen grosse Anerkennung.

Was eine „gute“ Zeitschrift ist, wird üblicherweise durch den Impact Factor bestimmt, d.h. wie oft im Durchschnitt Artikel in einer Zeitschrift im Zeitraum von zwei Jahren nach deren Veröffentlichung zitiert werden. Die entsprechenden Daten werden vom Institute for Scientific Information (ISI) Web of Knowledge erhoben. Der Impact Factor wird von ISI als verlässlich dargestellt: „a systematic and objective means to critically evaluate the world's leading journals“. Entsprechend wird eine wissenschaftliche Zeitschrift mit einem hohen Impact Factor als „gut“ taxiert, eine solche mit tiefem Impact Factor als wenig bedeutend. Diese Interpretation hat sich international durchgesetzt (z.B. Archambault und Larivière 2009, Jarwal, Brion und King 2009).

Diese Art der Qualitätsbewertung entscheidet heute massgeblich über die Verteilung der staatlichen Mittel und über die Karrieren von Wissenschaftlerinnen und Wissenschaftlern. An vielen Universitäten werden die Zeitschriften anhand des Impact Factors mit Punkten bewertet, so gibt zum Beispiel eine A+ Publikation 7 Punkte, eine A Publikation 4 Punkte, eine B Publikation 3 Punkte und eine C Publikation 1 Punkt. Eine Habilitation ist weitgehend Formsache, wenn entsprechend diesen Rankings genügend Punkte erreicht werden. Ganz ähnlich wird bei Berufungen auf Professuren vorgegangen. Einige Universitäten zahlen auch Geldbeträge für Publikationen in „guten“ Journals.

Extremer Fall von Trittbrettfahrern

Einer sorgfältigen empirischen Forschung folgend erweist sich die Vorstellung, eine Publikation in einer „guten“ Zeitschrift sei auch eine „gute“ Publikation, als verfehlt.

Der Grund ist die extrem ungleiche Zahl von Zitierungen von einzelnen Artikeln in einer Zeitschrift. Die mittels Zitaten gemessene Bedeutung eines wissenschaftlichen Organs hängt massgeblich von einigen wenige Beiträgen ab. Aus diesem Grund kann nicht von der durchschnittlichen Zahl von Zitaten – dem Impact Factor – auf die Qualität einzelner Beiträge geschlossen werden. Die meisten Artikel, die in einer als „gut“ gerankten Zeitschrift publiziert werden, profitieren als Trittbrettfahrer von den wenigen Artikeln, die eine grosse Zahl von Zitaten auf sich ziehen. Die meisten Artikel werden hingegen kaum, oder auch überhaupt nie zitiert. Rousseuw (1991, p.41, vgl. auch Starbuck 2005, p. 18, Kieser 2012, p. 99) drückt dies drastisch aus:

„It is commonly known and a constant cause of frustration that even wellknown journals contain a large fraction of bad articles which are boring, repetitive, incorrect, redundant and harmful to science in general.“

Wer auch nur eine Grundausbildung in Statistik genossen hat, weiss, dass bei einer stark schiefen Verteilung nicht auf Eigenschaften einzelner Werte geschlossen werden darf – und dennoch werden „gut“ publizieren und „gute“ Publikationen weitgehend gleich gesetzt und entscheiden über Auswahl, Karriere und Einkommen in der Wissenschaft.

Empirische Evidenz

Eine von der International Mathematical Union IMU (2008) in Auftrag gegebene Studie vergleicht zur Illustration zwei Zeitschriften, die Proceedings of the American Mathematical Society (mit einem Impact Factor von 0.434) und die Transactions of the American Mathematical Society (mit einem fast doppelt so hohen Impact Factor von 0.846). Die Wahrscheinlichkeit, dass ein zufällig ausgewählter Artikel in der ersten Zeitschrift öfter zitiert wird als in der weit „besseren“ zweiten Zeitschrift ist nicht weniger als 62%! Wir irren uns somit in 62% der Fälle. Ein zufällig ausgewählter Artikel in den „schlechten“ Proceedings ist in deutlich mehr als der Hälfte der Fälle mindestens ebenso gut oder sogar besser als ein zufällig ausgewählter Artikel in den „guten“ Transactions. „We are more often wrong than right“ (IMU 2008, p.11/2). Viele Leute überrascht diese Folgerung. Sie ist die Folge einer extrem schiefen Verteilung und eines kurzen zweijährigen Zeitfensters, was dazu führt, dass viele Artikel kaum oder nie zitiert werden.

Aus einem Editorial in der Zeitschrift Nature (2005, p.1003/4) wird deutlich, wie ungleich die Verteilung gerade auch in den von vielen als Vorbild betrachteten Naturwissenschaften ist. 89% des Impact Factors für das Jahr 2004 wurde durch gerade 25% der in diesem Jahr in Nature veröffentlichten Aufsätze generiert. Allein in diesem Jahr wurde ein Artikel über das Genom der Mäuse über 1000 Mal zitiert, hingegen wurde die grösste Zahl der anderen Aufsätze weniger als 20 Mal zitiert.

Speziell mit wirtschaftswissenschaftlichen Publikationen beschäftigt sich Oswald (2007). Er vergleicht zwei Top-Zeitschriften (American Economic Review AER, Econometrica) mit zwei als gut eingestuften Zeitschriften (Economic Journal EJ, Journal of Public Economics JPubEcs) und zwei als wenig wichtig eingeschätzten Zeitschriften (Journal of Industrial Economics JIE, Oxford Bulletin of Economics and Statistics OBES). Gemäss den Zitationszahlen des Web of Science sind die AER und die Econometrica an Stelle 1 und 2, EJ und JPubEcs an Stelle 9 und 16, und JIE und OBES an Stelle 47 und 51. Oswald nimmt die Daten für das Jahr 1981. Er kann damit die Zitierhäufigkeit über ein Vierteljahrhundert verfolgen und damit auch stark verzögerte Einschätzungen der Artikel berücksichtigen.

Die durchschnittliche Zahl von Zitierungen liegt wegen der extremen Schiefe der Verteilung wesentlich oberhalb des Medians für die zitierten Aufsätze.

Eine sehr kleine Zahl von Aufsätzen generiert die grösste Zahl der Zitate. Die zwei am häufigsten zitierten Beiträge waren in der AER mit 411 und in Econometrica mit 355 Nennungen. Viele Artikel wurden hingegen in den 25 Jahren überhaupt nie zitiert. 15 wurden nie oder nur einmal zitiert. In der AER, die mit 68 Zitaten an der Spitze der Zeitschriften liegt, sind mehr als ein Drittel im Verlauf des Vierteljahrhunderts weniger als 20 Mal von anderen Wissenschaftlern erwähnt worden.

Die falsche Zuordnung einer hohen Qualität eines Artikels auf Grund des Impact Factors einer Zeitschrift kann auf folgende Weise illustriert werden (vgl. auch Starbuck 2005). Im Median vereinigen die zwei Top-Zeitschriften AER und Econometrica etwa je 22 Zitate auf sich. Es lässt sich nun fragen, wie viele Artikel in den vier anderen Zeitschriften mehr Zitate generiert haben. Es sind 16%, oder beinahe ein Artikel aus fünf. Im EJ sind es 4 von 15 (27%), im JPubEcon 1 von 6 (17%). Nachträglich betrachtet haben somit die Gutachter und Editoren der als „top“ oder „gut“ eingestuften Zeitschriften signifikante Fehler begangen. Eine beträchtliche Zahl von Beiträgen von hoher Qualität wurde in weniger guten Zeitschriften, und eine solche mit geringer Qualität in besonders guten Zeitschriften veröffentlicht. Aus dem Publikationsort lässt sich somit nicht ohne erhebliche Fehler auf die Qualität eines Beitrages schliessen. Die gleiche Folgerung zieht die umfassende Studie von Laband und Tollison (2003), die nicht zitierte Aufsätze als „dry holes“ bezeichnet, sowie die aktuelle Studie von Baum (2010) für die wichtigsten Zeitschriften im Bereich Organisation.

Folgerungen

Inhaltlich gute Beiträge mit Publikationen in den besten Fachzeitschriften zu identifizieren, ist wegen der stark schiefen Verteilung der Zitierungen unstatthaft; in mehr als der Hälfte der Fälle ist diese Folgerung verfehlt – und trotzdem wird es heute vielfach getan. Der Fehler geht nicht auf eine unzureichende Messung zurück, sondern ist grundsätzlich. Jedem in der gleichen Zeitschrift veröffentlichten Artikel den gleichen Wert zuzuordnen verbirgt die extreme Ungleichheit in den Zitaten. Viele Artikel sind Trittbrettfahrer von einer kleinen Anzahl oft zitierter Beiträge, die den Impact Factor einer Zeitschrift bestimmen. Die Wissenschaft wird dadurch in mehrfacher Hinsicht geschädigt. Die Auswahl der zukünftigen Forscherinnen und Forscher wird verzerrt zugunsten derjenigen, die das Glück hatten, einen oder mehrere Aufsätze in Spitzenzeitschriften zu veröffentlichen. Das Abstellen auf Impact Factors kann in keiner Weise eine sorgfältige inhaltliche Analyse der einzelnen Beiträge ersetzen. Setzen die Entscheidungsträger – wie heute vielfach üblich – Publikationen in „guten“ Zeitschriften mit der Qualität der einzelnen Artikel gleich, werden zu einem erheblichen Ausmass ungeeignete Habilitanden und Bewerber auf Professuren ausgewählt und belohnt. Den Forschenden werden falsche Anreize gesetzt (vgl. Kieser 2012, Frey und Osterloh 2012).

Die harten Einschätzungen einer solchen Politik, etwa durch die International Mathematical Union („vague and misleading“ 2008, p. 12), sollten zu denken geben. Zumindest in Australien scheint ein Umdenken in Gang gekommen zu sein. Der National Health and Medical Research Council nennt die Bewertung von Beiträgen aufgrund des Impact Factors der Zeitschrift „unfair and un scholarly“ und verbietet deren Verwendung in Anträgen.

Ein solcher Lernprozess wäre auch der Ökonomik zu wünschen.

Literatur

Archambault, É. & Larivière, V. (2009). History of the journal impact factor: contingencies and consequences. *Scientometrics*, 79(3), 639-653.

Baum, J. A. C. (2010). Free-Riding on Power Laws: questioning the validity of the Impact Factor as a measure of research quality in organization studies. *Organization*, 18 (4), 449-466.

Frey, B. S. & Osterloh, M. (2012). [Rankings: Unbeabsichtigte Nebenwirkungen und Alternativen. Ökonomenstimme \(/artikel/2012/02/rankings-unbeabsichtigte-nebenwirkungen-und-alternativen/\)](#), 17. Feb. 2012.

Internationalen Mathematical Union IMU (2008). Citation Statistics. A report. Corrected version, 16/12/08.

Jarwal, S. D., Brion, A. M., & King, M. L. (2009). Measuring research quality using the journal impact factor, citations and 'Ranked Journals': blunt instruments or inspired metrics? *Journal of Higher Education Policy and Management*, 31(4), 289-300.

Kieser, A. (2012). JOURQUAL – der Gebrauch, nicht der Missbrauch, ist das Problem. Oder: Warum Wirtschaftsinformatik die beste deutschsprachige betriebswirtschaftliche Zeitschrift ist. *Die Betriebswirtschaft*, 72, 93-110.

Laband, D. N. & Tollison, R. D. (2003). Dry holes in economic research. *Kyklos*, 56, 161-174.

National Health and Medical Research Council (2010). NHMRC Removes Journal Impact Factors from Peer Review of Individual Research Grant and Fellowship Applications. Australian Government.

Nature (2005). Editorial: Not-so-deep impact. *Nature* 435, 1003-1004.

Oswald, A. J. (2007). An Examination of the Reliability of Prestigious Scholarly Journals: Evidence and Implications for Decision-Makers. *Economica*, 74, 21-31.

Rousseeuw, P. J. (1991). A diagnostic plot for regression outliers and leverage points. *Computational Statistics & Data Analysis*, 11(1), 127-129.

Starbuck, W. H. (2005). How much better are the most prestigious journals? The statistics of academic publication. *Organization Science*, 16, 180-200.

©KOF ETH Zürich, 16. Mai. 2013