Five Essays on the Quantification and Measurement of Intellectual Achievements

Dissertation

zur Erlangung des akademischen Grades eines

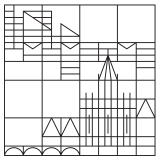
Doktors der Wirtschaftswissenschaften (Dr. rer. pol.)

vorgelegt von

Robert Hofmeister

an der

Universität Konstanz



Rechts-, Wirtschafts- und Verwaltungswissenschaftliche Sektion

Fachbereich Wirtschaftswissenschaften

Tag der mündlichen Prüfung: 13. Februar 2012

1. Referent: Prof. Dr. Heinrich W. Ursprung

2. Referent: Prof. Dr. Oliver Fabel

für Tina

Contents

List of Figures

List of Tables

Summary

This dissertation is a collection of five research papers written during my doctoral studies at the University of Konstanz between October 2007 and November 2011. The first four studies cover bibliometric topics whereas the last study deals with grading university students. All five studies contain empirical work, yet the main focus of the third and the fifth paper is theoretical. The first article analyses different aspects of the methodology of the *Handelsblatt Ranking* of the year 2007. The second article evaluates the research productivity of universities in business economics in Austria, Germany, and the German-speaking part of Switzerland. The third article presents a new method to evaluate research which provides a common basis for valuating research from different scientific disciplines. The fourth article investigates determinants of editing learned journals. Finally, the fifth paper presents a method to increase the precision of grades as a predictor of student ability. The paper also contains a case study. In the following I briefly summarize the main results.

Chapter 1 is a reprint of a joint article with Prof. Heinrich Ursprung (University of Konstanz). The article *Das Handelsblatt Ökonomen-Ranking 2007: Eine kritische Beurteilung* appeared in the *Perspektiven der Wirtschaftspolitik*, Vol. 9, No. 3, pp. 254-266, in 2008. We examine the methodology of the Handelsblatt Ranking which is currently the most recognized ranking of academic research in economics in Austria, Germany, and the German-speaking part of Switzerland and which is published on a yearly basis. Because it is frequently used as a research indicator the Handelsblatt ranking needs to be incentive compatible. We argue that (i) selecting a small set of journals instead of considering all journals, (ii) giving a bonus for co-authorship, and (iii) ignoring the length of an article distorts the research process and does not reflect of research achievements correctly. We analyze the latter for each of these aspects separately and show that the rankings of some researchers differ significantly depending on which method we use. We further show

that these differences persist also at the aggregate level of departments. Finally, we argue that the Handelsblatt should focus more on research output per professor than the total output of a department as the latter is highly driven by the number of department members and is not necessarily related to high average productivity.

Chapter 2 is joint work with Prof. Oliver Fabel (University of Vienna) and Miriam Henseler, née Hein (DFG). Our paper *Research Productivity in Business Economics: An Investigation of Austrian, German and Swiss Universities* was published in the *German Economic Review*, Vol. 9, No. 4, pp. 506-531, in the year 2008. We draw on a new dataset that collects the research output of business economists employed by Austrian, German and Swiss universities in spring 2008. The data set comprises publication records and personal data of roughly 1,800 scientists. We compute research rankings of departments and identify the leading departments in selected subdisciplines. Our results indicate that productivity differences between departments are relatively small and concentration of research output across departments is low. Using Tobit and Hurdle regressions, we investigate how institutional design and individual characteristics affect research productivity. We find that research productivity increases with department size as measured by the number of department members and with the number of a department's professors who actively publish. Moreover, productivity is higher in departments that run an economics study program. In line with the life cycle hypothesis we observe that the productivity of active researchers decreases with higher career age. Female business economists appear to be less productive than their male peers. It should be noted, that the paper has received considerable attention in the scientific community and gave rise to two comments. In two replies we show that our results are robust with regard to the points raised in these two comments.

Chapter 3 develops a new method to evaluating research and is also available as a working paper of the University of Konstanz. In this paper I adopt the notion that the ultimate aim of research is not to gain intellectual insights but to contribute to the well-being of mankind. Therefore I proposes a generational accounting approach to valuating research. Based on the flow of scientific results, a value-added (VA) index is developed that can, in principle, be used to assign a monetary value to any research result and, by aggregation, on entire academic disciplines or sub-disciplines. The basic idea of the VA-index is to distribute the value of all applications that embody research to the works

of research which the applications directly rely on, and further to the works of research of previous generations which the authors of the immediate reference sources have directly or indirectly made use of. Thereby a piece of research is valued at its final contribution to utility. The major contribution of the VA-index is to provide a measure of the value of research that is comparable across academic disciplines. To illustrate how the generational accounting approach works, I present a VA-based journal rating and a rating of the most influential recent journal articles in the field of economics.

Chapter 4 is an article written with Matthias Krapf (University of Vienna). The article *How Do Editors Select Papers, and How Good are They at Doing It?* appeared in *The B.E. Journal of Economic Analysis & Policy*, Vol. 11, Iss. 1 (Topics), Article 64. We use data on the B.E. Journals that rank articles into four quality tiers to examine the accuracy of the research evaluation process in economics. We find that submissions by authors with strong publication records and authors affiliated with highly-ranked institutions are significantly more likely to be published in higher tiers. Citation success as measured by RePEc statistics also depends heavily on the overall research records of the authors. Moreover, when controlling for the research topic as defined by JEL codes, we find that women receive significantly more citations than men. Finally and most importantly, we measure how successful the B.E. Journals' editors and their reviewers have been at assigning articles to quality tiers. While, on average, they are able to distinguish more influential from less influential manuscripts, we also observe many assignments that are not compatible with the belief that research quality is reflected by the number of citations.

Chapter 5 is the current version of a joint paper with Prof. Heinrich Ursprung (University of Konstanz) in which we develop a new method to standardize grades when not all students take the same exams. Our method uses regression analysis to control for differences in the difficulty of individual exams to arrive at more informative grades. Our approach relies on the idea that the difference between a student's ability and the received grade reflects the difficulty of the exam. We estimate the students' abilities via a system of equations, i.e. we derive the abilities endogenously. A key feature as compared to other methods to standardize grades is that we do not only compute a standardized grade point average (GPA) but also provide standardized grades at the level of individual exams. Using genuine examination data, we illustrate that our approach fares better than ECTS grading. Moreover, we show that grading standards differ significantly between different

exams and, arguably, different courses. If grades are not standardized in a sensible way, students are likely to choose soft courses. Standardizing grades eliminates such distortions and will lead to a more efficient selection of students into courses.

Chapter 1

Das Handelsblatt Ökonomen-Ranking 2007: Eine kritische Beurteilung

1.1 Einleitung

Dass Forschungsevaluationen unverzichtbare Führungsinstrumente darstellen, ist heutzutage praktisch unbestritten; zu kleineren Irritationen führt allenfalls noch der Umstand, dass mancherorts zu oft und zu wenig koordiniert evaluiert wird.[1] Was direkt betroffene Wissenschaftler jedoch immer wieder in Frage stellen, ist die *Qualität* von Evaluationen.[2] Da schlecht konzipierte Evaluationen nicht bloß ein Ärgernis für die direkt Betroffenen darstellen, sondern dem Wissenschaftssystem auch erheblichen Schaden zufügen können, hat der *Verein für Socialpolitik* vor zwei Jahren eine *Arbeitsgruppe Forschungsmonitoring* (AGFM) ins Leben gerufen. Die wichtigste Aufgabe der AGFM ist es, eine Forschungsdatenbank zu pflegen, mit deren Hilfe Forschungsevaluationen im Bereich der Wirtschaftswissenschaften überprüft und informiert kommentiert werden können.[3]

Anlässlich der Bayreuther Jahrestagung des *Vereins für Socialpolitik* haben wir in einer Paneldiskussion das *Handelsblatt Ökonomen-Ranking* 2006 kommentiert. Wir attestierten damals den Verantwortlichen, das mit Abstand beste privat (d.h. nicht im Wissenschaftssystem) erstellte Forschungsranking für die Volkswirtschaftslehre publiziert zu haben.[4] In der Zwischenzeit hat das *Handelsblatt* mit dem *Thurgauer Wirtschaftsinstitut*, das die Datenbank der AGFM verwaltet, eine Kooperation vereinbart. Diese Kooperation beinhaltet die Überprüfung ausgewählter VWL-Daten des *Handelsblatts* und die Erstellung eines äquivalenten Datensatzes für die Betriebswirtschaftslehre. Das *Handelsblatt Ökonomen-Ranking* 2007 beruht also auf Publikationsdaten, die mit dem Datensatz der AGFM abgeglichen worden sind. Das erste *Handelsblatt Betriebswirte-Ranking* wird mit der Zeit dieselbe Datenqualität aufweisen.

Die enorme Sichtbarkeit, verbunden mit der großen Akzeptanz bei den evaluierten Ökonomen (und vielleicht auch die Zusammenarbeit mit der AGFM) verleihen dem *Handelsblatt-Ranking* einen Stellenwert, der über die offenbar vorhandene ,,Newsworthiness" hinausgeht. Das *Handelsblatt-Ranking* wird insbesondere auch innerhalb desWissenschaftssys-

[1] Siehe z.B. Frey (2007).

[2] Siehe z.B. Ursprung (2003).

[3] Ebenfalls beunruhigt über die mangelnde Qualität breit publizierter Forschungsevaluationen hat der *Wissenschaftsrat* ein eigenes Forschungsrating von Universitäten und außeruniversitären Forschungseinrichtungen initiiert. Nach erfolgreichem Abschluss einer Pilotstudie sollen in einem rollierenden Turnus alle Fachgebiete abgedeckt werden. Zur ursprünglichen Empfehlung siehe http://www.wissenschaftsrat.de/texte/6285-04.pdf und zur laufenden Pilotstudie http://www.wissenschaftsrat.de/pilot start.htm.

[4] Eine Kritik am ersten *Handelsblatt Ökonomen-Ranking* vom Herbst 2005, das mit einer anderen Methode erstellt wurde, findet sich in Ursprung and Zimmer (2007).

tems rezipiert und übt so einen nicht zu unterschätzenden Einfluss auf die Profession aus. Gerade weil dem so ist, erscheint es uns wichtig, an dieser Stelle zu zeigen, dass das *Handelsblatt-Ranking*, trotz seiner anderweitigen Verdienste, kein geeignetes Führungsinstrument darstellt. Dabei wollen wir nicht auf die offensichtlichen Grenzen bibliometrischer Indikatoren eingehen, auf die das *Handelsblatt* ja auch immer wieder hinweist. Wir wollen in Abschnitt 1.2 vielmehr auf drei methodische Eigenheiten des *Handelsblatt-Rankings* hinweisen, die dysfunktionale Anreize zur Folge haben könnten, wenn der zentrale *Handelsblatt*-Indikator zum allgemeinen Evaluationsstandard erhoben würde. In Abschnitt 1.3 präsentieren wir den von der AGFM präferierten Indikator der Forschungsproduktivität und vergleichen dann in Abschnitt 1.4 die beiden Methoden anhand des so genannten Lebenswerk-Ranking. Dabei zeigt sich, dass die Evaluationsmethode des *Handelsblatts* zu verzerrten Messungen der Forschungsproduktivität führt, d.h. die Produktivität gewisser Forschertypen wird systematisch über- bzw. unterschätzt.

1.2 Zur Anreizinkompatibilität des Handelsblatt Ökonomen-Rankings

1.2.1 Die Zeitschriftenauswahl

Das *Handelsblatt Ökonomen-Ranking* 2007 setzt Forschungsoutput mit Artikeln gleich, die in einer Auswahl von 220 besonders angesehenen Fachzeitschriften erschienen oder zur Publikation angenommen worden sind. Das *Handelsblatt* erfasst also nur eine relativ kleine *Auswahl* von einschlägigen wissenschaftlichen Zeitschriften. Nun führt aber jede noch so wohlüberlegte Auswahl leicht zu Verhaltensänderungen, die weder beabsichtigt noch erstrebenswert sind.

Unschön ist also nicht nur der Umstand, dass der Forschungsoutput unvollständig und deshalb unter Umständen verzerrt erfasst wird; mit einer derartigen Auswahl nimmt man auch gewissen Ökonomen die Anreize, sich mit Problemen auseinanderzusetzen, die z.B. nur von lokalem Interesse sind und somit bloß in ,,minor journals" publiziert werden können. Darüber hinaus werden auch all diejenigen ,,journeymen" Forscher demotiviert, die das Handwerk der ökonomischen Forschung zwar gut beherrschen, deren Arbeit aber nicht die von renommierten Zeitschriften verlangte hohe Originalität aufweist. Ein Indikator, der aber nur einen speziellen Teil der ökonomischen Forschung erfasst, kann nicht der

ganzen Profession als Messlatte dienen.

Ein allgemein einsetzbarer Forschungsindikator muss notwendigerweise eine Vollerhebung des Forschungsoutputs anstreben. Dieses Ziel ist zugegebenermaßen nie wirklich erreichbar, da die Grenze zwischen Forschung und Anwendung wissenschaftlicher Methoden fließend ist und mit rein bibliometrischen Mitteln ohnehin nicht präzise identifiziert werden kann. Das Kriterium der Begutachtung durch anonyme Sachverständige ist jedoch wenig kontrovers und hat sich auch deshalb etabliert, weil es bibliometrisch relativ leicht umsetzbar ist. Für die Ökonomie bedeutet dies, dass man Forschung mit Publikationen in ökonomischen Fachzeitschriften mit Begutachtungsverfahren gleichsetzen kann; und da die von *EconLit* auferlegten Minimalstandards dem Begutachtungs-Kriterium recht nahe kommen, hat sich in der Bibliometrie die von *EconLit* erfasste Zeitschriftenliteratur als Standard durchgesetzt.

Die über 750 Zeitschriften, die heutzutage von *EconLit* erfasst werden, umfassen praktisch den gesamten Kernbereich der ökonomischen Forschung. Für gewisse Subdisziplinen, die an den Schnittstellen der Ökonomie und einer anderen Disziplin angesiedelt sind, ist der *EconLit*-Standard aber nicht befriedigend. Um den Forschungsoutput von Ökonometrikern besser abzubilden zu können, erfasst das *Handelsblatt* deshalb neuerdings auch Zeitschriften im Bereich der Statistik. Neben der Ökonometrie gibt es aber noch weitere Subdisziplinen, in denen häufig in nicht-ökonomischen Zeitschriften publiziert wird. Zu nennen sind hier z.B. die Politische Ökonomie, die Wirtschaftsgeschichte und die Verhaltensökonomik.

1.2.2 Die Koautorengewichtung

Die vom *Handelsblatt* verwendete Koautorengewichtung, die jedem der n Autoren eines Artikels den Anteil $2/(n+1)$ gutschreibt, ist aus verschiedenen Gründen problematisch. Im Vordergrund steht hier die mit dieser Gewichtung einhergehende Anreizinkompatibilität. Wenn ein von vielen Autoren verfasster Artikel bis zu zweimal mehr Gewicht erhält als ein von einem einzigen Autor verfasster Artikel, so werden Anreize für Koautorschaften geschaffen, ob dies im gegebenen Fall nun sinnvoll ist oder nicht. Im Extremfall kann diese Gewichtung sogar zu fingierten Koautorschaften führen, die im Logrollingverfahren abgesprochen werden, oder zu Koautorschaften ganzer Forschungseinheiten im Stil naturwissenschaftlicher Labors. Die Standardmethode der Koautorengewichtung ($1/n$ anstatt

$2/(n+1)$) weist keine derartigen dysfunktionalen Anreize auf. Darüber hinaus trägt die Standardmethode auch dem Umstand Rechnung, dass in den verschiedenen Subdisziplinen der Ökonomie womöglich eine unterschiedliche Anzahl von Autoren optimal und üblich ist: Wenn es nämlich derartige Unterschiede gibt, so zieht die Anrechnungsmethode des *Handelsblatts* systematische Verzerrungen nach sich.[5]

1.2.3 Die (vernachlässigte) Gewichtung des Umfangs von Forschungsresultaten

Man mag unsere Befürchtung, dass die Koautorengewichtung des *Handelsblatts* missbraucht werden könnte, vielleicht für übertrieben halten. Unser letzter Einwand, der sich gegen die Vernachlässigung jeglicher Gewichtung des Umfangs der erfassten Zeitschriftenartikel wendet, stützt sich jedoch auf ein Missbehagen, das in der Profession weit verbreitet ist. Wenn die (qualitätsgewichtete) Anzahl der Publikationen ganz allgemein als valider Indikator der Forschungsproduktivität interpretiert wird, so verleitet dies dazu, den Forschungsoutput in ,,kleinste publizierbare Einheiten" aufzuteilen. Abgesehen davon, dass eine derartige Fraktionierung die Kommunikation und Rezeption von Forschungsergebnissen behindert und den Begutachtungsprozess der Fachzeitschriften unnötig belastet, führt diese Bemessungsgrundlage auch zu verzerrten Messungen, weil eine Fraktionierung der Forschungsresultate nicht in allen Subdisziplinen im gleichen Umfang möglich ist.

Um dem Unwesen der Balkanisierung der Forschungsergebnisse zu begegnen, bietet es sich an, bei der Messung des Forschungsoutputs die Seitenzahl der Artikel zu berücksichtigen. Die Seitenzahl kann relativ einfach erhoben werden und reflektiert, weil der Zeitschriftenplatz beschränkt ist, die Einschätzung der Bedeutung der jeweiligen Artikel in den Augen der Herausgeber.[6] Diese Längenunterschiede sind nicht unbeträchtlich. So sind z.B. fast ein Drittel aller AER-Artikel, die von in Deutschland tätigen Ökonomen verfasst wurden, relativ kurz (bis zu 10 Seiten), währenddem doch etwa 20% mehr als 20 Seiten aufweisen.

Obwohl die meisten bibliometrischen Studien mit einer Seitenzahl-Gewichtung ar-

[5] Es ist richtig, dass Artikel, die von mehreren Autoren verfasst worden sind, ceteris paribus eine höhere Qualität aufweisen (siehe z.B. Rauber and Ursprung, 2008). Dies sollte aber nicht in der Koautorengewichtung berücksichtigt werden, da dieser Effekt ja schon im Qualitätsgewicht der publizierenden Zeitschrift zum Ausdruck kommt.

[6] Dass diese Einschätzung der Herausgeber durchaus fundiert ist, kann man daraus ersehen, dass längere Zeitschriftenartikel (cetris paribus) öfter zitiert werden als kürzere. Eine entsprechende Untersuchung findet sich in Ursprung and Zimmer (2007).

beiten, sollte diese Art der Gewichtung aber bloß als grober Indikator der Bedeutung betrachtet werden. Verschiedene Verfeinerungen sind denn auch in der bibliometrischen Literatur vorgeschlagen worden. So wird z.B. die Anzahl der Schriftzeichen pro Artikel verwendet, um unterschiedliche Seitenlayouts abzubilden, oder die Seitenzahl im Vergleich mit anderen Aufsätzen in der jeweiligen Zeitschrift bzw. Subdisziplin, um zeitschriftenspezifische oder subdisziplinspezifische Gepflogenheiten zu berücksichtigen.[7]

1.3 Die Standardmethode der Messung des Forschungsoutputs

Die Standardmethode der Messung des Forschungsoutputs, die mit den im vorhergehenden Abschnitt diskutierten Nachteilen nicht behaftet ist, beruht auf der folgenden Definition des Forschungsoutputs F_i von Forscher i:

$$F_i = \sum_{k \in EL} \frac{w_k p_k}{n_k}. \tag{1.1}$$

Dabei bezeichnet w_k die Qualität der Zeitschrift, in der Artikel k von Forscher i erschienen ist, p_k die Seitenzahl des Artikels und n_k die Anzahl der Autoren. Die Summe umfasst alle von *EconLit* (EL) erfassten Zeitschriftenartikel. Die entsprechende vom *Handelsblatt* verwendete Formel sieht folgendermaßen aus:

$$H_i = \sum_{k \in HB} \frac{2\hat{w}_k}{n_k + 1}, \tag{1.2}$$

wobei die Summe nun über alle Artikel k läuft, die in einer vom *Handelsblatt* (HB) erfassten Zeitschrift veröffentlicht oder zur Veröffentlichung angenommen worden sind.[8] Man beachte, dass die Zeitschriftenauswahl (HB) des *Handelsblatts* im Vergleich zur gesamten *EconLit* Literatur (EL) relativ klein ist. Die im folgenden Abschnitt betrachteten Ökonomen haben z.B. 8130 Artikel in 527 verschiedenen *EconLit* Zeitschriften

[7] Wenn man die Länge der Artikel mit der in der entsprechenden Zeitschrift üblichen Artikellänge vergleicht, so vermeidet man damit auch einen möglichen Fehler. Man kann nämlich einwenden, dass die Länge von Aufsätzen schon in der Qualitätsgewichtung der Zeitschrift enthalten ist, wenn die Qualitätsgewichtung den ,,impact" der Zeitschrift wiederspiegelt und der ,,impact" wiederum von der Länge abhängt.

[8] Die Zeitschriftengewichte wk in Formel 1.1 hat die AGFM bisher der Studie von Combes and Linnemer (2003) entnommen. Das *Handelsblatt* orientiert sich bei seiner Zeitschriftengewichtung $\hat{w}_k$ (Formel 1.2) ebenfalls weitgehend an dieser Studie. Wir gehen hier nicht auf den offensichtlichen Einfluss der Qualitätsgewichtung der Zeitschriften ein und lassen die Frage der adäquaten Gewichtung offen (siehe dazu auch die Beiträge von Klaus Ritzberger und Günther Schulze, Warning, and Wiermann in diesem Band).

veröffentlicht; davon werden jedoch nur 4365 Artikel in 168 Zeitschriften vom *Handelsblatt* erfasst.

Bei einer automatisierten Datenerhebung fallen keine beträchtlichen zusätzlichen Kosten an, wenn man die Methode des *Handelsblatts* durch die von der AGFM präferierten Standardmethode ersetzt. Die Koautorengewichtung kann man praktisch durch Knopfdruck verändern, und der *EconLit* Datensatz enthält alle Informationen über die vom *Handelsblatt* bisher nicht erfassten Zeitschriften und Seitenzahlen. Ein Wermutstropfen besteht allerdings darin, dass man die im Erscheinen begriffenen Arbeiten mit der Methode der AGFM nicht ohne weiteres berücksichtigen kann, weil die entsprechenden Seitenzahlen eben noch nicht bekannt sind.

Im vorhergehenden Abschnitt haben wir darauf hingewiesen, dass man mit der in Formel (1.2) zusammengefassten Methode des *Handelsblatts* neben den unerwünschten Anreizwirkungen auch Verzerrungen bei der Messung des Forschungsoutputs in Kauf nimmt. Wenn man zeigen könnte, dass diese Verzerrungen geringfügig sind, so fielen die Vorteile der doch etwas aufwändigeren Standardmethode (Formel 1.1) vielleicht nicht wirklich ins Gewicht. Dies gilt allerdings nur solange man lediglich die gewachsene Forschungslandschaft abbilden will und niemand den *Handelsblatt*-Indikator als Führungsinstrument missbraucht. Im nächsten Abschnitt vergleichen wir konkrete „Lebenswerk-Rankings", die mit der Standardmethode und Elementen der *Handelsblatt*-Methode erstellt wurden, und zeigen, dass die Verzerrungen, die man sich mit der *Handelsblatt*-Methode einhandelt, doch erheblich sind.

1.4 Das Handelsblatt Ranking auf dem Prüfstand

1.4.1 Das „Lebenswerk" Ranking

In Tabelle 1.1 präsentieren wir das nach der Standardmethode (Formel 1.1) erstellte Lebenswerk-Ranking der an deutschen Universitäten tätigen Volkswirte, deren Lebenswerk gemäß dem von Combes und Linnemer vorgeschlagenen Gewichtungsschema mehr als 100 AER-äquivalente Seiten aufweist ($F_i > 100$). Man beachte, dass dieses Ranking im Vergleich mit dem *Handelsblatt-Ranking*, das ja auch akzeptierte, aber noch nicht publizierte Artikel erfasst, etwas weniger aktuell ist.Wir ziehen das Lebenswerk-Ranking zur Illustration der methodischen Unterschiede heran, weil das *Handelsblatt* mit diesem Ranking

Tabelle 1.1: ,,Lebenswerk" Ranking

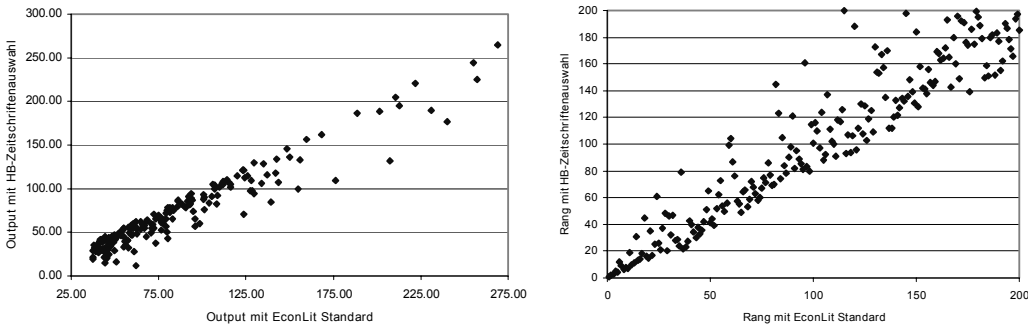
Rang	Name	Lebenswerk -Output F	Rang	Name	Lebenswerk -Output F
1	Hellwig, Martin	501,54	29	Funke, Michael	128,78
2	Sinn, Hans-Werner	362,01	30	Rees, Ray	128,1
3	Güth, Werner	257,53	31	Franz, Wolfgang	127,56
4	Konrad, Kai A.	255,15	32	Egger, Peter	125,79
5	von Hagen, Jürgen	239,94	33	Ursprung, Heinrich W.	124,31
6	Börsch-Supan, Axel	230,97	34	Janeba, Eckhard	123,93
7	Schmidt, Klaus	221,67	35	Belke, Ansgar	123,67
8	Ebert, Udo	212,88	36	Oechssler, Jörg	123,36
9	Bester, Helmut	210,47	37	Strausz, Roland	123,25
10	Vaubel, Roland	207,63	38	Wieland, Volker	119,91
11	Snower, Dennis	201,61	39	Pethig, Rüdiger	116,06
12	Moldovanu, Benny	188,83	40	Dawid, Herbert	115,13
13	Backhaus, Jürgen G.	176,14	41	Eichberger, Jürgen	114,07
14	Schweizer, Urs	168,14	42	Walz, Uwe	112,04
15	Bell, Clive	159,82	43	Schnitzer, Monika	111,67
16	Kraft, Kornelius	155,94	44	Röller, Lars-Hendrik	111,61
17	Komlos, John	155,04	45	Härdle, Wolfgang[a]	109,78
18	Gerlach, Stefan	150,22	46	Peitz, Martin	108,76
19	von Thadden, E.-L.	148,33	47	Buch, Claudia	108,38
20	Burda, Michael C.	143,54	48	Falk, Armin	106,97
21	Corneo, Giacomo	142,46	49	Riphahn, Regina T.	106,88
22	Zimmermann, Klaus F.	141,81	50	Puppe, Clemens	105,75
23	Endres, Alfred	139,26	51	Fuest, Clemens	105,35
24	Witt, Ulrich	137,06	52	Breyer, Friedrich	103,88
25	Schmitz, Patrick	134,97	53	Welsch, Heinz	101,24
26	Feld, Lars P.	134,1	54	Bolle, Friedel	100,97
27	Wagner, Joachim	129,76	55	Leininger, Wolfgang	100,76
28	Krebs, Tom	129,42	56	Klasen, Stephan	100,54

Quelle: Eigene Berechnungen.

[a] Wolfgang Härdle ist der einzige Ökonometriker, der auf dieser Liste auftaucht. Er und seine Kollegen werden von *EconLit* basierten Rankings unterbewertet.

seine ,,Top-Ökonomen" identifiziert und nicht etwa weil wir der Ansicht wären, dass derartige Lebenswerk-Rankings wirklich eine große Bedeutung hätten. Das Lebenswerk ist bestenfalls ein Maß für den Erfahrungsschatz eines Wissenschaftlers. Wenn man aus dem Lebenswerk Rückschlüsse auf die Leistungsfähigkeit ziehen will, so muss man entweder das Karrierealter in Rechnung stellen, so wie dies das *Handelsblatt* in seinem ,,Punkte pro Jahr" Ranking tut oder, besser noch, den Leistungsvergleich auf eine Alterskohorte beschränken, um den Einfluss des Alters adäquat zu berücksichtigen (siehe dazu Rauber and Ursprung, 2008).

Die von der AGFM präferierte Standardmethode (Formel 1.1) dient unserer Analyse als Vergleichsbasis. Davon ausgehend untersuchen wir den Einfluss der drei monierten Eigenheiten der *Handelsblatt*-Methode.Wir vergleichen also nicht das vom *Handelsblatt* publizierte Lebenswerk-Ranking mit alternativen Rankings. Wir nehmen vielmehr das in Tabelle 1.1 (trunkiert) präsentierte ,,Standardranking" und zeigen wie es sich verändert,



Quelle: Eigene Berechnungen.

Abbildung 1.1: Zeitschriftenauswahl, Streudiagramme a und b

wenn man nacheinander die Gewichtungen der Standardmethode durch die monierten Gewichtungen der *Handelsblatt*-Methode ersetzt. Der Vorteil dieses Vorgehens besteht darin, dass wir dazu nur den AGFM Datensatz benötigen.

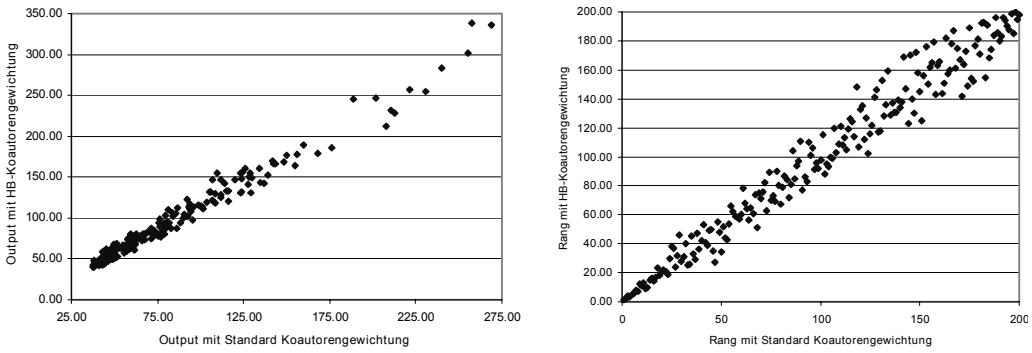
In einem ersten Schritt beschränken wir die Publikationserfassung auf diejenigen *EconLit*-Zeitschriften, die vom *Handelsblatt* ausgezählt werden.[9] Wir versehen also alle *EconLit*-Zeitschriften, die nicht in der *Handelsblatt*-Auswahl enthalten sind, mit dem Gewicht 0.

Die Streudiagramme a und b (Abbildung 1.1) zeigen wie sich beim Übergang von der ,,*EconLit* Vollerhebung" zur HB-Auswahl der gemessene Forschungsoutput und die sich daraus ergebenden Rangplätze ändern. Die ganz erheblichen Rangverschiebungen im Vergleich zur Vollerhebung weisen darauf hin, dass die Stichprobe des *Handelsblatts* in der Tat Verzerrungen zur Folge hat. Die Beschränkung auf die *Handelsblatt*-Auswahl führt zu individuellen Rangpunkteverbesserungen von bis zu 20% und zu noch sehr viel größeren Einbußen. Darüber hinaus zeigt sich, dass eben auch international durchaus sichtbare deutsche Volkswirte (alle Streudiagramme beschränken sich auf die Top-200 Ökonomen des Standard-Ranking)[10] einen erheblichen Teil ihres Oeuvres in ,,minor journals" publizieren.

Im zweiten Schritt analysieren wir den Einfluss der Koautorengewichtung. Die Streudiagramme a und b in Abbildung 1.2 dokumentieren wie sich die doch sehr eigenwillige Koautorengewichtung des *Handelsblatts* auf den gemessenen Forschungsoutput und die Rangierung der ,,Top-Ökonomen" auswirkt. Die Verzerrungen sind hier geringer als beim ersten Vergleich, aber immer noch durchaus beträchtlich. Eine genauere Betrachtung der

[9] Man beachte, dass wir die HB-Auswahl nicht genau reproduzieren können, weil die Datenbank der AGFM vorderhand nur die *EconLit*-Zeitschriften umfasst.

[10] Die beiden führenden Ökonomen fallen allerdings bei den Output-Abbildungen ,,aus dem Rahmen".

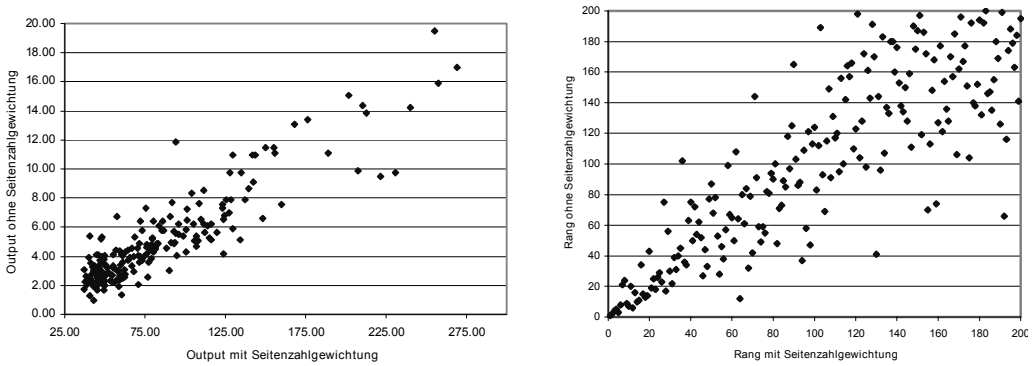


Quelle: Eigene Berechnungen.

Abbildung 1.2: Koautorengewichtung, Streudiagramme a und b

Datenpunkte bestärkt uns in der Vermutung, dass z.B. experimentelle Ökonomen, die regelmäßig mit mehreren Koautoren zusammenarbeiten, von der *Handelsblatt*-Methode profitieren, wohingegen Ökonomen, die unorthodoxe Forschungsstrategien verfolgen und somit eher als wissenschaftliche Einzelkämpfer auftreten, von der *Handelsblatt*-Methode unterbewertet werden.

Die markantesten Änderungen ergeben sich jedoch bei unserem letzten Vergleich, in dem wir wie das *Handelsblatt* auf jegliche Gewichtung des Umfangs der Forschungsergebnisse verzichten, d.h. wir geben die Seitenzahl-Gewichtung auf. Die Bedeutung der Seitenzahlgewichtung kommt in den Streudiagrammen a und b (Abbildung 1.3) sehr gut zum Ausdruck. Es gibt offenbar Ökonomen, die regelmäßig Forschungsergebnisse publizieren, deren Präsenation die regelmäßig Forschungsergebnisse publizierenverlangt und rechtfertigt, währenddem andere sich auf punktuelle Probleme spezialisieren, deren Gesamtzusam-



Quelle: Eigene Berechnungen.

Abbildung 1.3: Seitenzahl-Gewichtung, Streudiagramme a und b

Tabelle 1.2: Evaluation von Fachbereichen

Gesamtoutput		Durchschnittsproduktivität der Professoren	
Handelsblatt	Standardmethode	*Handelsblatt*	Standardmethode
1 Bonn	Bonn	Bonn	Bonn
2 LMU	LMU	Mannheim	LMU
3 Mannheim	Mannheim	Frankfurt/M	Oldenburg
4 Frankfurt/M	Frankfurt/M	Kiel	Mannheim
5 HU Berlin	FU Berlin	LMU	Hannover
6 Köln	Hannover	HU Berlin	Heidelberg
7 FU Berlin	HU Berlin	Heidelberg	Kiel
8 Kiel	Kiel	Bielefeld	Tübingen
9 Bielefeld	Köln	Dortmund	Frankfurt/M
10 Hannover	Heidelberg	Konstanz	Konstanz
11 Heidelberg	Bielefeld	Oldenburg	FU Berlin
12 Konstanz	Hamburg	Köln	Bielefeld
13 Dortmund	Tübingen	FU Berlin	HU Berlin
14 Hamburg	Konstanz	Hannover	Marburg
15 Karlsruhe	Erfurt	Tübingen	Frankfurt/O

menhang wohlbekannt ist und die deshalb relativ kurz abgehandelt werden können.

1.4.2 Fachbereichsrankings

Die Eigenheiten bibliometrischer Evaluationsmethoden machen sich natürlich auch im Aggregat, d.h. bei der Evaluation gesamter Fachbereiche bemerkbar. Tabelle 1.2 vergleicht das vom *Handelsblatt* an erster Stelle aufgeführte Fachbereichsranking, dem der gesamte Forschungsoutput (seit 1997) der gegenwärtigen Mitarbeiter zugrunde liegt, mit dem entsprechenden Ranking, das mit der Standardmethode erstellt wurde. Wir beschränken uns dabei auf eine Top-15 Rangliste der deutschen Fachbereiche. Weil der gesamte Forschungsoutput eines Fachbereichs natürlich ganz wesentlich von der Anzahl der ihm angehörenden Mitarbeiter abhängt, birgt dieses Ranking-Kriterium die Gefahr in sich, falsch interpretiert zu werden. In mancherlei Hinsicht aufschlussreicher sind Produktivitätsmaße, die eben nicht direkt von der Größe der Fachbereiche abhängen. Wir reproduzieren deshalb auch das Produktivitätsranking des *Handelsblatts* (Punkte pro Professor) und stellen es dem entsprechenden Ranking gegenüber, das mit der Standardmethode erstellt wurde. Während beim Ranking nach dem Gesamtoutput die Rangordnung nicht allzu stark von der verwendeten Methode abhängt, ergeben sich beim Produktivitätsranking doch große Unterschiede.

Nicht überraschend ist, dass sich die Ranglisten nach Gesamtoutput und Forschungsproduktivität stark unterscheiden. Es ist wohlbekannt, dass man umfassende Evaluationen nicht auf ein einziges Kriterium abstellen kann. Obwohl der ,,default mode" der *Handels-*

blatt-Homepage die Fachbereiche nach dem problematischen Kriterium ,,Gesamtoutput" auflistet, weist das *Handelsblatt* in vorbildlicherWeise weitere Indikatoren aus. Neben der schon angeführten Forschungsproduktivität geben diese zusätzlichen Indikatoren Hinweise auf die Qualität der Forschung (Top-5 Punkte) und auf die Streuung der Forschungsproduktivität innerhalb der jeweiligen Fachbereiche (Anteil des forschungsstärksten Professors an der Gesamtpunktezahl). Eine derartige mehrdimensionale Evaluation ist sehr nützlich und sinnvoll, weil die Interessen der Nutzer von Rankings eben sehr unterschiedlich sein können.

Obwohl gerade die Mehrdimensionalität des *Handelsblatt*-Fachbereichsranking lobend zu erwähnen ist, lassen sich die Aussagekraft des Qualitäts- und Streuungsindikators und die Präsentation der facettenreichen Information sicherlich noch verbes- sern. Wir schlagen in Abbildung 1.4 eine Präsentationsform vor, in der die vielleicht vier wichtigsten Indikatoren gleichberechtigt nebeneinander stehen und zudem auf einen Blick erfasst werden können. Alle vier Indikatoren wurden mit Hilfe der Standardmethode berechnet und beziehen sich auf alle Mitarbeiter, d.h. es wird auch bei der Berechnung der Produktivität (und somit anders als im Produktivitätsranking in Tabelle 1.2) der Forschungsoutput aller promovierten Mitarbeiter erfasst. Auf der Abszisse ist die durchschnittliche Produktivität der einem Fachbereich angehörenden Mitarbeiter abgetragen, auf der Ordinate die Durchschnittsqualität des Forschungsoutputs. Die Fläche der ,,Fachbereichsbubbles" entspricht dem Gesamtoutput (seit 1997) und die Länge der Balken ist proportional zur Standardabweichung der Forschungsproduktivitäten der einem Fachbereich angehörenden Mitarbeiter. Es werden die 15 nach dem Gesamtoutput größten Fachbereiche abgebildet.

Aus Abbildung 1.4 kann man z.B. entnehmen, dass die Fachbereichsgröße keinen Einfluss auf die Effizienz der Forschungsproduktion zu haben scheint. Sowohl der große Fachbereich der Universität Bonn, wie auch der mittelgroße Fachbereich der Universität Heidelberg und der kleine Fachbereich der Universität Oldenburg produzieren an der offenbarten Effizienzgrenze von Produktivität und Qualität. Die Belegschaft dieser Fachbereiche ist allerdings sehr heterogen. Homogener sind z.B. die ebenfalls recht effizienten Fachbereiche der Universitäten Frankfurt a.M. und Konstanz.

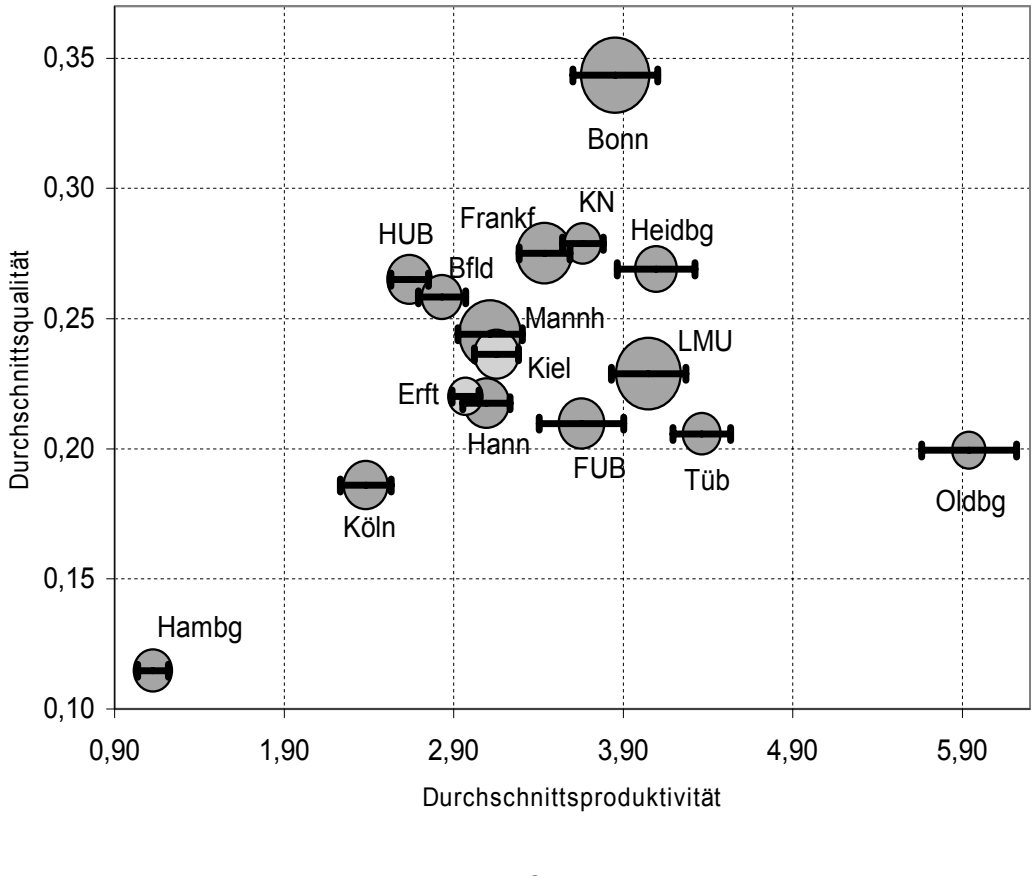


Abbildung 1.4: Fachbereichsranking

1.5 Schluss

Ziel unserer Ausführungen war es zu zeigen, dass das in seiner Art ausgezeichnete *Handelsblatt Ökonomen-Ranking* 2007 nicht wirklich anreizkompatibel ist und somit nicht als Führungsinstrument gebraucht werden sollte. Die vom *Handelsblatt* in den letzten zwei Jahren verwendete Methode führt darüber hinaus auch zu erheblich verzerrten Messungen der Forschungsproduktivität. Wir belegen diese Verzerrungen indem wir zeigen, wie sich ein konkretes Ranking ändert, wenn man die bibliometrische Standardmethode, die von der *Arbeitsgruppe Forschungsmonitoring* des *Vereins für Socialpolitik* präferiert wird, durch Elemente der *Handelsblatt*-Methode ersetzt. Es ging uns hier aber nicht darum, einen eigenen VfS-Standard zu entwerfen bzw. zu vertreten. Wir hoffen allerdings, dass ein entsprechender Diskurs mit der Zeit dazu führt, dass sich ein derartiger VfS-Standard herauskristallisiert.

Abschließend kann man festhalten, dass es dem *Handelsblatt* gelungen ist, das wegweisende Ranking aus dem Jahr 2006 weiter zu verbessern. Das neueste *Handelsblatt Ökonomen-Ranking* stützt sich auf verlässlichere Daten, ist differenzierter angelegt und wird einer der interdisziplinär ausgerichteten Subdisziplinen, nämlich der Ökonometrie, besser gerecht. Die Profession darf sich glücklich schätzen, dass ihre Forschungsleistungen von einem derart kompetenten privaten Anbieter einer breiten Öffentlichkeit vorgestellt werden. Wir haben hier auf die Grenzen dieses Rankings hingewiesen und Verbesserungsmöglichkeiten aufgezeigt, die es natürlich immer gibt. Wie das *Handelsblatt* darauf reagieren will, bleibt ihm überlassen. Die Profession der Volkswirte wartet auf alle Fälle gespannt auf das *Handelsblatt Ökonomen-Ranking* 2008.

Literaturverzeichnis

Combes, P.-P., and L. Linnemer (2003): "Where are the Economists Who Publish? Publication Concentration and Rankings in Europe Based on Cumulative Publications," *Journal of the European Economic Association*, 1(6), 1250–1308.

Frey, B. S. (2007): "Evaluierungen, Evaluierungen H Evaluitis," *Perspektiven der Wirtschaftspolitik*, 8(3), 207–220.

Rauber, M., and H. W. Ursprung (2008): "Evaluation of Researchers: A Life Cycle Analysis of German Academic Economists," in *Conferences on New Political Economy*, ed. by M. Albert, S. Voigt, and D. Schmidtchen, vol. 25 of *Conferences on New Political Economy*, pp. 101–122(2). Mohr Siebeck, Tübingen.

Ritzberger, K. (2008): "Eine invariante Bewertung wirtschaftswissenschaftlicher Fachzeitschriften," *Perspektiven der Wirtschaftspolitik*, 9(3), 267–285.

Schulze, G. G., S. Warning, and C. Wiermann (2008): "Zeitschriftenrankings für die Wirtschaftswissenschaften - Konstruktion eines umfassenden Metaindexes," *Perspektiven der Wirtschaftspolitik*, 9(3), 286–305.

Ursprung, H. W. (2003): "Schneewittchen im Land der Klapperschlangen: Evaluation eines Evaluators," *Perspektiven der Wirtschaftspolitik*, 4(2), 177–190.

Ursprung, H. W., and M. Zimmer (2007): "Who is the "Platz-Hirsch" of the German Economics Profession? A Citation Analysis," *Journal of Economics and Statistics (Jahrbuecher fuer Nationaloekonomie und Statistik)*, 227(2), 187–208.

Chapter 2

Research Productivity in Business Economics: An Investigation of Austrian, German and Swiss Universities

2.1 Introduction

The international exposure of economic research in continental Europe has certainly increased over the last two decades. This development has been accompanied by a growing interest in comparative evaluations of research institutions. Most of these evaluations have, however, focused on 'proper' economics (defined as the research program envisaged by classical political economists). Representative studies include Clemenz and Neusser (1991) for Austria, Combes and Linnemer (2001) for France, Guimarães (2002) for Portugal, Dolado, García-Romero, and Zamarro (2003) and Rodríguez (2006) for Spain, Cainelli, de Felice, Lamonarca, and Zoboli (2006) for Italy, Hein (2006) for Switzerland, Turnovec (2007) for the Czech Republic, and Rauber and Ursprung (2008a) for Germany. Some momentous ranking studies covering Europe as a whole have been published in a special issue of the Journal of the European Economic Association (2003).

The much younger subdiscipline of business economics has yet received very little attention. Clearly, this discipline that deals with the application of economic principles to firms or other management units attracts considerable public, commercial and academic interest - reflected, for example, in the growing number of professorships in business administration and the starting salaries of graduates. However, apart from Fabel and Heße (1999) we are not aware of any studies that evaluate research performance in this field. The above-mentioned ranking studies either do not consider this research at all or it is mingled with publications from the various subdisciplines of economics. However, due to differences in publication and citation cultures, blending across disciplines causes comparability problems.

In our study we therefore focus on research in the field of business economics, which, in our understanding, includes the subdiscipline *management*.We exploit a new and comprehensive dataset on the research output of academics in business economics who are employed at universities in Austria, Germany and (German-speaking) Switzerland. Research in economics and research in business economics are complementary. Lacking a business school tradition, business economics - with only few exceptions - constitutes an integral part of most economics faculties at Austrian, German and Swiss universities. This close relationship indicates that similar standards should be applied when evaluating research performance in economics and business economics.

In particular, it is evident that research success must be measured in terms of publica-

tions in journals that adhere to some minimum quality standard. For incentive-compatible performance measurement, it is then further necessary to account for quality differences between journals. By the same token, the evaluation strategy needs to be balanced across economics and business economics. Unfortunately, traditional ranking studies have often been tailored to meet the requirements of 'proper' economic research. Consequently, the publication data of business economists are underrepresented and the weighting schemes appear inappropriate. In contrast, our analysis reflects the publication habits in the field of business economics.

The paper is organized as follows. In the next section, we describe our dataset and our measures of research productivity. Instead of including a comprehensive literature survey, we discuss the relevant literature when we report our results in the following sections. In Section 3 we present our department rankings. In Section 4 we analyze institutional effects on research productivity and derive some conclusions concerning the training of junior scientists. The impacts of individual characteristics on research performance are analyzed in Section 5. The final section provides a brief outlook on important issues for future analysis.

2.2 Data and Methodology

We draw on a dataset collected under the auspices of the Committee for Research Monitoring of the German Economic Association (*Verein für Socialpolitik*). The dataset is housed by the Thurgau Institute of Economics and funded by the Association and the *Handelsblatt*, a leading German business newspaper. It comprises publication records and personal data of roughly 1,800 scientists in the field of Business Economics and Management who are employed by Austrian, German or (German-speaking) Swiss universities in spring 2008. Most of these researchers are employed by a full university.[1] However, we also include the academic staff of institutions that, by international standards, rather resemble business schools.[2] We focus on individuals who possess a doctor's degree and

[1] From the original list of university departments provided by the German Rectors' Conference (HRK), we exclude departments with less than four full professors in our sample - leaving out the International University Bruchsal, the Jacobs University Bremen, the Technical University Graz, the Universities of Erfurt, Hildesheim, Koblenz-Landau and Salzburg, the Kassel International Management School, the WHL Lahr and the International Graduate School (IHI) Zittau. We further omit the Dresden International University, the Steinbeis College Berlin and the Krems-Donau University because their staffs consist (almost exclusively) of academics from other universities on lecture contracts. Owing to its extreme specialization on health management, we also leave out the Medical University Hannover.

[2] The respective schools are legally entitled to award doctor's degrees.

whose principal occupation is academic research and teaching. Part-time lecturers with a primary nonuniversity employment are not included in the dataset.

Personal data and data on institutional characteristics of the departments are gleaned from the departments' homepages. The publications are collected from the EconLit and WISO databases. WISO indexes a large number of journals that publish articles in German. We account for differences in journal quality by using one of the journal meta-rankings proposed by Schulze, Warning, and Wiermann (2008). Meta-rankings are generated by imputing several journal weighting schemes that cover different but overlapping sets of journals. Specifically, we employ journal weights of the meta-ranking that uses Ritzberger's classification as the base scheme. Ritzberger calculates journal impact factors according to reciprocal citations for SSCI journals in the categories economics, business, finance, industrial relations, and labor, and for selected statistics journals.

Schulze, Warning, and Wiermann (2008) supplement this classification with additional journals that are not included in the SSCI but are ranked in questionnaire surveys conducted by Bräuninger and Haucap (2001), the German Academic Association for Business Research (VHB) and the Vienna University of Economics and Business Administration (WU Wien). While Bräuninger and Haucap's classification includes many economics journals that publish in German, the VHB and the WU Wien classifications introduce the business economics focus that we need for our analysis. The meta-ranking then classifies 2,825 journals (economics and business administration) by sorting them into six quality groups with group weights ranging from one to six.

Intuitively, it may appear more appropriate to use a meta-ranking that is based on the VHB or the WU Wien classification. However, such meta-rankings would virtually place all SSCI-listed journals into the top category. In contrast, using Ritzberger's list as the base scheme induces sufficient variation in the journal weights of the resulting meta-ranking. We admit that this procedure may induce a bias against management journals that have an interdisciplinary perspective. For our specific purpose, however, this feature is rather desirable because the results can be readily compared with the available rankings of economics departments. Such comparisons are interesting because pure business administration departments are the exception in Austria, Germany and Switzerland. The standard institutional set-up is rather a department of economic science that encompasses economics as well as business administration. Academics in business

administration are thus regularly subjected to research evaluations that fail to account for disciplinary differences.

To measure research performance, we assign a score pw/n to each publication in the sample where p denotes the number of pages, w is the journal weight and n the number of authors. A researcher's output is then defined as the sum of the scores of all articles written over his or her career. Individual research productivity is defined as output divided by career years. Because the weight of journals in the lowest quality category is one, the individual productivity measure can be interpreted as the average number of standardized pages in journals of the lowest quality category per career year.

We assume that the year in which a scientist is awarded the doctorate marks the beginning of his or her career. In cases where this information is missing, we use an estimate of the first career year: for all researchers whose first career year is known we compute the median time lag between the beginning of the career and the first publication. We then assume that this time lag should also apply to individuals for whom the information about the beginning of the career is missing. Department productivity is defined as the average of the productivities of its individual members. Thus, the department productivity measure can be interpreted as the average annual number of standardized pages in journals of the lowest quality category per department member.

Table 2.1 illustrates the distribution of the 2,825 journals and of the 20,879 articles in the dataset across the six quality categories. The distribution of the articles is bimodal. To test the hypothesis that this bimodality results from the interference of two distributions - one for top researchers and one for less prolific researchers - we compute the distribution of articles separately (1) for researchers who have achieved at least one publication in a top journal and (2) for researchers without a top publication. The last two columns of Table 2.1 reveal that individuals of both groups publish more articles in journals with a quality weight of four than in journals with quality weights of three and five. This observation does not support the above hypothesis. The observed bimodality is rather due to the way in which journals are assigned to quality categories. Journals in category four seem to be more popular research outlets for business economists in Austria, Germany and Switzerland.

Table 2.1 also provides information about the distribution of research output and the average number of authors per article across the six types of outlets. Comparing

Table 2.1: Distribution of journals, publications, scores, authors over journal classifications

Quality weight	% of journals	% of articles	% of output	Average authors per article	% of articles – at least one top publication	% of articles – without top publication
6	0.50	0.39	3.13	2.14	9.36	0.00
5	0.74	0.38	1.92	2.19	4.28	0.21
4	1.17	0.99	4.47	2.30	8.32	0.67
3	2.09	0.79	2.55	2.03	4.39	0.63
2	4.39	3.17	6.97	2.00	11.33	2.82
1	91.12	94.28	80.95	1.90	62.31	95.66
Number, average	2,825	20,879		1.91		

the distribution of the number of publications with the distribution of total output across quality categories illustrates the effect of the quality-weighting scheme. Most of the articles in our sample are either single (37%) or double authored (41%). The average number of authors appears to increase with journal quality.

One of our objectives is to investigate whether institutional and individual characteristics affect research productivity. Because almost 15% of the academics in our sample did not publish in our sample of journals, we then use Tobit regressions to identify the determinants of productivity. The descriptive statistics of the data used in our regression analyses of average department productivity (in Section 4) and of individual productivity (in Section 5) are detailed in Table 2.6.

2.3 Department Rankings

Table 2.7 reports department rankings according to research productivity. Table 2.7(a) includes only full professors and Table 2.7(b) includes full professors and junior staff. The leading department is at the University of Bonn. On average, full professors in Bonn publish the equivalent of almost 30 pages per career year (without co-authors) in journals of the lowest quality category. The departments at the universities of Mannheim and Vienna - respectively at the WHU Koblenz/Vallendar, when accounting for junior staff - are ranked second and third. Adopting a bird's-eye view, we cannot confirm a separation of research and teaching universities in Austria, Germany and Switzerland. This is in stark contrast to the situation in the United States. Research output is not concentrated on a select group of departments: the normalized Herfindahl index of 0.0088 (0.0086 for the ranking including junior staff) does not indicate a monopolization of the 'market for

Table 2.2: Rank correlations between productivity rankings using different journal weighting schemes (professors and junior staff)

	Whole sample	Quantile 1 (worst)	Quantile 2	Quantile 3	Quantile 4 (best)
VHB	0.8620	0.8079	0.3870	0.6364	0.4113
WU Wien	0.8012	0.7817	0.2043	0.5844	0.2641
Unweighted	0.8227	0.8827	0.4183	0.4632	0.4078
Combes/Linnemer	0.5549	0.3698	0.1609	0.4826	0.1957
Tinbergen	0.4084	0.2598	0.0960	0.1966	0.3101
No. of observations	89	23	22	22	22

publications'.

To judge the robustness of our results with respect to changes in the journal weighting scheme, Table 2.2 reports rank correlation coefficients between our ranking displayed in Table 2.7(b) and alternative rankings. Two of the alternative rankings are taken from Schulze, Warning, and Wiermann (2008) as well but use the VHB and the WU Wien classification as reference lists. We also compare our ranking with a ranking that uses no journal weights at all. For the whole sample the rank correlation between our preferred ranking and these three rankings is rather high. The rank correlations for the quantile 2-4 subsamples are, however, substantially lower, confirming that productivity differences between departments are relatively small.

There is much more disagreement in ranking departments that exhibit high productivity (quantile 4) than in ranking departments with less prolific members: the publication incidence in high-quality journals is actually only noticeable in good departments. Weightings induce shifts in rankings mainly at the top of the lists. This interpretation is confirmed by the rank correlation between our preferred ranking and the ranking computed with unitary quality weights. Again, the rank correlation is higher for low-productivity departments. Thus, high productivity and high quality are correlated.

Table 2.2 also displays rank-order correlations *vis-à-vis* productivity rankings based on the journal weighting schemes by Combes and Linnemer (2003) and the Tinbergen Research Institute at the Erasmus University, Rotterdam. Both classifications focus on journals in 'proper' economics (EconLit). Hence, they do not account for most business journals that we include in our ranking. The correlations between our preferred ranking and these two rankings are - not surprisingly - significantly lower than the correlations discussed above. This finding indicates that publications in WISO journals that are not listed in EconLit cannot be neglected in a well-balanced ranking for the business economics

profession. Although EconLit covers the most important and influential economics journals, business economists very often choose other publication outlets. Only 21% of the publications in our dataset are recorded by EconLit. Restricting the analysis to these journals would thus seriously distort the evaluation of research in business economics.

Rauber and Ursprung (2008a) propose to control for cohort effects if evaluating departments with different age structures. Following their method, we therefore define an individual's cohort by the group of peers who received their doctor's degree up to two years before or after the reference individual. We then order the peers in each cohort according to research productivity and assign the appropriate quantile to each individual. In a last step each department's score is calculated as the mean of the quantile values of its individual members.

Our cohort ranking based on the sample including junior staff is presented in Table 2.7(c). The leading department according to this ranking is at the University of Konstanz followed by the departments of the Technical University of Braunschweig and the Ludwig-Maximilians-University München. The rank correlation coefficient between the productivity and the cohort ranking is 0.7983. However, cohort rankings do not use information on the absolute differences of productivities within cohorts. Furthermore, not every additional publication increases the score. Thus, performance measurement using cohort rankings may provide somewhat weaker incentives to publish.

Using the departments' web pages, 1,490 individuals can be assigned to subdisciplines. In Table 2.3 we report top-five department lists for the subdisciplines "Financial Markets and Corporate Finance', 'Managerial Accounting', 'Marketing and Sales', 'Organization, Personnel and Strategy' and 'Financial Accounting, Auditing and Taxation'. Initially, we identified two more subdisciplines. Yet, we exclude the field 'Production, Cost Accounting and Industrial Management' because we are too often unable to differentiate this field from business information systems. We also exclude the subdiscipline 'Public Enterprise Management' due to an insufficient number of observations.

Only four departments, the departments of the universities of Jena, Mannheim, Köln and Paderborn, make it into the top-five lists in two subdisciplines. No department can claim more than two top rankings. This observation suggests that business economics research is rather specialized. Or phrased in terms of current German higher education politics, centers of excellence are not concentrated in a small number of locations.

Table 2.3: Top-five department by fields of research

Rank	Financial Markets and Corporate Finance	Managerial Accounting	Marketing and Sales	Organization, Personnel, and Strategy	Financial Accounting, Auditing, and Taxes
1	Mannheim University	Wien University	Darmstadt TU	Würzburg University	Saarbrücken University
2	Ulm University	Koblenz/ Vallendar WHU	Koblenz-Landau University	Bonn University	Paderborn University
3	Karlsruhe University	Ilmenau TU	Augsburg University	Paderborn University	Köln University
4	Dortmund University	Graz University	Jena University	Köln University	Hannover University
5	Jena University	Bremen	Mannheim University	Braunschweig TU	Trier University

Table 2.8 provides a ranking of departments such that research output is assigned to the individual's original training department - defined either as the department that granted the researcher's doctor's degree or *venia legendi* - instead of the department that the researcher is currently affiliated with. Unfortunately, we are unable to obtain information concerning the training department for all individuals in our sample. We only include departments in which at least four professors received their training. Professors who received their doctor's degree from the Humboldt University in Berlin, the University of Bonn and the University of Hagen are most productive (on average). The Technical University of Vienna, the University of Bonn and the University of Passau awarded the *venia legendi* to the most productive researchers in our sample.

The University of Bonn, which is the top university in terms of current department productivity, also belongs to the most successful training institutions. The other leading training departments do not stand out as high-productivity departments in Table 2.7(b). Generally, rank correlations between the rankings based on current affiliations and training institutions are moderate. The rank correlation between the productivity ranking reported in Table 2.7(b) and the productivity rankings in Table 2.8 is slightly higher when focusing on the doctor's degree 0.5234 than on the *venia legendi* 0.4799.

According to Davies, Kocher, and Sutter (2008) and Kocher and Sutter (2001), the concentration of research output across universities is higher if the research output is assigned to the department that granted the researcher's doctor's degrees than if it assigned to the researcher's current affiliation. The same holds true for our sample. However, the normalized Herfindahl index is still very low: the respective values are 0.0220 (doctorate)

and 0.1835 (*venia legendi*). Interpreting this information with due care suggests that the market for junior business economists is not very concentrated in the German-speaking area. We cannot single out a small group of departments that train the most productive individuals. Thus, it does not appear to be a promising strategy to concentrate recruiting on a few prestigious departments when hiring new faculty.

2.4 Institutional Effects

In this section we investigate whether institutional characteristics affect the research productivity of entire departments. Research productivity is measured as the average of the productivities of department members including junior staff. Table 2.4 reports the results of a Tobit regression analysis. We present results for two subsamples. Because the variable 'number of students' is not available for Austrian departments, only German and Swiss departments are considered in subsample 1, while subsample 2 also includes the Austrian departments.[3]

Table 2.4: Regression output of Tobit regressions for university sample (professors and junior staff)

	(1)		(2)	
	Without Austrian departments		All departments	
Dependent variable: department productivity	Coefficient	Standard error	Coefficient	Standard error
Size	0.1191	0.0549**	0.2095	0.0925**
Size squared			0.0019	0.0010*
No. of non-publishing professors	-1.6020	0.3993***	-1.5339	0.3743***
Dummy: economics	1.4983	0.9578	1.7729	0.8472**
No. of students per professor	-0.0009	0.0029		
Dummy: Switzerland	0.7088	1.8940	1.6192	1.8016
Dummy: Austria			-3.2536	1.7379*
Ratio Dr/Prof.	-0.8193	1.4770	-1.5035	1.3434
Constant	8.3686	1.0510***	7.4355	1.0942***
No. of observations	79		89	
Pseudo-R^2	0.0427		0.0515	

Notes: ***Significant at the 1% level, **Significant at the 5% level, *Significant at the 10% level.

We find that research productivity increases with department size as measured by the number of department members (see Table 2.4). Using subsample 2 that includes the Austrian departments (see Table 2.4, column 2), the effect of department size on

[3]For the same reason we must also exclude three German business schools (ESCP-EAP Berlin, Frankfurt School of Finance and Management, and Zeppelin University) from subsample 1.

productivity is actually positive but diminishing. Only when department size exceeds 55 persons, productivity begins to decline. There are only two departments with such a large faculty: the department of the WU Wien and the department of the University of St. Gallen. This finding is perfectly in line with Cainelli, de Felice, Lamonarca, and Zoboli (2006), who show that average research output of Italian economics departments is higher in larger departments. The positive correlation between productivity and size may reflect either increasing returns in research production (conceivably due to more peer pressure) or the selection of more successful individuals into larger and potentially more prestigious departments.

Cainelli, de Felice, Lamonarca, and Zoboli (2006) also report that research output is highly concentrated within Italian economics departments, a result that is confirmed by Australian evidence (see Neri and Rodgers, 2006). According to Cainelli, de Felice, Lamonarca, and Zoboli (2006)), this result reflects the division of labor that allows some individuals to specialize in research while others assume teaching and administrative duties. To investigate this issue, we use the Gini coefficient as a measure for the concentration of research output within departments. The average of the Gini coefficients over all departments is 0.22, indicating that concentration of research within departments is moderate. Specifically, the Gini coefficients in our sample are much lower than the Gini coefficients reported by Neri and Rodgers (2006) for Australian economics departments. Furthermore, we find virtually no correlation between concentration of research output and productivity. Division of labor thus does not necessarily induce better research performance.

Our next estimate shows that productivity is lower in departments with a higher number of non-publishing professors. Whether this confirms the finding of Taylor, Fender, and Burke (2006), who claim that researchers with productive peers are more productive themselves, remains questionable: in our computations department productivity is defined as the average over all individual productivities. Thus, this average also includes the unproductive members. We return to this issue in the next section where we analyze the determinants of individual research productivities.

Most programs in business economics and management in Austria, Germany and Switzerland are associated with economics departments. Interdisciplinary collaboration and interdisciplinary competition are likely to have an impact on productivity of business economists. In fact, our estimates show that productivity is higher in departments that

also run an economics study program (see Table 2.4, column 2).

According to Maske, Durden, and Gaynor (2003) and Taylor, Fender, and Burke (2006), higher teaching loads and/or more administrative duties reduce research productivity. We attempt to proxy the teaching load by the total number of students who major in business economics and management, economics or a related discipline and divide this number by the number of faculty members. Unfortunately, we were not able to uncover federal statistics on student numbers in Austria. The estimate for the subsample that includes only German and Swiss departments suggests, however, that higher teaching loads in terms of class sizes do not deter research productivity (see Table 2.4, column 1).

Research grants are provided with the intention to enhance research productivity. Often, past research performance is appreciated and used as a predictor for future research performance. We therefore expect a positive correlation between research grants per capita and department productivity. In 2005 the German CHE Consult (an organization that is specialized on advising institutes of higher education) collected data on research grants per researcher for a large number of German universities (see Berghoff, Federkeil, Giebisch, Hachmeister, Hennings, and Müller-Böling, 2006). The respective figures for Austria and Switzerland were released by the Austrian Agency for Quality Assurance and the swissUp project in Basel.[4] Owing to missing observations for some universities in our sample, we do not use this information in our regression analysis. Instead, we only compute the correlation coefficient. The coefficient value is 0.0931, indicating only a weak impact of research grants per capita on department research performance. This observation is in line with results of Arora, David, and Gambardella (1998) and Jacob and Lefgren (2007). Their explanation emphasizes that research grants only displace other sources of funding without actually improving total research funding.

According to Combes and Linnemer (2003), total publication output and publication output per capita are higher for German departments than for Swiss departments. The respective figures for Austrian departments are even lower. In contrast, Eichenberger, Meier, and Arpagaus (2000) find that, upon controlling for differences in population size, Austrian and Swiss departments exhibit higher research productivities than German departments. Both of these country comparisons consider only articles published within a rather restricted period of time. Eichenberger, Meier, and Arpagaus (2000) further focus

[4] See http://www.hochschulranking.ac.at and http://www.rankingswissup.ch, respectively.

their analysis on a small subset of journals. We find no significant differences in productivity between German and Swiss departments and significantly lower productivities for Austrian departments (see Table 2.4).

Finally, our Tobit regression reveals that the share of post-docs in a department does not significantly affect the average department productivity. Mentoring of post-docs does not seem to conflict with the research performance of professors.

2.5 Determinants of Individual Research Productivity

In this section we investigate the effects of institutional determinants and personal characteristics on individual research productivity. The results of a Tobit regression analysis for two different subsamples consisting of all faculty members (column 1) and of full professors only (column 2) are reported in Table 2.5. Additionally, we use a Hurdle model to analyze the propensity to publish and the productivity given publication incidence separately. We specify the initial binary choice in the first tier of the Hurdle model by a Probit model. For the second tier, rather low productivities of many researchers in our sample suggest the log-transformation of the productivity index. Following Wooldridge (2002, pp. 536-538), we therefore assume a log normal distribution of individual productivities of active researchers and use the OLS estimator for the second tier of the Hurdle model. The results of the Hurdle model are presented in Table 2.9. Again, we distinguish two subsamples: the subsample of all faculty members [Table 2.9(a)] and the subsample of full professors [Table 2.9(b)].

Individual productivity is affected by institutional determinants. Researchers in larger departments are more productive. However, the size effect on individual productivity is non-linear: the coefficient associated with the square of size is significantly negative. The effect reaches its maximum for researchers in departments with about 61 persons. However, we are reluctant to interpret this number as an optimal department size because all departments in our sample except for the departments of the University of St. Gallen and of the WU Wien are smaller - and both, by international standards, resemble business schools. The size effect rather indicates that potential returns to scale in research production are positive but diminishing.

Researchers from departments with a larger share of junior scientists exhibit lower productivity on average. The Hurdle model reveals that this effect is not due to significant

Table 2.5: Tobit regressions for individual sample

	(1)		(2)	
	All researchers		Only full professors	
Dependent variable: individual productivity	Coefficient	Standard error	Coefficient	Standard error
Size	0.2748	0.0546**	0.2829	0.0698**
Size squared	-0.0025	0.0005**	-0.0027	0.0007**
No. of non-publishing professors	-1.3432	0.2082**	-1.2921	0.2441**
Dummy: economics	2.4063	0.6201**	2.8688	0.7791**
Dummy: Switzerland	0.5649	1.0334	1.5135	1.3429
Dummy: Austria	-1.1933	1.0625	0.1279	1.5112
Ratio Dr/Prof.	-1.9402	0.6945**	-1.7928	0.9664*
Career age	-0.3304	0.0358**	-0.3126	0.0391**
Dummy: Prof. PhD	4.5697	2.4948*	4.5878	2.4531*
Dummy: Juniorprofessor	-2.4386	1.6604		
Dummy: Privatdozent	-3.6370	1.2039**		
Dummy: Dr	-7.8833	0.7739**		
Dummy: PhD	-15.7329	5.5552**		
Dummy: ao. Prof.	-3.9336	2.1236*		
Dummy: gender (female = 1)	-4.5103	0.7230**	-3.5733	1.0868**
Constant	13.1355	1.0461**	12.0470	1.2232**
No. of observations	1,482		870	
Pseudo R^2	0.0236		0.0194	

Notes: **Significant at the 1% level, *Significant at the 10% level.

differences in the propensity to publish but to a lower productivity of researchers who are publishing. Active post-docs in particular seem to profit from mentoring or from exchange with experienced colleagues. Informal collaboration between professors and post-docs within the same department is likely to be more developed in departments in which the share of post-docs is smaller. In any event, it does not seem to be the case that the research productivity of the senior faculty suffers when the junior faculty is sizable.

Recall from the previous section that productivity is lower in departments with a high number of non-publishing professors. We can now confirm that active researchers with less productive peers are less productive themselves. Taylor, Fender, and Burke (2006) suggest that research is valued more strongly, more resources are devoted to research, and opportunities for formal or informal collaboration are better in departments with a larger share of publishing academics. Also, this finding may reflect peer effects. In particular, when recruiting new faculty, superior research productivity may be of minor value or even an impediment if incumbent professors want to control internal research competition. Alternatively, however, the effect may be attributed to a selection bias: highly productive researchers may avoid becoming affiliated with departments with a large share of inactive colleagues.

Members of departments that also run economics study programs are more productive. The Hurdle model reveals that this finding can be attributed mainly to higher productivity of active scientists. Thus, professional exchange and competition with economists are particularly conducive to the productivity of researchers who already have some publication experience.

To account for life cycle effects, we define 'career age' as the number of years since obtaining the doctor's degree. Individual productivity then decreases with career age.[5] Remarkably, we find a negative effect of career age on the propensity to publish for the subsample of full professors [see Table 2.9(b)]. Because our estimates are based on aggregated data, professors of a higher career age who had more opportunities to publish than peers with shorter careers are actually less likely to have at least one journal publication during their whole career. Possibly, this finding is due to a change in publication behavior from books and collective volume articles to journal articles that has taken place in more recent times. For the subsample that also includes junior scientists we identify a positive effect of career age on the propensity to publish. The non-linearity of the effect indicates that it is harder for older scientists to publish their first journal article.

The decrease in the productivity of active researchers is in line with the life cycle hypothesis. For (younger) economists who are employed at German universities, Rauber and Ursprung (2008b) report that publication behavior follows a characteristic life cycle: productivity increases in the first years of an academic career, reaches a peak six to eight years after the onset of the academic career and begins to decline afterwards. Explanations of the decline in productivity of tenured professors include the lack of career incentives, the increased obsolescence of knowledge and an increased preference for non-research activities.

We also include dummies for an individual's highest academic degree in our regressions. The negative dummy coefficients for young researchers (see Table 2.5) indicate a lower productivity compared with full professors. With the exception of so-called 'Juniorprofessoren' and 'Privatdozenten' (staff without and with *venia legendi*, both non-tenured), lower productivity is at least partly due to a smaller propensity to publish [see Table 2.9(a)]. Because careers of younger scientists are shorter and many journals exhibit considerable publication lags they simply have had fewer opportunities to publish than professors. 'Juniorprofessoren' and 'Privatdozenten' still need to pass a rigorous competitive assessment

[5] We tested whether the age effect is non-linear but the coefficients of higher-order polynomials of the variable 'career age' turned out to be insignificant.

on the basis of their publication record when applying for a full professorship. Their propensity to publish does not significantly differ from full professors.

Within the group of active researchers, non-professors are ceteris paribus less productive [see Table 2.9(a), column 2]. However, finding lower productivities for non-professors who are of the same career age as full professors is not surprising. It only shows that promotions are actually at least partly granted on the basis of an assessment of past research success. To compare the productivities of active young researchers and full professors, we have to account for the fact that the careers of junior researchers are shorter than the careers of full professors. Comparing productivities of median aged junior researchers and median aged full professors, the junior scientists exhibit a higher productivity.

Although we do not know the country in which the academic training took place, we attempt to address the effect of having obtained academic training outside of the German-speaking region. Until very recently the short form for the doctor's degree awarded by Austrian, German and Swiss universities was 'Dr'. Thus, it is likely that individuals whose homepages report a 'PhD' degree have received their academic training abroad. Comparing full professors only (see Table 2.5, column 2), those who obtained a 'PhD' degree are more productive than researchers holding a 'Dr' degree. In his study on Portuguese economists, Guimarães (2002) reports that there are no significant differences in the propensity to publish in international journals between scientists who obtained their doctorate in Portugal and scientists who received their academic training abroad. There is one notable exception: academics who obtained their PhD in the United States are more likely to publish in international journals than their peers. This finding may either reflect better training, an advantage of US-based departments in the competition for top junior researchers, or the cultivation of a home bias of US-based top journals (see e.g. Hodgson and Rothman, 1999; Kocher and Sutter, 2001). Finally, we find evidence for gender differences in the publication behavior. Female business economists appear less productive than their male peers. Such differences have also been reported for 'proper' economics research (see e.g. Maske, Durden, and Gaynor, 2003; Rauber and Ursprung, 2008b; Taylor, Fender, and Burke, 2006). Rauber and Ursprung (2008b) show that female researchers are less likely to publish but that women who publish are just as productive as their male peers. In contrast, our Hurdle model reveals that active women exhibit a lower productivity than men. Moreover, we actually find no significant differences in the

propensity to publish between male and female professors. When using cross-sectional data, lower research output during career interruptions (e.g. during maternity leaves) implies lower overall productivity. In contrast, such events are likely to affect only the publication propensity in the years on leave when using panel data. Hence, there may be a rather simple explanation for the difference between our result and Rauber and Ursprung (2008b).

2.6 Outlook

Drawing on a new comprehensive dataset that collects the research output of roughly 1,800 business economists working at Austrian, German and Swiss universities, we provide research rankings of university departments and analyze the determinants of research performance. We find that individual research productivity - and consequently departmental research productivity - is affected by institutional and personal characteristics. Most of our findings appear to be in line with previous findings from studies on 'proper' economics that exist for various countries. A direct comparison of research performance between the disciplines economics and business economics would certainly be promising - and possible, given the new data.

Another issue that may be addressed in the future is the problem of adequately accounting for interdisciplinary research. It remains to be tested, for instance, whether the gender differences with regard to publication performance are due to restrictions imposed by the publication data. Women's choices of study programs are known to be biased toward the arts and cultural studies (see BMFSFJ 2005). Consequently, female academics in business economics and management may tend to specialize on interdisciplinary research that is certainly underrepresented in our publication data. Also, business school-type universities may be underrated in our ranking because both teaching and research may have a more interdisciplinary orientation than research undertaken at full universities in which economic science departments offer joint study programs in economics and business administration. Further, because business school-type institutions specialize in supplying a broad and basic business education, teaching possibly obtains greater relative importance and staff may be more specialized on this task than in 'full' universities.

These open issues are clearly just as important for evaluations of research in 'proper' economics and in (business) economic disciplines that engage in developing quantitative

research methods. Interdisciplinary research in these fields may be published in science journals that are not included in either EconLit or WISO. In any event, measuring research performance in areas that are inherently interdisciplinary requires the collection of even more comprehensive data and more elaborate evaluation methods. We hope that the German Economic Association's research monitoring group will be able to tackle these issues in the near future.

2.7 Appendix

Table 2.6: Descriptive statistics

	Mean	Standard deviation	Mean	Standard deviation
Sample: university data	Without Austrian universities		All universities	
No. of observations	79		89	
Productivity	9.9065	4.3238	9.5493	4.3185
Dummy: economics	0.6456	0.4814	0.6180	0.4886
Size (no. of faculty members)	16.6582	10.3166	18.2472	13.6242
No. of students per professor	172.8767	166.0015		
No. of non-publishing professors	0.5823	1.2771	0.6180	1.2294
Ratio Dr/Prof.	0.4777	0.3524	0.4840	0.3695
Dummy: Switzerland	0.0633	0.2450	0.0562	0.2316
Dummy: Austria			0.0787	0.2707
Sample: individual data	All researchers		Only full professors	
No. of observations	1,482		870	
Productivity	8.8581	10.4480	10.1363	10.4765
Dummy: economics	0.6815	0.4660	0.6632	0.4729
Size (no. of faculty members)	28.0331	21.6013	24.6035	18.6563
No. of non-publishing professors	0.8516	1.5190	0.8989	1.6883
Ratio Dr/Prof.	0.7318	0.5204	0.6050	0.4624
Dummy: Switzerland	0.0877	0.2830	0.0805	0.2722
Dummy: Austria	0.1619	0.3685	0.1011	0.3017
Career age	14.1754	9.5608	18.8575	9.0384
Dummy: Prof. PhD	0.0115	0.1065	0.0195	0.1385
Dummy: Juniorprofessor	0.0290	0.1679		
Dummy: Privatdozent	0.0587	0.2352		
Dummy: Dr	0.3023	0.4594		
Dummy: PhD	0.0034	0.0580		
Dummy: a.o. Prof.[a]	0.0196	0.1386		
Dummy: gender (female = 1)	0.1808	0.3850	0.1138	0.3177
Dummy: publication	0.9325	0.2509	0.9805	0.1385

'a.o.' indicates 'extraordinary professorship', i.e. tenured or non-tenured professorship achieved without undergoing formal application procedures.

Table 2.7: (a) Productivity ranking of departments (full professors only)

Rank	University	Productivity	Rank	University	Productivity
1	Bonn University	29.70	46	Berlin FU	8.92
2	Mannheim University	19.85	47	Gießen University	8.92
3	Wien University	19.21	48	Wuppertal University	8.90
4	Saarbrücken University	17.51	49	Dresden TU	8.71
5	Koblenz/Vallendar WHU	17.48	50	Hamburg University	8.47
6	Augsburg University	16.49	51	Magdeburg University	8.41
7	Frankfurt/Main University	16.21	52	Berlin TU	8.25
8	Konstanz University	16.20	53	Zürich ETH	8.02
9	Köln University	16.12	54	Oestrich-Winkel EBS	7.79
10	München TU	15.87	55	Mainz University	7.69
11	Braunschweig TU	15.75	56	Oldenburg University	7.67
12	München LMU	15.60	57	Bremen University	7.50
13	Ulm University	15.43	58	Marburg University	7.41
14	Dortmund University	15.30	59	Wien WU	7.38
15	Basel University	14.76	60	Eichstätt KU	7.26
16	Jena University	14.66	61	Clausthal TU	7.23
17	Aachen RWTH	14.08	62	Siegen University	7.03
18	Würzburg University	13.97	63	München UniBW	6.82
19	Bern University	13.86	64	Hohenheim University	6.81
20	Kiel University	13.72	65	Zeppelin University	6.37
21	Darmstadt TU	13.69	66	Düsseldorf University	6.27
22	Zürich University	13.06	67	Innsbruck University	6.26
23	Regensburg University	12.80	68	Frankfurt School of F&M	6.20
24	Paderborn University	12.76	69	Witten/Herdecke University	6.00
25	Hannover University	12.04	70	Leipzig University	5.68
26	Karlsruhe University	11.96	71	Frankfurt/Oder University	5.59
27	Bamberg University	11.81	72	Bielefeld University	5.28
28	Bochum University	11.71	73	Potsdam University	5.05
29	Kaiserslautern TU	10.99	74	Chemnitz TU	4.93
30	Passau University	10.99	75	Ilmenau TU	4.91
31	Stuttgart University	10.60	76	Cottbus BTU	4.71
32	Münster University	10.30	77	Osnabrück University	4.44
33	Graz University	10.18	78	Rostock University	4.26
34	Erlangen-Nürnberg University	10.05	79	Kassel University	4.15
35	Duisburg-Essen University	9.73	80	Berlin ESCP-EAP	4.03
36	Greifswald University	9.67	81	Hamburg TU	3.93
37	Tübingen University	9.64	82	Bayreuth University	3.44
38	Göttingen University	9.37	83	Linz University	3.38
39	St.Gallen University	9.34	84	Hamburg UniBW	3.29
40	Freiburg University	9.26	85	Halle-Wittenberg University	3.22
41	Hagen FernUni	9.21	86	Flensburg University	2.54
42	Trier University	9.16	87	Freiberg TU	2.41
43	Wien TU	9.01	88	Lüneburg Leuphana University	2.18
44	Berlin HU	8.96	89	Klagenfurt University	2.06
45	Leipzig HHL	8.95			

Table 2.7: (b) Productivity ranking of departments (professors and junior staff)

Rank	University	Productivity	Rank	University	Productivity
1	Bonn University	24.01	46	Wien TU	8.92
2	Mannheim University	18.86	47	Wuppertal University	8.90
3	Koblenz/Vallendar WHU	17.81	48	St.Gallen University	8.86
4	Köln University	16.64	49	Dresden TU	8.70
5	Saarbrücken University	16.37	50	Berlin TU	8.33
6	Konstanz University	16.03	51	Berlin HU	8.30
7	München TU	15.87	52	Gießen University	8.21
8	Ilmenau TU	15.76	53	Hamburg University	8.17
9	Braunschweig TU	15.75	54	Magdeburg University	7.70
10	Frankfurt/Main University	15.72	55	Mainz University	7.56
11	Ulm University	15.56	56	Bremen University	7.44
12	München LMU	15.33	57	Marburg University	7.41
13	Basel University	14.76	58	Zürich ETH	7.16
14	Jena University	14.66	59	Siegen University	7.16
15	Wien University	14.21	60	Eichstätt KU	7.14
16	Würzburg University	13.97	61	Innsbruck University	7.11
17	Kiel University	13.72	62	München UniBW	7.11
18	Augsburg University	13.16	63	Graz University	6.87
19	Zürich University	13.06	64	Clausthal TU	6.83
20	Aachen RWTH	12.96	65	Hohenheim University	6.81
21	Chemnitz TU	12.69	66	Frankfurt/Oder University	6.65
22	Darmstadt TU	12.53	67	Oestrich-Winkel EBS	6.62
23	Regensburg University	12.50	68	Osnabrück University	6.37
24	Bern University	12.49	69	Zeppelin University	6.37
25	Dortmund University	12.16	70	Witten/Herdecke University	6.30
26	Karlsruhe University	11.96	71	Düsseldorf University	6.27
27	Bamberg University	11.81	72	Frankfurt School of F&M	6.20
28	Hannover University	11.76	73	Leipzig University	6.14
29	Paderborn University	11.65	74	Berlin ESCP-EAP	5.99
30	Greifswald University	11.59	75	Wien WU	5.91
31	Passau University	10.99	76	Bielefeld University	5.28
32	Tübingen University	10.95	77	Cottbus BTU	4.71
33	Stuttgart University	10.60	78	Potsdam University	4.69
34	Münster University	10.25	79	Kassel University	4.15
35	Berlin FU	10.13	80	Rostock University	3.95
36	Kaiserslautern TU	10.12	81	Hamburg TU	3.93
37	Duisburg-Essen University	9.71	82	Hamburg UniBW	3.72
38	Oldenburg University	9.56	83	Bayreuth University	3.44
39	Erlangen-Nürnberg University	9.32	84	Linz University	3.41
40	Bochum University	9.30	85	Halle-Wittenberg University	3.22
41	Freiburg University	9.26	86	Lüneburg Leuphana University	2.67
42	Hagen FernUni	9.21	87	Flensburg University	2.54
43	Trier University	9.16	88	Freiberg TU	2.41
44	Göttingen University	9.11	89	Klagenfurt University	2.29
45	Leipzig HHL	8.95			

Table 2.7: (c) Cohort rankings (professors and junior staff)

Rank	University	Average cohort-quantile	Rank	University	Average cohort-quantile
1	Konstanz University	0.84	46	Bochum University	0.60
2	Braunschweig TU	0.83	47	Leipzig HHL	0.59
3	München LMU	0.81	48	Düsseldorf University	0.59
4	München TU	0.80	49	Bremen University	0.59
5	Koblenz/Vallendar WHU	0.79	50	Wuppertal University	0.58
6	Kiel University	0.79	51	Dresden TU	0.57
7	Bonn University	0.78	52	Eichstätt KU	0.57
8	Frankfurt/Main University	0.78	53	Magdeburg University	0.56
9	Basel University	0.77	54	Frankfurt/Oder University	0.56
10	Mannheim University	0.76	55	Karlsruhe University	0.56
11	Regensburg University	0.75	56	St.Gallen University	0.56
12	Freiburg University	0.75	57	Berlin TU	0.55
13	Würzburg University	0.74	58	Chemnitz TU	0.55
14	Köln University	0.74	59	Siegen University	0.53
15	Passau University	0.73	60	Witten/Herdecke University	0.52
16	Ulm University	0.73	61	Zeppelin Universityversity	0.52
17	Stuttgart University	0.73	62	Clausthal TU	0.51
18	Dortmund University	0.72	63	Osnabrück University	0.50
19	Berlin FU	0.71	64	Hamburg University	0.50
20	Greifswald University	0.71	65	Leipzig University	0.49
21	Tübingen University	0.70	66	Zürich ETH	0.49
22	Bamberg University	0.70	67	Bielefeld University	0.49
23	Hannover University	0.70	68	Mainz University	0.49
24	Zürich University	0.69	69	Hohenheim University	0.48
25	Kaiserslautern TU	0.69	70	Innsbruck University	0.48
26	Aachen RWTH	0.69	71	Wien TU	0.47
27	Saarbrücken University	0.68	72	Wien WU	0.46
28	Wien University	0.68	73	Berlin ESCP-EAP	0.45
29	Hagen FernUni	0.66	74	Kassel University	0.45
30	Münster University	0.65	75	Potsdam University	0.43
31	Erlangen-Nürnberg University	0.65	76	Hamburg TU	0.43
32	Augsburg University	0.65	77	Oestrich-Winkel EBS	0.43
33	München UniversityBW	0.64	78	Graz University	0.42
34	Jena University	0.64	79	Bayreuth University	0.40
35	Trier University	0.63	80	Cottbus BTU	0.40
36	Paderborn University	0.63	81	Hamburg UniversityBW	0.39
37	Oldenburg University	0.63	82	Halle-Wittenberg University	0.38
38	Marburg University	0.62	83	Linz University	0.37
39	Darmstadt TU	0.62	84	Rostock University	0.37
40	Bern University	0.61	85	Frankfurt School of F&M	0.32
41	Duisburg-Essen University	0.61	86	Freiberg TU	0.30
42	Göttingen University	0.61	87	Lüneburg Leuphana University	0.28
43	Ilmenau TU	0.61	88	Klagenfurt University	0.28
44	Berlin HU	0.61	89	Flensburg University	0.27
45	Gießen University	0.61			

Table 2.8: (a) Productivity rankings according to training location (professors and junior staff) – doctor's degree

Rank	University granting doctor's degree	Productivity	Rank	University granting doctor's degree	Productivity
1	Berlin HU	21.10	30	Aachen RWTH	8.04
2	Bonn University	17.37	31	München LMU	7.98
3	Hagen FernUni	16.51	32	Karlsruhe University	7.97
4	Passau University	15.95	33	Magdeburg University	7.42
5	Mannheim University	15.36	34	Innsbruck University	7.41
6	Braunschweig TU	15.16	35	Wien WU	7.35
7	Kiel University	14.03	36	Berlin FU	7.24
8	Koblenz/Vallendar WHU	13.63	37	Duisburg-Essen University	7.14
9	Kaiserslautern TU	12.45	38	Hohenheim University	7.04
10	Saarbrücken University	12.38	39	Zürich University	6.87
11	Dortmund University	12.06	40	Paderborn University	6.86
12	Frankfurt/Main University	11.86	41	Basel University	6.75
13	Oldenburg University	11.74	42	Münster University	6.56
14	Würzburg University	11.48	43	Bochum University	6.41
15	Augsburg University	10.99	44	Erlangen-Nürnberg University	6.29
16	Trier University	10.84	45	Gießen University	6.23
17	St.Gallen University	10.79	46	Graz University	6.18
18	Hannover University	10.76	47	Berlin TU	5.92
19	Bielefeld University	10.70	48	Rostock University	5.55
20	Hamburg University	10.68	49	Bayreuth University	4.78
21	Köln University	10.67	50	Stuttgart University	4.61
22	Wien TU	10.66	51	Linz University	3.63
23	Regensburg University	10.44	52	Bremen University	3.59
24	Marburg University	10.37	53	Bamberg University	3.38
25	Tübingen University	10.14	54	München TU	2.72
26	Freiburg University	9.92	55	Zürich ETH	2.62
27	Wien University	9.73	56	Klagenfurt University	2.19
28	Freiberg TU	8.79	57	Oestrich-Winkel EBS	1.82
29	Göttingen University	8.64			

Table 2.8: (b) Productivity rankings according to training location (professors and junior staff) – *venia legendi*

Rank	University granting *venia legendi*	Productivity	Rank	University granting *venia legendi*	Productivity
1	Wien TU	25,55	28	Darmstadt TU	9,22
2	Bonn University	22,87	29	Dortmund University	9,07
3	Passau University	17,07	30	Bochum University	8,71
4	Hamburg UniBW	17,05	31	Erlangen-Nürnberg University	8,51
5	Basel University	16,96	32	München LMU	8,37
6	Bielefeld University	16,83	33	Innsbruck University	8,15
7	Koblenz/Vallendar WHU	16,32	34	Wien University	7,93
8	Lüneburg Leuphana University	15,91	35	München TU	7,64
9	Kiel University	15,69	36	Karlsruhe University	7,41
10	Würzburg University	14,98	37	Aachen RWTH	7,40
11	Hamburg University	14,71	38	Stuttgart University	7,23
12	Kaiserslautern TU	14,13	39	Berlin TU	7,18
13	Mannheim University	13,52	40	Wien WU	6,90
14	Saarbrücken University	13,51	41	Münster University	6,69
15	Köln University	13,46	42	Graz University	6,66
16	Berlin HU	12,96	43	Gießen University	6,50
17	Zürich University	12,93	44	Bayreuth University	6,23
18	Frankfurt/Main University	12,17	45	Paderborn University	5,92
19	Regensburg University	11,85	46	Eichstätt KU	5,80
20	Augsburg University	11,42	47	Hannover University	5,53
21	Trier University	10,92	48	Berlin FU	5,23
22	Hohenheim University	10,92	49	Göttingen University	4,43
23	Oldenburg University	10,54	50	Bremen University	3,50
24	Duisburg-Essen University	10,42	51	Oestrich-Winkel EBS	3,21
25	St.Gallen University	10,26	52	Klagenfurt University	3,17
26	Tübingen University	9,30	53	Linz University	2,64
27	Freiburg University	9,27			

Table 2.9: Hurdle model: (a) whole sample and (b) only full professors

	(1) 1. stage: Probit		(2) 2. Stage: OLS	
Dependent variable	Dummy: publication		Log productivity	
	Coefficient	Standard error	Coefficient	Standard error
(a) whole sample[a]				
Size	0.0075	0.0041*	0.0235	0.0057***
Size squared	-	-	-0.0002	0.0001***
No. of nonpublishing professors	-0.1490	0.0358***	-0.1709	0.0242***
Dummy: economics	-0.1703	0.1417	0.3194	0.0642***
Dummy: Switzerland	-0.1262	0.2126	0.1545	0.1128
Dummy: Austria	-0.3255	0.2024	-0.2461	0.1174**
Ratio Dr/Prof.	-0.0298	0.1368	-0.2298	0.0751***
Career age	0.3809	0.0863***	-0.1643	0.0329***
Career age2	-0.0313	0.0095***	0.0064	0.0018***
Career age3	0.0009	0.0004**	-0.0001	0.0000***
Career age4	0.0000	0.0000**	-	-
Dummy: Prof. PhD			0.1395	0.3355
Dummy: Junior professor	-0.3113	0.3895	-0.6439	0.1790***
Dummy: Privatdozent	0.0868	0.4176	-0.4273	0.1189***
Dummy: Dr	-1.1625	0.2126***	-1.1225	0.0994***
Dummy: PhD	-2.5850	0.6380***	-0.8595	0.1606***
Dummy: a.o. Prof.[b]	-0.4328	0.4871	-0.4404	0.2350*
Dummy: gender (female = 1)	-0.4613	0.1308***	-0.4659	0.0725***
Constant	1.3155	0.3201***	3.0683	0.1999***
No. of observations	1,482		1,382	
Pseudo-R^2	0.2378		0.2264	
(b) Only full professors[c]				
Size	0.0648	0.0253***	0.0206	0.0073***
Size squared	-0.0005	0.0002**	-0.0002	0.0001**
No. of non-publishing professors	-0.2554	0.0645***	-0.1512	0.0260***
Dummy: economics	-0.0917	0.2777	0.3326	0.0780***
Dummy: Switzerland			0.1412	0.1430
Dummy: Austria	-0.2374	0.3932	-0.2882	0.1722*
Ratio Dr/Prof.	-0.1561	0.3485	-0.1241	0.0996
Career age	-0.0268	0.0122**	-0.0408	0.0044***
Dummy: Prof. PhD			0.1889	0.3177
Dummy: gender (female = 1)	-0.1439	0.3228	-0.4274	0.1028***
Constant	2.1855	0.3291***	2.3325	0.1272***
No. of observations	870		853	
Pseudo-R^2	0.1528		0.1817	

Notes: ***Significant at the 1% level, **Significant at the 5% level, *Significant at the 10% level.
[a] All persons with the title Professor PhD published at least one article during their career.
[b] 'a.o.' indicates 'extraordinary professorship', i.e. tenured or non-tenured professorship achieved without undergoing formal application procedures.
[c] All professors from Switzerland and all persons with the title Professor PhD published at least one article during their career.

Bibliography

Arora, A., P. David, and A. Gambardella (1998): "Reputation and Competence in Publicly Funded Science: Estimating the Effects on Research Group Productivity," *Annales d'Economie et de Statistique*, 49-50, 163–198.

Berghoff, S., G. Federkeil, P. Giebisch, C.-D. Hachmeister, M. Hennings, and D. Müller-Böling (2006): "Das CHE ForschungsRanking deutscher Universitäten 2006," Working Papers 79, Centrum für Hochschulentwicklung, Gütersloh.

BMFSFJ (2005): *Gender Datenreport 2005*. Bundesministerium für Familie, Senioren, Frauen und Jugend, section 1.7.

Bräuninger, M., and J. Haucap (2001): "Was Ökonomen lesen und schätzen: Ergebnisse einer Umfrage," *Perspektiven der Wirtschaftspolitik*, 2(2), 185–210.

Cainelli, G., A. de Felice, M. Lamonarca, and R. Zoboli (2006): "The Publications of Italian Economists in ECONLIT. Quantitative Assessment and Implications for Research Evaluation," *Economia Politica*, 0(3), 385–423.

Clemenz, G., and K. Neusser (1991): "International impact of Austrian economists," *Empirica*, 18, 201–219, 10.1007/BF00924912.

Combes, P.-P., and L. Linnemer (2001): "La publication d'articles de recherche en économie en France," *Annales d'Economie et de Statistique*, 62, 5–47.

——— (2003): "Where are the Economists Who Publish? Publication Concentration and Rankings in Europe Based on Cumulative Publications," *Journal of the European Economic Association*, 1(6), 1250–1308.

Davies, J. B., M. G. Kocher, and M. Sutter (2008): "Economics research in Canada:

a long-run assessment of journal publications," *Canadian Journal of Economics*, 41(1), 22–45.

Dolado, J. J., A. García-Romero, and G. Zamarro (2003): "Publishing performance in economics: Spanish rankings (1990-1999)," *Spanish Economic Review*, 5, 85–100.

Eichenberger, R., U. Meier, and R. Arpagaus (2000): "Ökonomen, Publikationen und Zitationen: Ein europäischer Vergleich," *Perspektiven der Wirtschaftspolitik*, 1(2), 143–160.

Fabel, O., and F. Hesse (1999): "Befragungsstudie vs. Publikationsanalyse: Zur Interpretation von Ranglisten der Forschungsaktivitäten deutscher betriebswirtschaftlicher Fachbereiche," *Die Betriebswirtschaft*, 59, 196–204.

Guimarães, P. (2002): "The State of Portuguese Research in Economics: An Analysis based on Publications in International Journals," *Portuguese Economic Journal*, 1, 3–25.

Hein, M. (2006): "How Did Economic Research in Switzerland Develop since the Beginning 90s?," *Swiss Journal of Economics and Statistics (SJES)*, 142(II), 285–306.

Hodgson, G. M., and H. Rothman (1999): "The Editors and Authors of Economics Journals: A Case of Institutional Oligopoly?," *Economic Journal*, 109(453), F165–86.

Jacob, B., and L. Lefgren (2007): "The Impact of Research Grant Funding on Scientific Productivity," NBER Working Papers 13519, National Bureau of Economic Research, Inc.

Kocher, M. G., and M. Sutter (2001): "The Institutional Concentration of Authors in Top Journals of Economics during the Last Two Decades," *Economic Journal*, 111(472), F405–21.

Maske, K. L., G. C. Durden, and P. E. Gaynor (2003): "Determinants of Scholarly Productivity among Male and Female Economists," *Economic Inquiry*, 41(4), 555–564.

Neary, J. P., J. A. Mirrlees, and J. Tirole (2003): "Evaluating Economics Research in Europe: An Introduction," *Journal of the European Economic Association*, 1(6), 1239–1249.

Neri, F., and J. R. Rodgers (2006): "Ranking Australian Economics Departments by Research Productivity," *The Economic Record*, 82(s1), S74–S84.

Rauber, M., and H. W. Ursprung (2008a): "Evaluation of Researchers: A Life Cycle Analysis of German Academic Economists," in *Conferences on New Political Economy*, ed. by M. Albert, S. Voigt, and D. Schmidtchen, vol. 25 of *Conferences on New Political Economy*, pp. 101–122(2). Mohr Siebeck, Tübingen.

——— (2008b): "Life Cycle and Cohort Productivity in Economic Research: The Case of Germany," *German Economic Review*, 9, 431–456.

Ritzberger, K. (2008): "A Ranking of Journals in Economics and Related Fields," *German Economic Review*, 9, 402–430.

Rodríguez, D. (2006): "Publishing Performance of Spanish Academics: 1970-2004," *Spanish Economic Review*, 8, 255–270.

Schulze, G. G., S. Warning, and C. Wiermann (2008): "Zeitschriftenrankings für die Wirtschaftswissenschaften - Konstruktion eines umfassenden Metaindexes," *Perspektiven der Wirtschaftspolitik*, 9(3), 286–305.

Taylor, S. W., B. F. Fender, and K. G. Burke (2006): "Unraveling the Academic Productivity of Economists: The Opportunity Costs of Teaching and Service," *Southern Economic Journal*, 72(4), 846–859.

Turnovec, F. (2007): "Publication Portfolio of the Czech Economists and Problems of Rankings," *Ekonomický Casopis/Journal of Economics*, 55, 623–645.

Wooldridge, J. M. (2007): *Econometric Analysis of Cross Section and Panel Data*. MIT Press, Cambridge, MA.

Chapter 3

Measuring the Value of Research: A Generational Accounting Approach

3.1 Introduction

Wilhelm Conrad Röntgen discovered the X-rays (in German referred to as Röntgen rays) on 8 November 1895. He observed that the rays illuminated a fluorescent screen and noted that placing an object between the screen and the source emitting the rays reduced the effect but did not fully stop it. Röntgen published his first results on X-rays in the very last days of 1895. In this publication, Röntgen mentioned that photographic plates are sensitive to X-rays and that he had produced pictures of various objects, including a photograph showing the bones of his wife's hand. This news spread so fast, that the first radiological laboratories opened already in March 1896, only a few months after Röntgen's publication.[1] In 1901 Röntgen received the very first Nobel Prize in physics "in recognition of the extraordinary services he has rendered by the discovery of the remarkable rays subsequently named after him".[2]

The discovery of the X-rays illustrates how research benefits mankind, namely through applications. The use of X-rays improved the quality of medical treatment. Subsequent research has increased the value of X-rays further, by allowing for higher quality pictures, lower radiation and additional applications such as X-ray computed tomography. Although Röntgen's work constitutes only a small fraction of the research on X-rays that has been carried out ever since, his contribution left its imprint on any subsequent piece of research: X-ray computed tomography and other subsequent applications profess Röntgen's legacy. The value of Röntgen's discovery for mankind thus derives from direct and indirect contributions.

Alfred Nobel wanted his prize to be awarded "to those who, during the preceding year, shall have conferred the greatest benefit on mankind." Awarding committees are therefore obliged to assess the overall impact of a researcher's work on welfare. It is thereby not important whether the honored research generates direct valuable applications, as in the case of the X-rays, or whether such applications are likely to be forthcoming after further research and development as it was the case with regard to Einstein's theory of relativity. What is supposed to count for winning the Nobel Prize is that research will eventually give rise to applications, thereby improving the human condition.

Although the Nobel Prize is the highest distinction in science today, Alfred Nobel's

[1] see Glasser (1995) and the reprint of Röntgen's original publication in Glasser (1995).
[2] Nobel Prize homepage (http://nobelprize.org/nobel_prizes/physics/laureates/1901/rontgen.html)

selection criterion plays almost no role when it comes to judging less outstanding research. Evaluations of researchers nowadays measure a scientist's research performance in terms of citations to his or her work, the number of published papers, the outlets in which he or she has published, or some combination thereof. These criteria, however, provide no information on the intrinsic value of research since they measure only the creation of knowledge per se and not necessarily the direct or indirect creation of social welfare. The informative value of traditional research performance indices is thus limited, i.e. these indices are not helpful in guiding the science system towards an efficient use of research time and funds.

One of the arguably most important issues in research policy is to determine the overall amount invested in basic and applied research and to distribute these funds across the various disciplines and specific fields of investigation. The composition of the public research budget is however in large parts historically determined and, moreover, subject to rent-seeking activities. It would therefore be surprising if the marginal return of research, for example, in quantum physics was the same as in archeology.

This study proposes a research performance indicator that can, in principle (i.e. if the required information is acquired), provide researchers, managers of research institutions, and politicians in charge of research policy with an instrument that measures how efficiently scarce resources are used in the science system. Before presenting the basic idea of this indicator, I briefly survey the state of the art of research evaluation in order to show why the existing research performance indicators fail to assess the relative contribution of different scientific disciplines towards generating social welfare benefits.

Traditional measures of research production

Citations represent the basic ingredient of all traditional measures of research performance. It is therefore not surprising that the most frequently mentioned obstacle to comparing research productivity across disciplines consists in differences in discipline-specific citation habits. In particular the number of studies referenced in the average research item varies a great deal across disciplines. Standardizing indicators based on citation counts by normalizing discipline averages does of course not solve the problem as it renders all disciplines equally important by definition.

Instead of normalizing research indicators ex post, one can control for reference in-

tensity when calculating an indicator. In practice these two approaches produce different results, because of cross-citations between fields. Such interdisciplinary citation flows can be used to illustrate how research is influenced by different fields. The invariant method developed by Pinski and Narin (1976), for example, is a method to measuring journal quality that corrects for reference intensity, i.e. it corrects for the average number of references per study across disciplines.[3] In addition, the invariant method also weights citations according to the quality of the citing source. Palacios-Huerta and Volij (2004) show that the invariant method is the only method satisfying a set of desirable properties. This axiomatic foundation has certainly greatly contributed to establishing the invariant method as the state-of-the-art method of rating journals.

The quality weights that result from applying the invariant method to a set of journals measure the "intellectual influence" (Palacios-Huerta and Volij, 2004) of a research outlet, i.e. a journal's quality weight represents the journal's relative importance for subsequent research. Ritzberger (2008), for example, ranks economics journals using the invariant method. He finds that business-related fields rank significantly lower than other sub-fields of economics. This finding is due to the fact that business-related research is more applied and therefore contributes less input for subsequent research than more theoretical research. Consequently, the invariant method ranks applied journals systematically lower than journals specializing in publishing theoretical results or new research methods. Knowledge and utility are, however, two distinct concepts which do not need to coincide. It is therefore an open question whether the relatively poor performance of down-stream journals also holds in terms of utility.

It has repeatedly been stressed that there are many ways of evaluating research and that the selection of a specific evaluation method is often rather arbitrary.[4] Assessing research in terms of utility restricts this arbitrariness. Utility is, for example, independent of the prevailing citation habits. Invariance to reference intensity is thus not only a nice theoretical concept, it is an indispensable constituent of any utility based research assessment.

[3] The invariant method is also at the heart of Google's search algorithm.

[4] see, among others, Neary, Mirrlees, and Tirole (2003), Palacios-Huerta and Volij (2004), Kóczy and Strobel (2010), and Ravallion and Wagstaff (2010)

The basic idea of the VA-index

I now return to fleshing out the general idea of measuring research productivity by tracing, generation by generation, social welfare gains to their intellectual sources.

The purpose of research is not to have a direct effect on social welfare, but to provide insights to politicians, entrepreneurs and the individual how to achieve higher welfare. Research is thus an investment into future utility. Different research fields will, typically, affect different aspects of utility. Allocating resources in academia is, therefore, a classical investment decision in which preferences play a role. A rational decision maker needs to maximize his or her expected utility, given his or her information. To do so efficiently, a decision maker needs to know how research maps into utility.

Understanding how research affects utility goes beyond knowing in which fields applications emerge and how valuable these applications are. By restricting a discipline's relevance to applications, one ignores, for example, the fundamental contribution of mathematics to almost all disciplines. What one really needs to understand is how applications emerge, i.e. who contributed how much to the creation of a particular application.

Reconsider X-ray computed tomography, which was developed by Allan M. Cormack and Godfrey N. Hounsfield. Cormack and Hounsfield received the Nobel Prize in Physiology or Medicine in 1979. Besides Cormack and Hounsfield, Wilhelm C. Röntgen certainly also has an intellectual share in the creation of X-ray computed tomography. Moreover, Röntgen did not discover X-rays out of the blue: he relied on earlier research. More so, Cormack and Hounsfield relied not only on Röntgen's work, but also on other findings. The intellectual achievement of X-ray computed tomography is, therefore, not one of Cormack and Hounsfield alone, but one of numerous researchers, Röntgen included. The intellectual credit of X-ray computed tomography and all other inventions thus belongs to numerous researchers. Understanding how research maps into utility means to understand how and how much the individual idea has contributed to applications and thus to utility.

In the following I propose a generational accounting approach, that identifies the intellectual sources of research generated welfare. Moreover, this approach measures the intellectual influence of any research result on other research results and assigns the parent result a corresponding share of influence. The accounting feature of this method is that all shares of influence of a research result add up to one. The sources of intellectual contribution are thus exactly identified, measured, and distributed.

The proposed method allows to assess research performance in terms of utility by combining the value of applications with the corresponding shares of research influence. Measuring the value of applications is unfortunately way beyond the scope of this paper. Nederhof and Meijer (1995) argue that research does not improve living conditions directly, but that research results need to be transferred to users outside the science system to generate useful applications. Transferring knowledge to users outside the science system can therefore be regarded as final product of any scientific endeavor. The raison d'être of the VA-index is to identify the sources of value created by scientist for a given valuation of the observed research transfers.

The next section presents the theoretical concept of the VA-index. I also discuss some problems of applying the method to existing data. Section three presents a first implementation for economics, albeit a purely bibliometric one, since we do not have, so far, data on research transfer and a convincing exogenous measure for the value of applications. The bibliometric application ranks economics journals and individual articles published in the 1986-2004 period according to their respective influence on subsequent economic research. Section four concludes.

3.2 Research accounting

The VA-index is designed to measure the contribution of individual research results to the creation of value, i.e. to the utility accruing to the users of applications that embody these research results. The idea of the VA-index rests on the insight that research production requires research as one of its main production factors and this production process can be investigated and portrayed with standard economic techniques. In this respect the VA-index follows the basic idea of endogenous growth theory (see, for example, Romer, 1990).

Research provides two kinds of services: transfer of knowledge to the applied sector and intellectual input into further research. Transfer of knowledge to the applied sector constitutes the final product of academic research, i.e. the output. Intellectual input into further research, on the other hand, is an intermediate good in the production of knowledge. The value or revenue generated by research results thus depends on the knowledge transferred to the applied sector and on its contribution to subsequent research. The intellectual revenue of a research result is illustrated with the help of the right-hand account

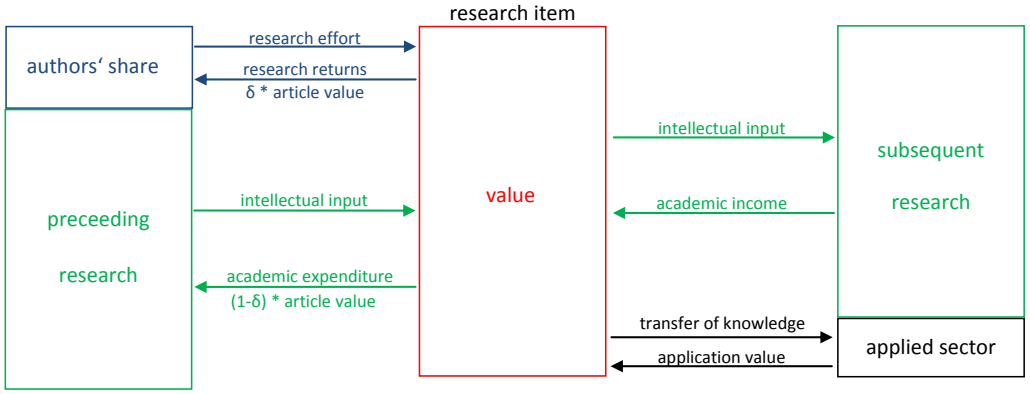


Figure 3.1: Research accounting at article level

in Figure 3.1.

Research relies on previous research results. Stigler and Friedland (1975), for example, refer to citations as "a form of intellectual collaboration". Researchers can therefore only claim a fraction δ of the value of their research results as their own intellectual contribution, i.e. their value added, while the remaining value belongs to the preceding literature upon which the results are built. The expenditure part of an article is portrayed by the left-hand account in Figure 3.1.

The VA-index identifies all direct intellectual sources of the research results and also the relevance of each individual source. It then follows a simple rule: the value of a research item is distributed to all direct intellectual inputs (researchers own input and preceding research), according to their respective influence on the result at hand. Applying this concept for all generations of research allows to identify all *indirect* intellectual sources of any research result because indirect sources are linked to the final research result via a finite chain of direct inputs. Consider a research result of a given value. A fraction of this result's value remains with the authors while the remaining part is passed on to the research results that provided direct inputs to this study. The value of the direct inputs is then also distributed on the research items that these direct inputs are based on. Thus, any transfer from a research result to its direct inputs will not only affect the direct inputs but also the direct inputs of the direct inputs. By continuing this process, value is mapped from one generation of research to the previous one, and ultimately to all items which provided direct and indirect inputs to a research result at some point of time.

Figure 3.2 illustrates how the value of research result A is mapped to its intellectual sources. Research results are depicted by two rectangles, a large one representing the value of the item and a smaller shaded one (in the upper left corner of the large rectangle),

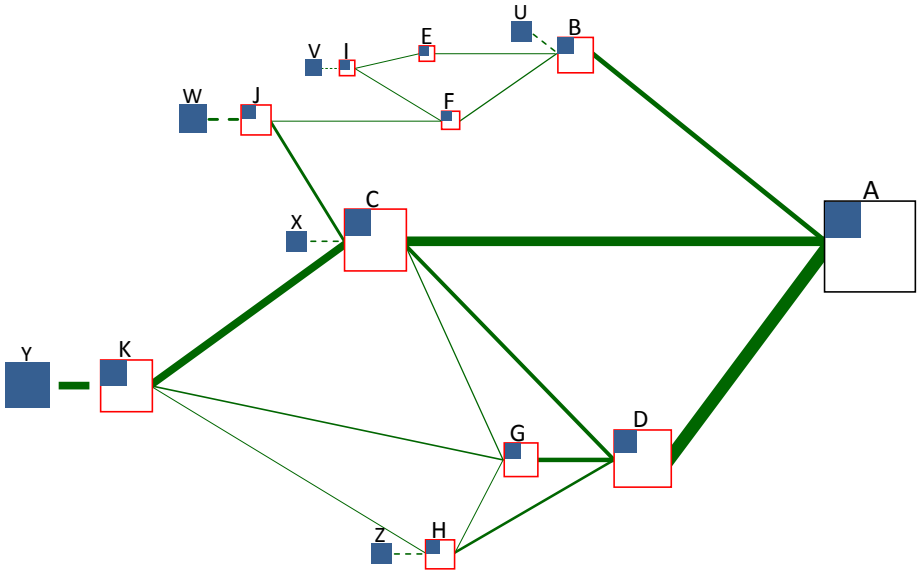


Figure 3.2: Research accounting between research items

representing the authors' contribution. The six items U to Z represent research results whose "ancestors" cannot be traced and therefore constitute starting points of the research production process. Connections between research results indicate intellectual input from the left to the right and the corresponding transfer of value in the opposite direction. The size of the rectangles and the width of the lines are proportional to the corresponding values and flows.

Research results B, C and D are the direct intellectual sources of A. C, moreover, is also an indirect source of A as it contributes to A via D and via G and D. The direct inputs produced by C imply that A uses some results from C, whereas indirect input means that A relies on results to which C has contributed, namely results D and G. Research result B receives a smaller transfer from A (represented by a thinner connection) than results C and D because its contribution to A is assumed to be smaller. Furthermore, the authors' shares (as a fraction of the result's value) also vary across items. Comparing results K and D illustrates that the authors of K have contributed more to their own results than the authors of D in relative as well as in absolute terms. Finally, all shaded areas add up to A's article value, so A's article value is completely distributed.

Whether a transfer stems from subsequent research or directly from an application plays no role for the construction of the VA-index. Figure 3.3 illustrates the crucial role of applications. Application *A1* builds on research results B and C, whereas applications *A2* and *A3* build on results C and D, respectively. The value of B (D), of course, corresponds

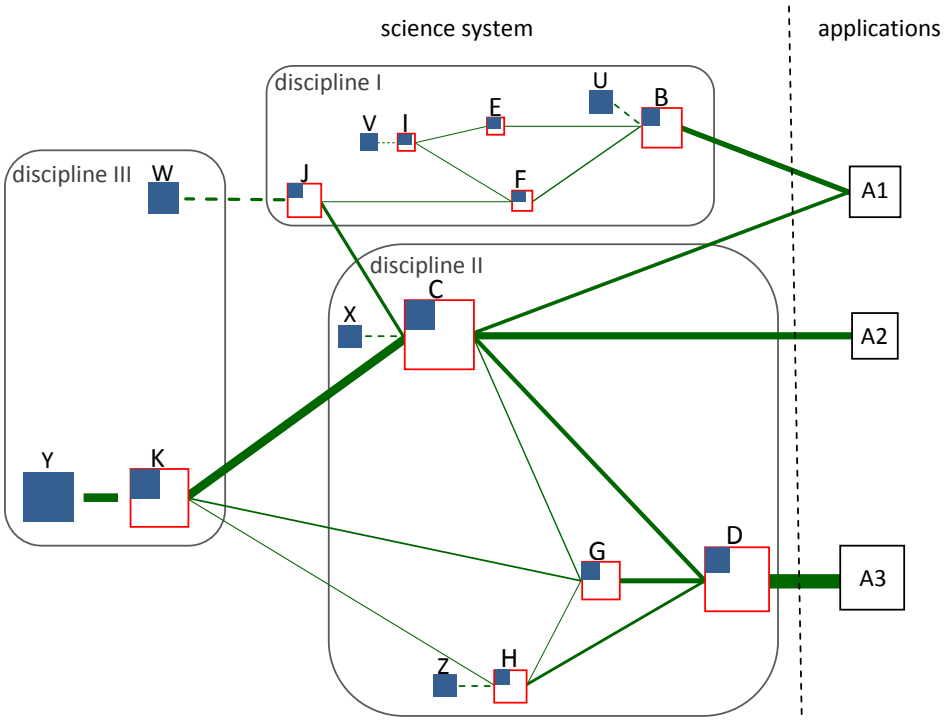


Figure 3.3: Research accounting from applications to academia

to its contribution to application *A1* (*A3*), whereas the value of C corresponds to its contributions to *A1* and *A2*.

In Figure 3.3 research is divided in three disciplines. The contribution of a discipline to the society's welfare gain equals the sum of value added (the shaded areas) of the individual research results belonging to this discipline. Note that *discipline III* has the largest impact on the creation of welfare, although it does not directly lead to any application.

Transferring knowledge to the applied sector is a dynamic process. Research results are absorbed with time lags and it takes time for applications to emerge. One would therefore expect that the (materialized) research value increases over time. Similarly, if a research study combines two or more results, the dynamics of value realization can change the relative importance of the study's intellectual sources. Consider, for example, a study that introduces a new econometric method and, by applying this method, also provides new insights into the mechanism of money supply. Central bankers are able to use the findings directly, whereas it takes some time for the new econometric method to give rise to other applications. The money supply innovation will, therefore, be most relevant in the beginning.

In the next subsection I outline a formal representation of generational research accounting and the VA-index.

Formal description

Let Γ be the set of intellectual contributions at a given point of time and let Γ comprise a total of N research results (items).

The value v_i of item i is split between the item's own contribution amounting to the fraction δ_i of the value and the contribution to the preceding literature that receives, via transfers, the remaining value in return for its intellectual input.

Let $c_{i\to j}$ denote the value of the direct intellectual input provided by item j to item i and the corresponding value transfer from i to j; $c_{i\to i}$ is zero by definition. The vector $c_i = (c_{i\to 1} \ \dots \ c_{i\to N})$ collects all direct inputs provided to item i. The transfer matrix $C = (c_1; c_2; \dots; c_N)$ describes all inputs and thus accounts for all inter-generational transfers of value added.

When mapping value transfers across many generations of items, it is convenient to work with transfers in relative terms. The share of total value of item i transferred to item j is given by $d_{i\to j} \equiv c_{i\to j}/v_i$. By construction $\sum_{j=1}^{N} d_{i\to j} = 1 - \delta_i$, since the authors' contribution δ_i is not passed further on. Moreover, let $d_i = (d_{i\to 1} \ \dots \ d_{i\to N})$, $D = (d_1; d_2; \dots; d_N)$ and $a_{i\to j}$ be a dummy variable assuming the value 1 if research result j contributes to research result i ($c_{i\to j} > 0$) and $a_{i\to j} = 0$ otherwise.

An indirect link of order q connecting item i to item j is a chain $a_{i\to k_1} \ a_{k_1\to k_2} \ \dots \ a_{k_q\to j}$, with no zeros.[5] That is, an indirect link of order q is a link passing through q generations of research results.

Note that research results can be linked through many different chains or paths connecting item i with item j. For example items A and K in Figure 3.2 are indirectly linked of order 1 via $A \to C \to K$, of order 2 via $A \to D \to C \to K$, $A \to D \to G \to K$ and $A \to D \to H \to K$, and of order 3 via $A \to D \to G \to C \to K$ and $A \to D \to G \to H \to K$.

Indirect links reveal the channels through which research results affect each other. The order of an indirect link indicates the importance of the link. The higher the order of a link, the smaller is the value passing through this channel, because each generation of items between the two ends keeps a fraction of value.

The share of indirect transfer via a link of order $q \geq 1$ between item i and item j is

[5] Formally, if $i = k_0$ and $j = k_{q+1}$, then an indirect link of order q between i and j via $k_1, \dots, k_q$ exists, if $\prod_{n=0}^{q} a_{k_n\to k_{n+1}} = 1$.

given by

$$\begin{aligned} id_{i\to j,q} &= id_{i\to 1,q-1}\, d_{1\to j} + \ldots + id_{i\to N,q-1}\, d_{N\to j} \\ &= \sum_{k=1}^{N} id_{i\to k,q-1}\, d_{k\to j}. \end{aligned} \tag{3.1}$$

Adding over all research results in Γ implies that all potential paths are taken into account. Because a share δ_k of value at level $q-1$ remains with the contributing research result, only the remaining fraction is passed on. Direct transfers are indirect transfers of order zero, i.e. $id_{i\to j,0} \equiv d_{i\to j}(1-\delta_i)$. Finally, define $id_{i,q} = (id_{i\to 1} \quad \ldots \quad id_{i\to N})$ and $ID_q = (id_{1,q}; \ldots; id_{N,q})$.

Using the concept of indirect transfers, one can express the transfer of value from applications to their intellectual sources and the resulting measures of value-added in the following manner:

$$\mathtt{va} = \Delta \left[\mathtt{u} + \sum_{q=0}^{Q} ID_q' \mathtt{u} \right], \tag{3.2}$$

where $\mathtt{va}$ is the resulting vector of value added, Δ is a diagonal matrix with the individual δs on the main diagonal, $\mathtt{u}$ is the utility gain deriving from each item's knowledge transfer to the applied sector, and Q is the order of the longest indirect link.

$\mathtt{va}$ is a measure of valuation of individual research results. Based on this measure one can derive (1) valuations of individual scientists, (2) entire research units, (3) evaluations of journal quality, and (4) evaluations of entire disciplines.

The contribution of an arbitrary research entity e is given by:

$$\mathtt{va}_e = \sum_{i\in E} \mathtt{va}_i * s_{i,e}, \tag{3.3}$$

where E is the set of research projects which entity e was involved in, and $s_{i,e}$ captures the entity's share in result i. If a result has three authors, each of them might be credited by $s = \frac{1}{3}$.

Since the creation of research results and the creation of the utility related to these results are separated in time, the costs and benefits are not directly comparable. One therefore may want to account for time preferences by adding a discount factor to equation 3.3. Discounting valuations is also appropriate when comparing different disciplines. Suppose an average result in mathematics has a larger impact on welfare than a result

in economics. If the average impact time in mathematics is much longer than in economics, the effective returns from mathematics can actually be smaller than those from economics.[6]

Many theoretical concepts in economics face problems when it comes to implementing those concepts. Generational research accounting makes no exception. First, measuring the contribution u of a research result to the applications is a challenging task. Second, some sources that provided intellectual input to a research result may not be documented, be it because they were forgotten or, as for example in the case of the Nash equilibrium (Nash, 1950), because an explicit reference appears to be redundant. Due to a lack of better alternatives one is, unfortunately, often forced to use citations. In bibliometrics this is common practice, but nevertheless it is a rather crude escape.[7] The third problem of implementing generational research accounting concerns the relevance of the sources for a research result. Citations only indicate the incidence of intellectual input, not its impact. Even worse, Serrano (2004) points out that citations sometimes refer to bad contributions to a topic. Finally, the share of a research item's value that the authors can claim as their contribution is hard to measure and likely to differ from item to item.

All these problems have to be treated appropriately, if the full potential of the VA-index is to be reached. Yet, even while many tough problems remain, the VA-index can already now provide valuable services to the scientific community. Comparing the performance of different social science disciplines, for example, is less demanding than comparing the value added of the social sciences with the value added of the natural sciences, because all social sciences use similar channels of transferring their output to the applied sector, implying that measurement errors will affect all sub disciplines in the same manner.

One way of circumventing many of these problems is to compile VA-indices that portray only the science system, i.e. indices that do not attempt to measure utility in terms of application values but only the value of research for subsequent research as measured by standard bibliometric data. The next section presents such an indicator which is based on citation flows only. Although the resulting measures cannot be regarded as reliable measures of the value of research for the society at large, one can gain from this exercise first insights into how the VA-index performs as compared to other bibliometric measures

[6] Note, that the VA-index allows not only to compare the returns from different areas of research, but also to compare the returns from research to returns from other activities, e.g. from infrastructure projects.

[7] Ritzberger (2008), for example, provides a comprehensive list of shortcomings of relying on citation data.

of research performance.

3.3 The intellectual sources of research in economics: applying the VA-index to citation data

In this section I apply the VA-index using only standard citation data and then compare the resulting evaluations of journal quality and article impact with alternative bibliometric measures. Since we are, at the time being, not able to measure the application value of economic research, the presented VA-indices are restricted to identifying the intellectual sources of research in economics conducted between 2006 and 2008.

The data

I use citation data collected from the Social Science Citation Index (SSCI) for the years 1986 to 2008.[8] The 1986 edition also indexes some journal articles published in 1985. I only consider articles published in journals which were classified by the SSCI as economics journals in 2010. The resulting set of items is limited to items classified as an article, discussion, letter, note, reprint or review. This restriction yields a set of 162,745 research items. Within this set I identify 972,815 citations. For the most recent years almost thirty percent of the references can be identified (the remaining references are not indexed in the considered subsample of the SSCI).

Basic assumptions

I assume that citations indicate direct intellectual input from the cited article to the citing article. Since citations only indicate the incidence of intellectual input between articles, I further assume that all sources that an article cites are equally important. Let $a_{i\to j}$ indicate whether article i cites article j ($a_{i\to j}=1$) or not ($a_{i\to j}=0$), i.e.

$$a_{i\to j}=\begin{cases}1 & \text{if } c_{i\to j}>0\\ 0 & \text{if } c_{i\to j}=0.\end{cases}$$

Given a complete set of literature which covers all citable items, any reference can

[8] Ursprung and Zimmer (2007) show that citation data is prone to errors and that such errors can have a significant effect on citation-based ratings. Readers should be aware of the possibility of such errors when it comes to interpreting the results below.

be matched with the corresponding article. The data I use is, however, incomplete: the SSCI mainly covers journal articles, not all journals are indexed by the SSCI, and the electronic record upon which I rely dates back only to 1985. The last limitation causes older articles to receive a disproportionately higher share of citations: an article published in 2008 can cite articles published between 1985 and 2008, whereas an article published in 1985 can only cite articles published in the same year. Old articles will therefore receive more citations and appear to be more influential than they actually are. To control for this bias one can normalize citations by the number of references instead of the number of citations (identified in the dataset). Normalizing transfers by references causes, however, the fraction of value that an article passes to the previous generation to vary with the share of identified references. I evade this problem by introducing a year dependent correction factor that normalizes the citation to reference ratio to the ratio in 2008.[9] The fraction of value of article i transferred to article j is then given by:

$$d_{i\to j} = cf_i \; a_{i\to j} / \left(\sum_{k=1}^{N} a_{i\to k} \right),$$

where cf_i is the correction factor of the year in which article i was published.

I calculate shares of indirect transfer using $id_{i\to j}$ as given in equation 3.1 and assume that the authors' shares are constant across articles, i.e. $\delta_i = \delta \; \forall i$. Finally, I discount the value of a publication by 3.065 percent a year which corresponds to the average real interest on ten year US treasury bonds between 1986 and 2008.

3.3.1 Rating journal quality

The invariant method

Evaluating journal quality on the basis of citations has been popular among scientists ever since Garfield (1972) proposed the journal impact factor. While the impact factor considers only the number of citations, Liebowitz and Palmer (1984) proposed to take the citing journal's quality into account. The invariant method developed by Pinski and Narin (1976) controls not only for citation quality but also for reference intensity, i.e.

[9] In 2008 almost 30% of the references can be identified as citations. In 1992 this share is 15%. Thus, I assume that another 15% of references from 1992 articles could be identified if the dataset covered 16 more years. The number of citations would be twice the number reported. Therefore items which are cited by articles published in 1992 will only receive half the transfer they would have obtained if the citing article was published in 2008.

the number of items an article refers to. Palacios-Huerta and Volij (2004) establish the invariant method as current the state-of-the-art method to rate journals by showing that the invariant method is the only method that satisfies a set of desirable properties.

I therefore regard the invariant method to represent the natural benchmark for the VA-index. Invariant quality weights of those journals which have had full coverage in the SSCI between 2002 and 2008 are reported in the second column of Table 3.4 in the appendix. Table 3.1 below presents a selection of the top-ranked journals.[10] In contrast to most other applications of the invariant method, I control for reference intensity at the article level and not at the aggregate journal level. Palacios-Huerta and Volij (2004) point out that controlling for reference intensity at the article level gives an equal vote to each article in a journal, whereas controlling for reference intensity at the journal level causes articles to receive a disproportional high (low) weight if they have relatively many (few) references. The citing and the cited period are both 2006-2008. I omit self-citations at the level of journals, because some journals, for whatever reasons, show excessive rates of self-citations. The resulting rating with the three year window 2006-2008 is labeled *Inv 3y*. Journal weights are normalized so that the top journal has a score of 100.

The journal rating *Inv 3y* corresponds to the state of the art in bibliometrics, but it is nevertheless not the best rating of the invariant type. In the following I will elaborate on the shortcomings of the standard invariant rating and present a superior invariant rating.

Journal ratings usually rely on rather short cited periods.[11] Limiting the cited period gives, however, editors and thereby researchers an incentive to focus on hot topics which are likely to attract citations quickly. Promising research, which is however not en vogue, is more likely to remain unpublished, since the journal's rating will not benefit from a paper's late success. Furthermore, manipulation attempts have a larger effect on the resulting rating if it relies on a small data set: short cited periods thus increase the incentives for strategic manipulations.

Differences in journal coverage are not a big issue when ratings are based on short cited periods because journals with an incomplete citation record are either excluded or treated as if all data were available. For long cited periods, however, differences in

[10] The restriction on the coverage period causes that the Journal of the European Economic Association (JEEA) is not included in the primary set of journals. However, a ranking of the JEEA is provided on a less restrictive basis.

[11] Palacios-Huerta and Volij (2004) use seven years, Kalaitzidakis, Mamuneas, and Stengos (2003) five years and Ritzberger (2008) three years

journal coverage become more problematic, since more journals are affected and the effects resulting from incomplete coverage become more pronounced. Figure 3.4 illustrates how a journal's imputed quality depends on its coverage if journal quality is measured as the total number of citations divided by the total number of articles published in the relevant period. The first panel in Figure 3.4 shows the distribution of citations in articles published in 2006-2008 of articles published in 1985-2008. The second panel presents the corresponding cumulative density function. The last panel reports citations per article for a fictitious journal with coverage from 2008 back to the year indicated on the x-axis. Citations per article peak for a journal with a track record of 12 years (i.e. a journal whose coverage commences in 1997). Comparing the rating of a journal with twelve years of data coverage to the rating of the same journal with 23 years of coverage suggests that articles of the "older" journal have almost 20% lower quality. These differences do, however, not reflect differences in journal quality, they rather arise because of the specific pattern of the citation life cycle.

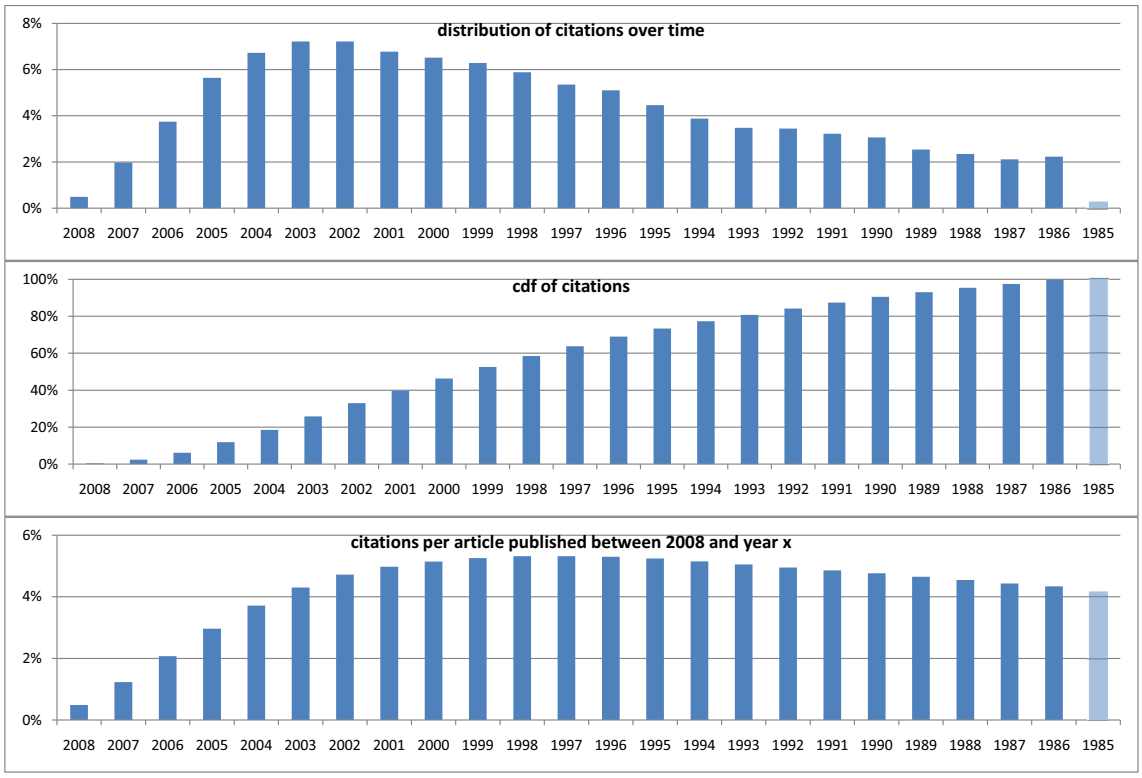


Figure 3.4: Journal coverage and citation life cycle

To control for variations in journal coverage I weight citations by a year specific weight: all articles published in a year receive a weight proportional to the fraction ϕ of citations to articles published in this year as represented by the first panel of Figure 3.4. Each

individual article then receives a weight equal to ϕ/n, where n is the number of articles published in the respective journal and year. If a journal does not have full coverage, I rescale the journal's year specific weights to add up to one.

Column three in Table 3.1 reports the journal rating using the described correction for coverage. It is based on citations by articles published between 2006 and 2008 of articles from all years covered in the data set. The results would be identical to the standard invariant method, (1) if the journal size was constant over time, (2) if all journals had full coverage, and (3) if all years received the same number of citations. The rating is labeled *Inv* 24y*.

Table 3.1: Comparison of journal ratings

Journal	Inv 3y	Inv* 24y	VA-index w	VA-index u
QUARTERLY JOURNAL OF ECONOMICS	96.37 (2)	100 (1)	100 (1)	100 (1)
JOURNAL OF POLITICAL ECONOMY	86.97 (3)	81.59 (3)	71.78 (2)	69.3 (3)
ECONOMETRICA	100 (1)	85.23 (2)	70.23 (3)	68.65 (4)
JOURNAL OF ECONOMIC LITERATURE	51.13 (6)	73.02 (4)	68.36 (4)	99.87 (2)
AMERICAN ECONOMIC REVIEW	75.23 (5)	59.81 (6)	56.23 (5)	64.35 (5)
REVIEW OF ECONOMIC STUDIES	80.71 (4)	63.5 (5)	53.04 (6)	48.3 (7)
BROOKINGS PAPERS ON ECONOMIC ACTIVITY	7.1 (56)	35.27 (7)	40.31 (7)	52.51 (6)
JOURNAL OF ECONOMIC PERSPECTIVES	34.95 (11)	27.27 (10)	28.19 (8)	39.25 (8)
JOURNAL OF FINANCIAL ECONOMICS	28.74 (15)	27.08 (11)	27.66 (9)	35.6 (11)
JOURNAL OF ECONOMIC GROWTH	24.18 (22)	28.05 (9)	25.64 (10)	37.67 (10)
JOURNAL OF MONETARY ECONOMICS	25.89 (18)	24.7 (13)	25.32 (11)	32.3 (12)
RAND JOURNAL OF ECONOMICS	40 (8)	30.01 (8)	24.68 (12)	30 (13)
JOURNAL OF ECONOMIC THEORY	39.11 (9)	26.61 (12)	21.55 (13)	21.2 (28)
ECONOMIC POLICY	27.26 (16)	18.69 (18)	20.7 (14)	38.74 (9)
JOURNAL OF LABOR ECONOMICS	31.2 (14)	20.6 (14)	20.12 (15)	24.84 (21)
REVIEW OF ECONOMICS AND STATISTICS	40.65 (7)	20.4 (15)	20.07 (16)	29.58 (14)
INTERNATIONAL ECONOMIC REVIEW	35.82 (10)	19.93 (16)	17.86 (17)	20.3 (32)
JOURNAL OF HUMAN RESOURCES	24.26 (21)	16.32 (23)	17.54 (18)	23.82 (22)
JOURNAL OF LAW & ECONOMICS	31.92 (13)	17.53 (20)	16.96 (19)	20.45 (31)
GAMES AND ECONOMIC BEHAVIOR	32.69 (12)	19.16 (17)	16.7 (20)	16.84 (40)
AER Papers & Proceedings	18.12 (30)	15.81 (24)	16.62 (21)	21.59 (25)
JOURNAL OF ECONOMETRICS	23.92 (23)	18.06 (19)	16.2 (22)	26 (18)
JOURNAL OF INTERNATIONAL ECONOMICS	26.48 (17)	14.58 (28)	16.08 (23)	26.7 (17)
REVIEW OF ECONOMIC DYNAMICS	20.86 (26)	16.57 (21)	15.09 (24)	16.13 (44)
ECONOMIC JOURNAL	25.56 (20)	15.26 (25)	14.66 (25)	25.76 (19)
JOURNAL OF FINANCIAL AND QUANTITATIVE ANALYSIS	25.8 (19)	16.43 (22)	14.46 (26)	17.53 (37)
JOURNAL OF ACCOUNTING & ECONOMICS	7.87 (52)	13.85 (29)	13.65 (27)	13.73 (51)
JOURNAL OF BUSINESS & ECONOMIC STATISTICS	22.97 (24)	14.87 (27)	13.41 (28)	21.34 (26)
JOURNAL OF LAW ECONOMICS & ORGANIZATION	8.9 (48)	13.65 (30)	12.62 (29)	17.27 (38)
JOURNAL OF PUBLIC ECONOMICS	19.33 (27)	12.38 (32)	12.39 (30)	19.98 (33)
JOURNAL OF INDUSTRIAL ECONOMICS	21.46 (25)	13.56 (31)	11.93 (31)	21.13 (29)
JOURNAL OF APPLIED ECONOMETRICS	18.2 (29)	15.04 (26)	11.77 (32)	20.66 (30)
WORLD BANK ECONOMIC REVIEW	10.55 (39)	10.4 (39)	11.39 (33)	26.96 (16)
MATHEMATICAL FINANCE	10.2 (42)	10.89 (36)	10.54 (35)	26.97 (15)
JOURNAL OF HEALTH ECONOMICS	6.25 (65)	8.49 (43)	8.87 (38)	21.31 (27)
ECONOMIC THEORY	19.09 (28)	11.1 (35)	8.29 (41)	11.41 (72)
WORLD BANK RESEARCH OBSERVER	10.42 (41)	7.87 (44)	8.23 (42)	22.28 (24)
JOURNAL OF ECONOMIC GEOGRAPHY	4.56 (75)	5.59 (60)	5.61 (54)	23.3 (23)
JOURNAL OF ENVIRONMENTAL ECONOMICS AND MANAGEMENT	6.28 (64)	5.7 (57)	5.48 (56)	24.95 (20)

Figure 3.5 illustrates the differences between *Inv 3y* and *Inv* 24y*. Differences can reflect field specific citation life cycles, changes in journal quality, stochastic variations, or a combination thereof. The Brookings Papers on Economic Activity, the Journal of Economic Growth, and the Journal of Risk and Uncertainty are among those journals that profit most from extending the cited period. Many journals related to finance also profit from extending the cited period, whereas many journals specializing in international economics and economic development fare better in the first rating (*Inv 3y*).

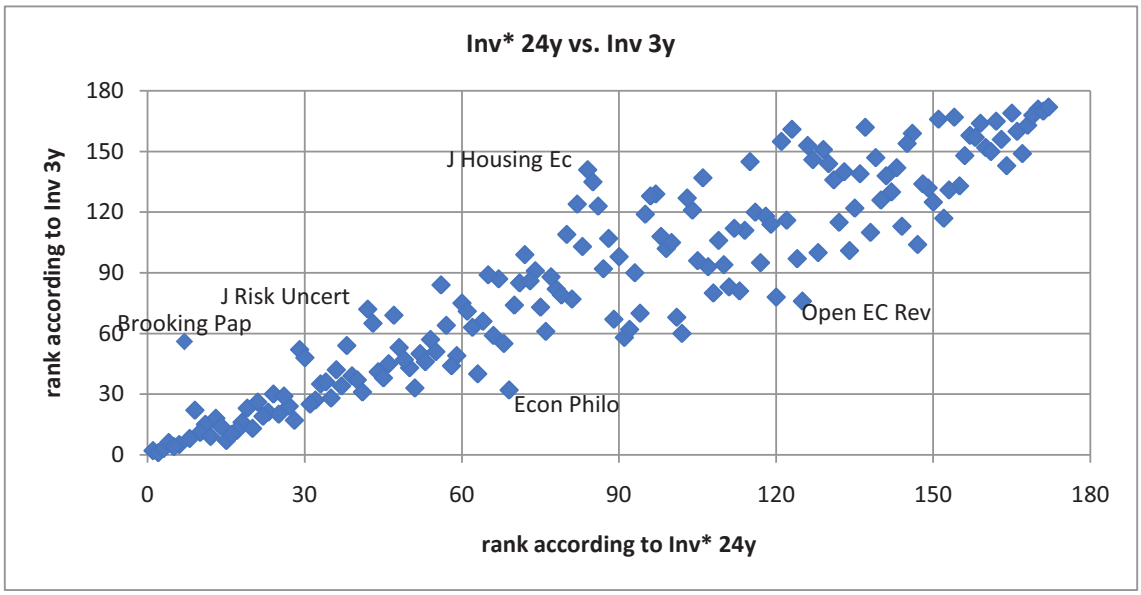


Figure 3.5: Scatter plot of journal ranks according to Inv* 24 and Inv 3y

The VA method

Columns four and five of Table 3.1 present journal ratings based on the VA-index. Assuming that all direct citations of an article are equally important causes older articles to have on average larger VA-indices, because older articles receive both direct as well as indirect returns for their contributions. This article, for example, cites Garfield (1972) and several articles citing Garfield (1972). The direct contribution of Garfield (1972) to this paper, however, is less important than its refinement by Pinski and Narin (1976).[12] Assuming that all citations are equally important will, in general, lead to an overestimation of the direct contribution of older studies. In computing the VA-index, I therefore use yearly weights as in the case of *Inv* 24y*. I also discount article scores by 3.065 percent a year. Moreover, I normalize article scores such that the sum of the article scores in each year equals the number of articles published in this year. Finally, I set δ equal to 0.164, because this value implies that the growth rate of knowledge roughly equals the growth rate of the economy.[13]

I present two versions of the VA-index: Both versions assign exogenous application

[12]Note, that this statement is different from claiming that the overall contribution of Pinski and Narin (1976) is larger than the one by Garfield (1972) as the statement above relates only to direct and not to overall contributions.

[13]The average citation time in my dataset is 12.7 years including citations to articles published before 1985. The average annual productivity growth in OECD countries between 1985 and 2006 has been around 1.2%. If research drives productivity the two growth rates should be the same in the long run. For an illustrative exercise as this one the exact value is not so important, as the estimates at journal level are quite robust to changes in δ. The correlation between the ratings using $\delta = 0.164$ and $\delta = 0.1$ is 0.99998, the correlation between ratings using $\delta = 0.164$ and $\delta = 1/3$ is still 0.99989.

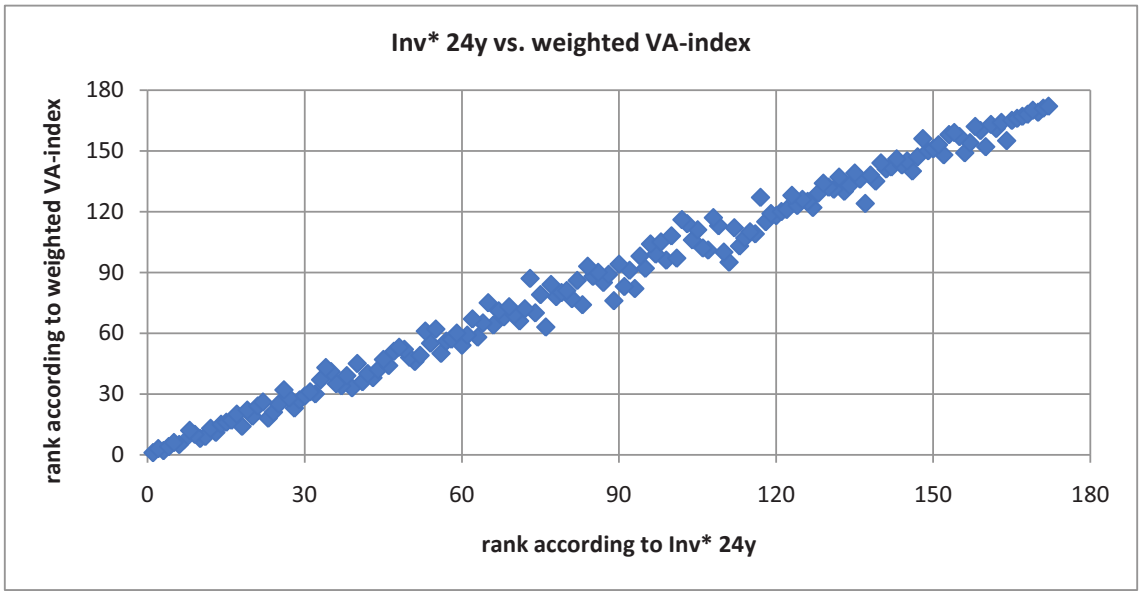


Figure 3.6: Scatter plot of journal ranks according to Inv* 24 and weighted VA-index

values to articles published between 2006 and 2008 and then use citations by all articles in the sample to identify the intellectual sources. For the *VA-index w* I assume that a 2006-2008 article's value equals the quality weight of the journal in which the article was published. Journal quality weights are thus determined by a system of equations. The *VA-index u*, on the other hand, relies on unweighted 2006-2008 article value, i.e. every article's application value is normalized to one. In both cases the VA-indices of the quality of articles published between 1986 and 2008 are calculated according to equation 3.2 and subsequently discounted.

Figure 3.6 shows that the journal rating based on the *VA-index w* is very close to the rating based on *Inv* 24y*. The correlation between the two ratings is 0.9946 (based on journal scores, not on journal ranks). The ratings of Economic Policy (*Inv* 24y* rank 18; *VA-index w* rank 14), the Journal of Human Resources (23; 18), and the Journal of International Economics (28; 23) increase significantly if value added is taken into account. Many journals devoted to economic theory, however, rank lower according to the *VA-index w*.

Investigating the effects of discounting and of applying the VA-method separately suggests that both have a small but systematic negative effect on the rating of theory journals (see Table 3.5 in the appendix for the effect of discounting, and Table 3.6 for the effect of the VA-method). Other differences between the two methods used to compile *Inv* 24* and *VA-index w* seem to disappear at the aggregation level of journals.

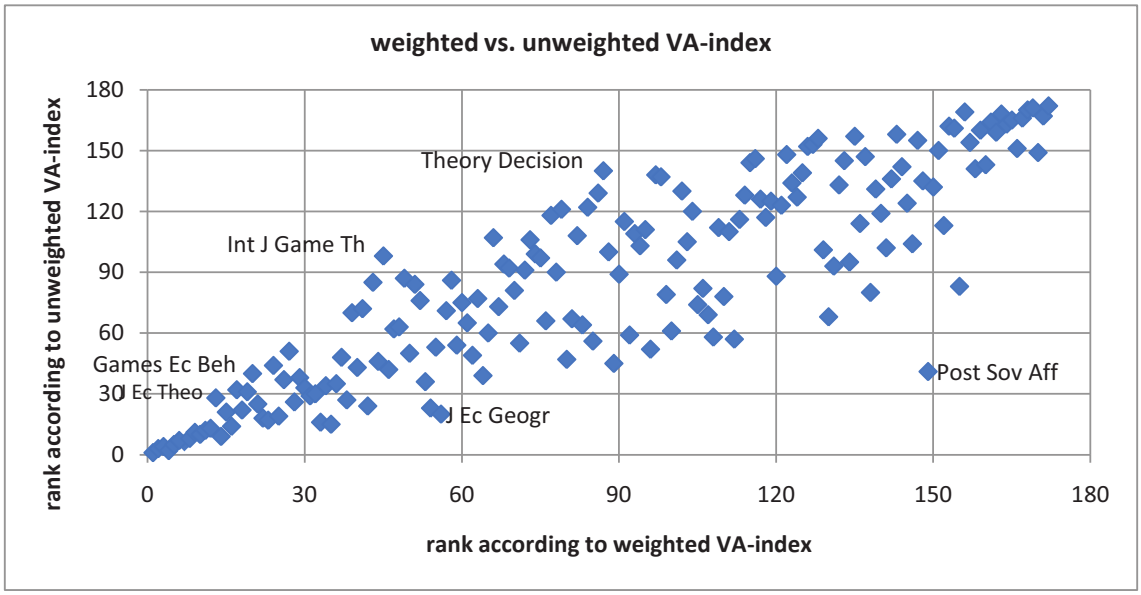


Figure 3.7: Scatter plot of journal ranks according to weighted and unweighted VA-index

One can argue that traditional journal ratings discriminate against more applied journals because the transfer of knowledge to the applied sector and the creation of knowledge are treated as being one thing. Therefore *VA-index u* presents a reference case giving equal voting power to all articles. The *VA-index u* rating is reported in the last column of Table 3.1.

Figure 3.7 illustrates the impact of weighting the value of the target items (2006-2008) on the resulting rating. It is no coincidence that the four journals which profit most from weighting are theory journals. Economic theory is the beneficiary of weighting. This finding is perfectly intuitive. Journals specializing in research on transition economics, development, geography, agriculture, or natural resources, on the other hand, fare much better if all 2006-2008 items have equal voting power. By comparing the two ratings one can easily identify the most applied fields of economics, i.e. those fields which provide relatively little input into the progress of general economic thought.

Table 3.2 provides the coefficients of correlation between the different ratings. All four ratings are highly correlated. This is a typical feature of journal ratings. *Inv* 24* and *Imp w* are the two most highly correlated ratings.

In concluding let me emphasize again that the purely bibliometric application of the VA-index presented in this section may provide more reliable ratings than the invariant method because the VA-index controls for citation quality at article level. Controlling for citation quality at article level is a logical and important step in computing journal

Table 3.2: Correlation of journal ratings

	Inv 3y	Inv* 24y	Imp w	Imp u
Inv 3y	1	0.9530	0.9369	0.8632
Inv* 24		1	0.9946	0.9433
Imp w			1	0.9589
Imp u				1

ratings. Oswald (2007), for example, shows that article quality varies substantially within journals. Palacios-Huerta and Volij (2004) point out that applying the invariant method to individual articles is not possible because older articles cannot cite newer ones, implying that a reciprocal relationship cannot be established. If, however, one uses indirect citations of degree one or higher degrees one can very well construct a mapping from journals to journals which takes article specific citation quality into account. In this case, an item does not profit from being directly cited but merely links direct citations to its own references.

3.3.2 Rating article impact

Measuring the value of individual articles is far more challenging than measuring the quality of journals. Data errors, for example, often cancel out at the journal level. At the level of individual articles, data errors translate, however, directly into erroneous results. The results presented in this subsection therefore need to be interpreted with great caution. In any event, my aim is to illustrate the VA method as applied to the evaluation of individual research items.

Table 3.3 reports normalized scores of journal articles according to the number of citations, citations weighted by journal quality, and the VA-index. The citing period is again 2006 to 2008. Moreover, all three article scores are discounted by three percent a year. Contrary to the journal ratings presented in the last subsection, I had to include self citations, because I can only identify self citations at the journal level and not at the level of individual authors. The table reports all articles which make it into the top twenty according to at least one of the three criteria.

It is not surprising that the rank variation at the article level is much higher than the rank variation at the journal level. The differences are remarkable considering that all ratings are based on the same data. Still, nine articles are among the top twenty according to all three criteria. Some articles even receive quite similar ratings, for example the 1998

Table 3.3: Rank comparison at article level

article	#citations	weighted citations	VA-index
Arellano, Bond (Rev Ec Stud, 1991)	100 (1)	51.63 (17)	48.47 (24)
Artzner, Delbaen, Eber, Heath (Math Finance, 1999)	99.9 (2)	39.64 (32)	40.37 (38)
Engle, Granger (ECONOMETRICA, 1987)	93.02 (3)	25.7 (119)	51.74 (20)
Clarida, Gali, Gertler (JEL, 1999)	90.65 (4)	58.05 (11)	61.7 (12)
La Porta, Lopez-de-Silanes, Shleifer, Vishny (JPE, 1998)	84.82 (5)	66.7 (7)	80.78 (3)
Acemoglu, Johnson, Robinson (AER, 2001)	78.94 (6)	95.34 (2)	67.33 (7)
Newey, West (ECONOMETRICA, 1987)	77.2 (7)	83.34 (3)	100 (1)
Romer (JPE, 1990)	73.99 (8)	35.29 (40)	61.5 (13)
Lucas (JME, 1988)	71.89 (9)	29.29 (74)	82.61 (2)
Bollerslev (J Econometrics, 1986)	69.11 (10)	28.41 (86)	41.15 (34)
Johansen (J EC Dyn Contr, 1988)	68.82 (11)	16.44 (378)	27.04 (107)
Bertrand, Duflo, Mullainathan (QJE, 2004)	68.3 (12)	100 (1)	65.68 (9)
Christiano, Eichenbaum, Evans (JPE, 2005)	66.64 (13)	70.06 (4)	42.82 (28)
Im, Pesaran, Shin (J Econometrics, 2003)	65.94 (14)	16.6 (365)	10.36 (886)
Staiger, Stock (ECONOMETRICA, 1997)	63.15 (15)	66.08 (8)	62.27 (11)
Fama, French (J Financial Ec, 1993)	62.47 (16)	69.86 (5)	67.7 (6)
Kwiatkowski, Phillips, Schmidt, Shin (J Econometrics, 1992)	62.27 (17)	20.13 (228)	16.35 (358)
Clarida, Gali, Gertler (QJE, 2000)	61.4 (18)	59.66 (9)	56.61 (17)
Fehr, Schmidt (QJE, 1999)	60.8 (19)	69.63 (6)	56.38 (18)
Hall, Jones (QJE, 1999)	59.06 (20)	48.66 (20)	48.92 (23)
Romer (JPE, 1986)	58.69 (21)	21.71 (190)	70.31 (4)
Jensen (AER Papers Proc, 1986)	48.27 (30)	42.41 (25)	56.62 (16)
Bolton, Ockenfels (AER, 2000)	45.65 (35)	52.48 (15)	42.01 (32)
Melitz (ECONOMETRICA, 2003)	43.56 (38)	57.16 (13)	38.19 (42)
Laibson (QJE, 1997)	41.25 (41)	57.82 (12)	60.06 (14)
Grossman, Hart (JPE, 1986)	37.62 (46)	34.34 (45)	66.18 (8)
Barro (QJE, 1991)	36 (52)	13.19 (598)	62.34 (10)
Bils, Klenow (JPE, 2004)	31.92 (67)	54.31 (14)	35.09 (51)
Gul, Pesendorfer (ECONOMETRICA, 2001)	25.74 (118)	58.44 (10)	42.58 (30)
Katz, Murphy (QJE, 1992)	24.79 (130)	28.35 (87)	70.03 (5)
Summers, Heston (QJE, 1991)	23.79 (141)	13.14 (605)	58.9 (15)
Miguel, Kremer (ECONOMETRICA, 2004)	12.01 (673)	50.07 (18)	31.18 (77)
Shimer, Smith (ECONOMETRICA, 2000)	11.2 (799)	49.51 (19)	31.51 (76)
Strömberg (QJE, 2004)	10.89 (861)	51.97 (16)	35.31 (50)
Summers, Heston (Rev Inc Wealth, 1988)	5.81 (3114)	3.11 (5749)	52.43 (19)

JPE article by La Porta and co-authors.

"The Penn World Table (Mark 5)" published by Summers and Heston in the Quarterly Journal of Economics (QJE) is a different case. "The Penn World Table (Mark 5)" takes rank fifteen according to the VA-index, but is ranked only 141st according to the number of citations and 605th according to weighted citations. This implies that the Penn World Tables are less intensively used in current research but they nevertheless made an important indirect contribution to the literature which is still noticeable. The indirect contributions are even more pronounced for the last article in Table 3.3 which is an earlier version of the Penn World Tables. The earlier versions of the Penn World Tables laid the foundation for the version published 13 years later in the QJE. Ratings based the two traditional indices fail, however, to capture this important input. The ratings of the two articles by Romer (JPE 1986 and 1990) tell a similar story. Romer's 1990 article "Endogenous Technological Change" ranks higher according to the two traditional indices, but not according to the VA-index. Here, the VA-index again captures the contribution of the earlier and maybe more innovative work.

Given the variation of article quality within journals, the VA-index is likely to be more reliable than ratings based on unweighted or weighted citations. Moreover, the VA-index also reduces incentives for strategic manipulations since the cited source receives only a

fraction of the credit and it would be rather tedious to figure out who else is going to profit from such a manipulation.

3.4 Conclusion

In this paper I have presented a theoretical basis for an ideal measure of research performance, the VA-index. I have also provided a first implementation of the index to research in economics, which illustrates some features of the index. The implementation illustrates the gap that exists between the current potential of bibliometric data and the informational requirements of the VA-index. Closing this gap is a challenge for future research. In the meantime, the theoretical concept of the VA-index provides a point of reference for existing indicators.

Fortunately, improvements to research assessment are fairly easy to make. Campbell (2008) mentions that Nature's editorial policy followed some medical journals in now giving authors the option to report contributions of particular authors to a piece of research. Moreover, he mentions the possibility of not citing entire papers but to refer to particular (sub)sections of a paper. Such local citations allow, for example, differentiating citations due to methodological aspects from citations due to a paper's main conclusion. A related option is to use local references that indicate the part of an article for which a reference was relevant. Classifying references into, say, essential, important, and useful references could be another simple, but important step towards better metrics.

An aspect often ignored in evaluating research is that measuring the success of research and rewarding it are two related but not necessarily identical issues. Measuring research success is necessary for developing a rational research policy. Only when we understand how much different parts of science add to human well-being, how these parts interact, how the realization of value-added evolves over time, and how risky the production of science is, will we be able to establish efficient institutions for research. Consider for example the risk in producing research. If no risk were involved in producing research, a linear reward scheme can be optimal, whereas in a purely stochastic environment fixed rewards would suffice.

Understanding and improving the science system, however, clearly goes beyond assessing research output. We also need to gain a better understanding of the determinants of research success. Ellison (2002), for example, claims that, in economics, a paper's main

contribution has become less important for publication success than other aspects of quality (such as generality, robustness checks, extensions, discussions of related literature, etc.). According to Ellison, this shift reflects changing norms in the economics profession and provides little benefit at high cost. Sunderland, Sunderland-Groves, Shanley, and Campbell (2009) point at a 'research-implementation gap' in conservation biology and suggest ways to improve the transmission of knowledge from researchers to practitioners and also a transmission of experience in the opposite direction. Shanley and López (2009) find that contemporary researchers in conservation biology face wrong incentives and propose that research institutions should "restructure institutional incentive structures to take into account actual 'impact' rather than solely 'high impact' journals". This is exactly what the VA-index does by linking actual impact to its scientific sources.

3.5 Appendix

Table 3.4 presents a comparison of journal ratings for the full set of journals.

Table 3.4: Comparison of journal ratings

Journal	Inv 3y	Inv* 24y	VA-index w	VA-index u
QUARTERLY JOURNAL OF ECONOMICS	96.37 (2)	100 (1)	100 (1)	99.75 (2)
ECONOMETRICA	100 (1)	79.96 (2)	74.18 (2)	71.15 (4)
JOURNAL OF POLITICAL ECONOMY	86.97 (3)	79.63 (3)	73.34 (3)	71.23 (3)
JOURNAL OF ECONOMIC LITERATURE	51.13 (6)	73.03 (4)	69.3 (4)	100 (1)
AMERICAN ECONOMIC REVIEW	75.23 (5)	59.46 (6)	56.56 (5)	64.35 (5)
REVIEW OF ECONOMIC STUDIES	80.71 (4)	61.6 (5)	53.95 (6)	48.32 (7)
BROOKINGS PAPERS ON ECONOMIC ACTIVITY	7.1 (56)	35.23 (7)	41.11 (7)	52.92 (6)
JOURNAL OF FINANCIAL ECONOMICS	28.74 (15)	25.88 (11)	29.61 (8)	36.58 (11)
JOURNAL OF ECONOMIC PERSPECTIVES	34.95 (11)	28.06 (9)	27.63 (9)	38.37 (9)
JOURNAL OF MONETARY ECONOMICS	25.89 (18)	24.29 (13)	26.12 (10)	33.05 (12)
JOURNAL OF ECONOMIC GROWTH	24.18 (22)	27.7 (10)	26.03 (11)	37.53 (10)
RAND JOURNAL OF ECONOMICS	40 (8)	28.57 (8)	25.5 (12)	30.35 (13)
JOURNAL OF ECONOMIC THEORY	39.11 (9)	25.1 (12)	22.51 (13)	21.31 (28)
ECONOMIC POLICY	27.26 (16)	18.94 (17)	20.69 (14)	38.89 (8)
JOURNAL OF LABOR ECONOMICS	31.2 (14)	20.83 (15)	20.24 (15)	24.91 (21)
REVIEW OF ECONOMICS AND STATISTICS	40.65 (7)	21.24 (14)	19.34 (16)	28.51 (14)
GAMES AND ECONOMIC BEHAVIOR	32.69 (12)	17.89 (18)	17.94 (17)	17.09 (39)
JOURNAL OF HUMAN RESOURCES	24.26 (21)	16.5 (21)	17.68 (18)	24.02 (22)
INTERNATIONAL ECONOMIC REVIEW	35.82 (10)	20.02 (16)	17.66 (19)	19.81 (32)
JOURNAL OF ECONOMETRICS	23.92 (23)	17.28 (20)	17.08 (20)	26.57 (16)
JOURNAL OF LAW & ECONOMICS	31.92 (13)	17.78 (19)	16.76 (21)	20.28 (31)
AER Papers & Proceedings	18.12 (30)	15.88 (24)	16.58 (22)	21.57 (25)
JOURNAL OF INTERNATIONAL ECONOMICS	26.48 (17)	15.11 (26)	15.69 (23)	26.05 (17)
REVIEW OF ECONOMIC DYNAMICS	20.86 (26)	16.2 (22)	15.42 (24)	16.05 (43)
JOURNAL OF FINANCIAL AND QUANTITATIVE ANALYSIS	25.8 (19)	16.1 (23)	14.75 (25)	17.27 (38)
JOURNAL OF ACCOUNTING & ECONOMICS	7.87 (52)	13.28 (30)	14.52 (26)	13.96 (50)
ECONOMIC JOURNAL	25.56 (20)	15.6 (25)	14.36 (27)	25.18 (19)
JOURNAL OF BUSINESS & ECONOMIC STATISTICS	22.97 (24)	14.65 (27)	13.78 (28)	21.36 (27)
JOURNAL OF LAW ECONOMICS & ORGANIZATION	8.9 (48)	13.08 (31)	13.12 (29)	17.52 (37)
JOURNAL OF PUBLIC ECONOMICS	19.33 (27)	12.64 (32)	12.2 (30)	19.66 (33)
JOURNAL OF APPLIED ECONOMETRICS	18.2 (29)	14.56 (28)	12.13 (31)	20.51 (30)
JOURNAL OF INDUSTRIAL ECONOMICS	21.46 (25)	13.65 (29)	11.73 (32)	20.58 (29)
MATHEMATICAL FINANCE	10.2 (42)	9.92 (39)	11.68 (33)	27.61 (15)
WORLD BANK ECONOMIC REVIEW	10.55 (39)	10.91 (35)	11.14 (34)	26.04 (18)
EUROPEAN ECONOMIC REVIEW	12.19 (34)	11.06 (33)	10.76 (35)	18.57 (34)
JOURNAL OF MONEY CREDIT AND BANKING	17.54 (31)	10.2 (36)	10.44 (36)	17.95 (36)
JOURNAL OF ECONOMICS & MANAGEMENT STRATEGY	11.27 (35)	11.03 (34)	10.35 (37)	14.5 (47)
ECONOMETRIC THEORY	7.46 (54)	10.1 (38)	9.38 (38)	11.68 (61)
JOURNAL OF HEALTH ECONOMICS	6.25 (65)	8.49 (43)	9.07 (39)	21.56 (26)
ECONOMIC THEORY	19.09 (28)	10.18 (37)	8.89 (40)	11.35 (67)
JOURNAL OF RISK AND UNCERTAINTY	4.83 (72)	9.25 (41)	8.88 (41)	16.56 (41)
JOURNAL OF MATHEMATICAL ECONOMICS	11.25 (36)	9.61 (40)	8.68 (42)	10.51 (78)
WORLD BANK RESEARCH OBSERVER	10.42 (41)	8.21 (44)	8.09 (43)	21.76 (24)
INTERNATIONAL JOURNAL OF GAME THEORY	10.97 (37)	8.93 (42)	7.88 (44)	9.27 (94)
JOURNAL OF URBAN ECONOMICS	9.31 (45)	7.35 (45)	7.23 (45)	15.27 (46)
JOURNAL OF ECONOMIC DYNAMICS & CONTROL	10.78 (38)	7.33 (46)	6.85 (46)	11.8 (60)
JOURNAL OF DEVELOPMENT ECONOMICS	14.12 (33)	6.95 (47)	6.84 (47)	16.13 (42)
ECONOMIC INQUIRY	10.09 (43)	6.67 (48)	6.3 (48)	11.64 (63)
SOCIAL CHOICE AND WELFARE	5.12 (69)	6.39 (51)	6.22 (49)	10.19 (82)
JOURNAL OF ECONOMIC HISTORY	8.64 (50)	6.38 (52)	6 (50)	9.73 (84)
MACROECONOMIC DYNAMICS	8.93 (47)	6.45 (50)	5.85 (51)	10.94 (76)
IMF STAFF PAPERS	3.9 (84)	5.93 (55)	5.81 (52)	13.73 (51)
JOURNAL OF ECONOMIC SURVEYS	7.51 (53)	6.59 (49)	5.81 (53)	18.15 (35)
JOURNAL OF ECONOMIC GEOGRAPHY	4.56 (75)	5.56 (60)	5.69 (54)	23.43 (23)
LABOUR ECONOMICS	6.97 (57)	6.09 (54)	5.65 (55)	13.26 (52)
JOURNAL OF ENVIRONMENTAL ECONOMICS AND MANAGEMENT	6.28 (64)	5.7 (57)	5.6 (56)	24.93 (20)
JOURNAL OF ECONOMIC BEHAVIOR & ORGANIZATION	8.73 (49)	5.52 (61)	5.32 (57)	10.79 (77)
SCANDINAVIAN JOURNAL OF ECONOMICS	9.38 (44)	5.84 (56)	5.22 (58)	11.08 (71)
INTERNATIONAL JOURNAL OF INDUSTRIAL ORGANIZATION	9.07 (46)	6.14 (53)	5.21 (59)	11.49 (64)
OXFORD ECONOMIC PAPERS-NEW SERIES	4.83 (71)	5.7 (58)	5.15 (60)	12.91 (54)
EXPLORATIONS IN ECONOMIC HISTORY	10.5 (40)	5.29 (62)	4.96 (61)	9.55 (91)
OXFORD BULLETIN OF ECONOMICS AND STATISTICS	8.01 (51)	5.61 (59)	4.9 (62)	14.2 (49)
REVIEW OF INCOME AND WEALTH	6.84 (61)	3.82 (73)	4.74 (63)	11.13 (69)
REGIONAL SCIENCE AND URBAN ECONOMICS	6.22 (66)	4.59 (64)	4.54 (64)	11.68 (62)
ENERGY JOURNAL	6.89 (59)	4.55 (65)	4.47 (65)	16.91 (40)
NATIONAL TAX JOURNAL	3.68 (85)	4.05 (71)	4.35 (66)	8.16 (103)
ECONOMICA	6.3 (63)	4.89 (63)	4.32 (67)	11.11 (70)

Journal	Inv 3y	Inv* 24y	VA-index w	VA-index u
JOURNAL OF PRODUCTIVITY ANALYSIS	3.64 (87)	4.14 (69)	4.09 (68)	13.13 (53)
JOURNAL OF POPULATION ECONOMICS	4.7 (74)	4.14 (70)	3.94 (69)	9.56 (90)
CANADIAN JOURNAL OF ECONOMICS-REVUE CANADIENNE D ECONOMIQUE	7.35 (55)	4.4 (66)	3.88 (70)	9.13 (97)
JOURNAL OF BANKING & FINANCE	2.58 (99)	3.74 (74)	3.87 (71)	9.51 (92)
OXFORD REVIEW OF ECONOMIC POLICY	3.13 (91)	3.89 (72)	3.77 (72)	10.22 (81)
JOURNAL OF THE JAPANESE AND INTERNATIONAL ECONOMIES	2.44 (103)	2.91 (83)	3.73 (73)	9.13 (96)
ECONOMICS AND PHILOSOPHY	17.04 (32)	4.24 (67)	3.65 (74)	7.94 (106)
JOURNAL OF REGULATORY ECONOMICS	3.54 (89)	4.19 (68)	3.51 (75)	9.07 (98)
PUBLIC CHOICE	4.8 (73)	3.62 (75)	3.3 (76)	7.14 (119)
ECONOMICS OF EDUCATION REVIEW	4.46 (77)	3.12 (81)	3.21 (77)	7.38 (117)
INTERNATIONAL TAX AND PUBLIC FINANCE	4.19 (82)	3.33 (78)	3.21 (78)	9.63 (86)
JOURNAL OF RISK AND INSURANCE	2.15 (109)	3.14 (80)	3.12 (79)	11.01 (73)
INTERNATIONAL REVIEW OF LAW AND ECONOMICS	1.34 (124)	2.83 (88)	3.11 (80)	6.62 (126)
JOURNAL OF COMPARATIVE ECONOMICS	5.66 (67)	3.1 (82)	3.11 (81)	10.98 (74)
RESOURCE AND ENERGY ECONOMICS	4.25 (79)	3.32 (79)	3.02 (82)	14.46 (48)
JOURNAL OF POLICY ANALYSIS AND MANAGEMENT	3.54 (90)	2.55 (93)	3.01 (83)	7.88 (108)
HEALTH ECONOMICS	3.01 (92)	2.85 (84)	3 (84)	12.66 (56)
ECONOMICS LETTERS	3.54 (88)	3.35 (77)	2.98 (85)	7.06 (122)
ECONOMIC DEVELOPMENT AND CULTURAL CHANGE	6.94 (58)	2.8 (90)	2.87 (86)	11.37 (65)
THEORY AND DECISION	3.67 (86)	3.61 (76)	2.85 (87)	5.39 (139)
JOURNAL OF REAL ESTATE FINANCE AND ECONOMICS	1.1 (135)	2.84 (85)	2.75 (88)	8.91 (100)
LAND ECONOMICS	2.25 (107)	2.83 (87)	2.66 (89)	15.34 (45)
REAL ESTATE ECONOMICS	1.36 (123)	2.81 (89)	2.66 (90)	9.6 (88)
JOURNAL OF HOUSING ECONOMICS	0.91 (141)	2.83 (86)	2.56 (91)	8.11 (104)
ECONOMICS OF TRANSITION	1.64 (119)	2.4 (94)	2.51 (92)	12.01 (59)
STUDIES IN NONLINEAR DYNAMICS AND ECONOMETRICS	2.58 (98)	2.59 (92)	2.45 (93)	8.47 (101)
SOUTHERN ECONOMIC JOURNAL	6.59 (62)	2.67 (91)	2.44 (94)	7.48 (115)
REVIEW OF WORLD ECONOMICS	4.18 (83)	1.87 (108)	2.18 (95)	7.62 (112)
JOURNAL OF INSTITUTIONAL AND THEORETICAL ECONOMICS	4.95 (70)	2.4 (95)	2.16 (96)	5.71 (134)
ECONOMIC GEOGRAPHY	2.51 (102)	2.22 (97)	2.1 (97)	12.81 (55)
JOURNAL OF EVOLUTIONARY ECONOMICS	2.25 (108)	1.97 (101)	2.08 (98)	11.04 (72)
KYKLOS	2.96 (93)	1.83 (109)	2.07 (99)	9.03 (99)
ENERGY ECONOMICS	1.22 (129)	2.27 (96)	2.04 (100)	10.06 (83)
REVIEW OF INDUSTRIAL ORGANIZATION	1.3 (128)	2.15 (98)	2.02 (101)	7.22 (118)
JOURNAL OF DEVELOPMENT STUDIES	2.87 (94)	1.9 (106)	2.01 (102)	11.37 (66)
ECONOMIC HISTORY REVIEW	5.5 (68)	2.12 (99)	2.01 (103)	5.39 (138)
JOURNAL OF TRANSPORT ECONOMICS AND POLICY	0.95 (137)	1.95 (103)	1.98 (104)	6.41 (130)
INDUSTRIAL AND CORPORATE CHANGE	1.4 (121)	1.9 (107)	1.92 (105)	10.29 (80)
WORLD ECONOMY	4.22 (81)	1.82 (110)	1.89 (106)	7.84 (109)
INSURANCE MATHEMATICS & ECONOMICS	1.56 (120)	1.47 (118)	1.89 (107)	7.61 (113)
ENVIRONMENTAL & RESOURCE ECONOMICS	2.31 (105)	2.03 (100)	1.86 (108)	12.12 (58)
WORLD DEVELOPMENT	2.12 (111)	1.79 (112)	1.78 (109)	10.98 (75)
CHINA ECONOMIC REVIEW	0.82 (145)	1.53 (116)	1.77 (110)	10.49 (79)
JOURNAL OF REGIONAL SCIENCE	2.75 (96)	1.94 (105)	1.66 (111)	7.69 (110)
INTERNATIONAL JOURNAL OF FORECASTING	1.31 (127)	1.95 (104)	1.65 (112)	6.43 (129)
AMERICAN JOURNAL OF AGRICULTURAL ECONOMICS	2.09 (112)	1.75 (114)	1.65 (113)	12.17 (57)
FISCAL STUDIES	2.25 (106)	1.79 (111)	1.65 (114)	7.44 (116)
JOURNAL OF ECONOMIC PSYCHOLOGY	1.71 (118)	1.51 (117)	1.64 (115)	5.31 (142)
INFORMATION ECONOMICS AND POLICY	4.24 (80)	1.76 (113)	1.6 (116)	6.68 (124)
JAPANESE ECONOMIC REVIEW	6.89 (60)	1.96 (102)	1.59 (117)	5.19 (146)
CONTEMPORARY ECONOMIC POLICY	4.41 (78)	1.45 (120)	1.52 (118)	7.11 (121)
REVIEW OF INTERNATIONAL POLITICAL ECONOMY	0.42 (155)	1.37 (121)	1.35 (119)	9.57 (89)
SCOTTISH JOURNAL OF POLITICAL ECONOMY	1.91 (114)	1.56 (115)	1.35 (120)	6.55 (128)
JOURNAL OF ECONOMIC EDUCATION	0.82 (146)	1.19 (127)	1.32 (121)	5.21 (145)
JOURNAL OF AFRICAN ECONOMIES	1.77 (116)	1.31 (124)	1.3 (122)	6.91 (123)
FEMINIST ECONOMICS	0.15 (162)	0.92 (138)	1.18 (123)	6.64 (125)
MANCHESTER SCHOOL	0.43 (153)	1.24 (126)	1.14 (124)	5.34 (141)
JOURNAL OF ECONOMICS	2.75 (95)	1.46 (119)	1.12 (125)	4.47 (152)
JOURNAL OF MACROECONOMICS	2.64 (97)	1.33 (122)	1.12 (126)	5.51 (136)
OPEN ECONOMIES REVIEW	4.55 (76)	1.27 (125)	1.09 (127)	4.48 (151)
ECONOMIC DEVELOPMENT QUARTERLY	0.21 (161)	1.32 (123)	1.08 (128)	4.37 (153)
CAMBRIDGE JOURNAL OF ECONOMICS	2.55 (100)	1.09 (128)	1.02 (129)	8.42 (102)
EUROPEAN REVIEW OF AGRICULTURAL ECONOMICS	0.92 (140)	1.01 (133)	0.98 (130)	11.15 (68)
AGRICULTURAL ECONOMICS	1.01 (136)	1.07 (130)	0.97 (131)	9.45 (93)
JOURNAL OF AGRICULTURAL AND RESOURCE ECONOMICS	0.48 (151)	1.05 (131)	0.94 (132)	9.19 (95)
JAPAN AND THE WORLD ECONOMY	1.85 (115)	1.02 (132)	0.93 (133)	5.14 (148)
ECONOMIC RECORD	2.54 (101)	1 (134)	0.93 (134)	5.18 (147)
ECONOMIC MODELLING	0.85 (144)	1.08 (129)	0.93 (135)	5.77 (132)
ECOLOGICAL ECONOMICS	0.93 (139)	0.92 (136)	0.92 (136)	7.69 (111)
ECONOMIST-NETHERLANDS	0.61 (147)	0.88 (139)	0.92 (137)	4.25 (156)
AUSTRALIAN JOURNAL OF AGRICULTURAL AND RESOURCE ECONOMICS	2.14 (110)	0.92 (137)	0.83 (138)	9.69 (85)
APPLIED ECONOMICS	1.38 (122)	0.95 (135)	0.82 (139)	5.99 (131)
ECONOMY AND SOCIETY	0.27 (159)	0.59 (146)	0.77 (140)	7.14 (120)

Journal	Inv 3y	Inv* 24y	VA-index w	VA-index u
SMALL BUSINESS ECONOMICS	1.32 (126)	0.81 (141)	0.77 (141)	5.36 (140)
JOURNAL OF POLICY MODELING	1.22 (130)	0.76 (142)	0.76 (142)	5.67 (135)
FOOD POLICY	0.95 (138)	0.83 (140)	0.73 (143)	8.05 (105)
JCMS-JOURNAL OF COMMON MARKET STUDIES	0.43 (154)	0.63 (145)	0.72 (144)	6.59 (127)
DEFENCE AND PEACE ECONOMICS	1.92 (113)	0.7 (143)	0.69 (145)	4.03 (158)
JOURNAL OF AGRICULTURAL ECONOMICS	0.9 (142)	0.7 (144)	0.65 (146)	7.93 (107)
AMERICAN JOURNAL OF ECONOMICS AND SOCIOLOGY	2.38 (104)	0.54 (147)	0.61 (147)	4.19 (157)
BULLETIN OF INDONESIAN ECONOMIC STUDIES	1.75 (117)	0.42 (150)	0.59 (148)	5.5 (137)
POST-SOVIET AFFAIRS	0.6 (148)	0.3 (156)	0.53 (149)	15.61 (44)
CANADIAN JOURNAL OF AGRICULTURAL ECONOMICS-REVUE CANADIENNE D AGROECONOMIE	1.16 (132)	0.49 (148)	0.4 (150)	5.72 (133)
APPLIED ECONOMICS LETTERS	1.32 (125)	0.44 (149)	0.38 (151)	4.78 (149)
VALUE IN HEALTH	0.12 (166)	0.4 (152)	0.37 (152)	3.76 (161)
HITOTSUBASHI JOURNAL OF ECONOMICS	1.13 (134)	0.41 (151)	0.35 (153)	2.91 (167)
EUROPE-ASIA STUDIES	0.47 (152)	0.24 (158)	0.34 (154)	7.5 (114)
SOUTH AFRICAN JOURNAL OF ECONOMICS	0.34 (158)	0.28 (157)	0.32 (155)	3.74 (162)
DEVELOPING ECONOMIES	1.14 (133)	0.3 (155)	0.31 (156)	4.37 (154)
JOURNAL OF POST KEYNESIAN ECONOMICS	1.19 (131)	0.35 (153)	0.3 (157)	5.29 (143)
EURASIAN GEOGRAPHY AND ECONOMICS	0.86 (143)	0.13 (164)	0.3 (158)	9.61 (87)
WORK EMPLOYMENT AND SOCIETY	0.1 (167)	0.3 (154)	0.28 (159)	3.93 (159)
POST-COMMUNIST ECONOMIES	0.13 (164)	0.23 (160)	0.24 (160)	5.25 (144)
JOURNAL OF ECONOMIC ISSUES	0.37 (157)	0.24 (159)	0.19 (161)	3.89 (160)
TIJDSCHRIFT VOOR ECONOMISCHE EN SOCIALE GEOGRAFIE	0.13 (165)	0.2 (161)	0.17 (162)	3.41 (164)
JAHRBUCHER FUR NATIONALOKONOMIE UND STATISTIK	0.56 (150)	0.19 (162)	0.17 (163)	2.86 (168)
EMERGING MARKETS FINANCE AND TRADE	0.37 (156)	0.15 (163)	0.12 (164)	3.47 (163)
FUTURES	0.06 (169)	0.12 (165)	0.11 (165)	3.4 (165)
EASTERN EUROPEAN ECONOMICS	0.26 (160)	0.09 (166)	0.1 (166)	4.33 (155)
JOURNAL OF MEDIA ECONOMICS	0.58 (149)	0.08 (167)	0.07 (167)	3.05 (166)
TRIMESTRE ECONOMICO	0.14 (163)	0.06 (168)	0.06 (168)	2.67 (170)
REVUE D ETUDES COMPARATIVES EST-OUEST	0 (171)	0.01 (170)	0.01 (169)	2.5 (171)
POLITICKA EKONOMIE	0.08 (168)	0.01 (169)	0.01 (170)	4.67 (150)
EKONOMICKY CASOPIS	0.05 (170)	0 (171)	0 (171)	2.83 (169)
EKONOMISKA SAMFUNDETS TIDSKRIFT	0 (172)	0 (172)	0 (172)	2.33 (172)

Table 3.5 presents two ratings according to the invariant method adapted to a cited period of 24 years and two ratings according to the VA-index. The first version of a rating uses nominal weights, whereas the second version uses discounted weights. The ordering of journals follows the relative loss in ranks a journal experiences when moving from nominal to discounted weights based on the invariant method.

Table 3.5: Nominal vs. discounted weighting

Journal	Inv 24 nom	Inv 24 disc	VA w nom	VA w disc
JOURNAL OF MATHEMATICAL ECONOMICS	34	40	42	43
JOURNAL OF ECONOMIC GROWTH	9	10	11	10
SOCIAL CHOICE AND WELFARE	47	51	49	51
MATHEMATICAL FINANCE	36	39	33	35
JOURNAL OF APPLIED ECONOMETRICS	26	28	31	32
INTERNATIONAL REVIEW OF LAW AND ECONOMICS	82	88	80	86
OXFORD BULLETIN OF ECONOMICS AND STATISTICS	55	59	62	62
JOURNAL OF LABOR ECONOMICS	14	15	15	15
GAMES AND ECONOMIC BEHAVIOR	17	18	17	20
ECONOMIC THEORY	35	37	40	41
JOURNAL OF ECONOMETRICS	19	20	20	22
INTERNATIONAL JOURNAL OF GAME THEORY	40	42	44	45
REVIEW OF ECONOMIC DYNAMICS	21	22	24	24
INFORMATION ECONOMICS AND POLICY	108	113	116	117
JOURNAL OF REGULATORY ECONOMICS	65	68	75	75
JOURNAL OF FINANCIAL AND QUANTITATIVE ANALYSIS	22	23	25	26
THEORY AND DECISION	73	76	87	87
REAL ESTATE ECONOMICS	86	89	90	90
JOURNAL OF ACCOUNTING & ECONOMICS	29	30	26	27
JOURNAL OF ECONOMIC BEHAVIOR & ORGANIZATION	59	61	57	60
JOURNAL OF LAW ECONOMICS & ORGANIZATION	30	31	29	29
JOURNAL OF EVOLUTIONARY ECONOMICS	98	101	98	105
JOURNAL OF ECONOMICS & MANAGEMENT STRATEGY	33	34	37	37
JOURNAL OF PRODUCTIVITY ANALYSIS	67	69	68	71
INDUSTRIAL AND CORPORATE CHANGE	104	107	105	106
JOURNAL OF BANKING & FINANCE	72	74	71	72
JOURNAL OF HOUSING ECONOMICS	84	86	91	93
JOURNAL OF ECONOMIC DYNAMICS & CONTROL	45	46	46	47
STUDIES IN NONLINEAR DYNAMICS AND ECONOMETRICS	90	92	93	94
REVIEW OF INDUSTRIAL ORGANIZATION	96	98	101	104
JOURNAL OF ECONOMIC SURVEYS	48	49	53	53
MACROECONOMIC DYNAMICS	49	50	51	52

Journal	Inv 24 nom	Inv 24 disc	VA w nom	VA w disc
HITOTSUBASHI JOURNAL OF ECONOMICS	148	151	153	156
KYKLOS	107	109	99	101
FISCAL STUDIES	109	111	114	113
AMERICAN JOURNAL OF AGRICULTURAL ECONOMICS	112	114	113	112
INSURANCE MATHEMATICS & ECONOMICS	116	118	107	109
JOURNAL OF ECONOMICS	117	119	125	127
JOURNAL OF AFRICAN ECONOMIES	122	124	122	121
ECONOMICA	62	63	67	67
JOURNAL OF AGRICULTURAL AND RESOURCE ECONOMICS	129	131	132	134
JOURNAL OF INSTITUTIONAL AND THEORETICAL ECONOMICS	94	95	96	98
INTERNATIONAL JOURNAL OF FORECASTING	103	104	112	114
CHINA ECONOMIC REVIEW	115	116	110	110
FEMINIST ECONOMICS	137	138	123	124
SMALL BUSINESS ECONOMICS	140	141	141	144
JOURNAL OF AGRICULTURAL ECONOMICS	143	144	146	146
VALUE IN HEALTH	151	152	152	153
JOURNAL OF ECONOMIC ISSUES	158	159	161	162
POST-COMMUNIST ECONOMIES	159	160	160	160
JAHRBUCHER FUR NATIONALOKONOMIE UND STATISTIK	161	162	163	163
ECONOMETRICA	2	2	2	3
JOURNAL OF FINANCIAL ECONOMICS	11	11	8	9
JOURNAL OF MONETARY ECONOMICS	13	13	10	11
PUBLIC CHOICE	75	75	76	79
INTERNATIONAL JOURNAL OF INDUSTRIAL ORGANIZATION	53	53	59	61
JAPAN AND THE WORLD ECONOMY	132	132	133	137
ECONOMETRIC THEORY	38	38	38	39
JOURNAL OF RISK AND INSURANCE	80	80	79	81
REGIONAL SCIENCE AND URBAN ECONOMICS	64	64	64	65
JOURNAL OF THE JAPANESE AND INTERNATIONAL ECONOMIES	83	83	73	74
REVIEW OF INTERNATIONAL POLITICAL ECONOMY	121	121	119	120
JOURNAL OF ECONOMIC EDUCATION	127	127	121	122
MANCHESTER SCHOOL	126	126	124	125
JCMS-JOURNAL OF COMMON MARKET STUDIES	145	145	144	145
DEVELOPING ECONOMIES	155	155	156	157
JOURNAL OF POST KEYNESIAN ECONOMICS	153	153	157	158
AMERICAN ECONOMIC REVIEW	6	6	5	5
AMERICAN JOURNAL OF ECONOMICS AND SOCIOLOGY	147	147	147	147
APPLIED ECONOMICS	135	135	139	139
BROOKINGS PAPERS ON ECONOMIC ACTIVITY	7	7	7	7
CAMBRIDGE JOURNAL OF ECONOMICS	128	128	129	129
CONTEMPORARY ECONOMIC POLICY	120	120	118	118
EASTERN EUROPEAN ECONOMICS	166	166	166	166
ECOLOGICAL ECONOMICS	136	136	136	136
ECONOMIC DEVELOPMENT QUARTERLY	123	123	128	128
ECONOMICS OF EDUCATION REVIEW	81	81	77	77
ECONOMY AND SOCIETY	146	146	140	140
EKONOMICKY CASOPIS	171	171	171	171
EKONOMISKA SAMFUNDETS TIDSKRIFT	172	172	172	172
EMERGING MARKETS FINANCE AND TRADE	163	163	164	164
ENVIRONMENTAL & RESOURCE ECONOMICS	100	100	108	108
EUROPEAN REVIEW OF AGRICULTURAL ECONOMICS	133	133	130	130
FUTURES	165	165	165	165
INTERNATIONAL TAX AND PUBLIC FINANCE	78	78	78	78
JOURNAL OF BUSINESS & ECONOMIC STATISTICS	27	27	28	28
JOURNAL OF ECONOMIC GEOGRAPHY	60	60	54	54
JOURNAL OF ECONOMIC LITERATURE	4	4	4	4
JOURNAL OF ECONOMIC THEORY	12	12	13	13
JOURNAL OF ENVIRONMENTAL ECONOMICS AND MANAGEMENT	57	57	56	56
JOURNAL OF MEDIA ECONOMICS	167	167	167	167
JOURNAL OF POLICY MODELING	142	142	142	142
JOURNAL OF POPULATION ECONOMICS	70	70	69	69
JOURNAL OF PUBLIC ECONOMICS	32	32	30	30
JOURNAL OF REAL ESTATE FINANCE AND ECONOMICS	85	85	88	88
JOURNAL OF REGIONAL SCIENCE	105	105	111	111
LABOUR ECONOMICS	54	54	55	55
NATIONAL TAX JOURNAL	71	71	66	66
POLITICKA EKONOMIE	169	169	170	170
POST-SOVIET AFFAIRS	156	156	149	149
QUARTERLY JOURNAL OF ECONOMICS	1	1	1	1
RAND JOURNAL OF ECONOMICS	8	8	12	12
REVIEW OF ECONOMIC STUDIES	5	5	6	6
REVUE D ETUDES COMPARATIVES EST-OUEST	170	170	169	169
TRIMESTRE ECONOMICO	168	168	168	168

Journal	Inv 24 nom	Inv 24 disc	VA w nom	VA w disc
WORK EMPLOYMENT AND SOCIETY	154	154	159	159
SOUTH AFRICAN JOURNAL OF ECONOMICS	157	157	155	154
ECONOMIC RECORD	134	134	134	133
OPEN ECONOMIES REVIEW	125	125	127	126
JAPANESE ECONOMIC REVIEW	102	102	117	116
ECONOMICS LETTERS	77	77	85	84
JOURNAL OF POLICY ANALYSIS AND MANAGEMENT	93	93	83	82
ECONOMIST-NETHERLANDS	139	139	137	135
EURASIAN GEOGRAPHY AND ECONOMICS	164	164	158	155
JOURNAL OF ECONOMIC HISTORY	52	52	50	49
WORLD BANK RESEARCH OBSERVER	44	44	43	42
RESOURCE AND ENERGY ECONOMICS	79	79	82	80
JOURNAL OF HEALTH ECONOMICS	43	43	39	38
AER Papers & Proceedings	24	24	22	21
ECONOMIC JOURNAL	25	25	27	25
INTERNATIONAL ECONOMIC REVIEW	16	16	19	17
JOURNAL OF POLITICAL ECONOMY	3	3	3	2
TIJDSCHRIFT VOOR ECONOMISCHE EN SOCIALE GEOGRAFIE	162	161	162	161
APPLIED ECONOMICS LETTERS	150	149	151	151
CANADIAN JOURNAL OF AGRICULTURAL ECONOMICS-REVUE CANADIENNE D AGROECONOMIE	149	148	150	150
DEFENCE AND PEACE ECONOMICS	144	143	145	143
FOOD POLICY	141	140	143	141
AUSTRALIAN JOURNAL OF AGRICULTURAL AND RESOURCE ECONOMICS	138	137	138	138
AGRICULTURAL ECONOMICS	131	130	131	131
ECONOMIC MODELLING	130	129	135	132
JOURNAL OF ECONOMIC PSYCHOLOGY	118	117	115	115
ENERGY ECONOMICS	97	96	100	99
ECONOMICS OF TRANSITION	95	94	92	92
SOUTHERN ECONOMIC JOURNAL	92	91	94	91
ECONOMIC DEVELOPMENT AND CULTURAL CHANGE	91	90	86	83
LAND ECONOMICS	88	87	89	89
EUROPE-ASIA STUDIES	160	158	154	152
BULLETIN OF INDONESIAN ECONOMIC STUDIES	152	150	148	148
ENERGY JOURNAL	66	65	65	64
EXPLORATIONS IN ECONOMIC HISTORY	63	62	61	58
JOURNAL OF MACROECONOMICS	124	122	126	123
WORLD DEVELOPMENT	114	112	109	107
IMF STAFF PAPERS	56	55	52	50
ECONOMIC HISTORY REVIEW	101	99	103	97
ECONOMIC GEOGRAPHY	99	97	97	96
JOURNAL OF URBAN ECONOMICS	46	45	45	44
JOURNAL OF RISK AND UNCERTAINTY	42	41	41	40
WORLD ECONOMY	113	110	106	103
REVIEW OF WORLD ECONOMICS	111	108	95	95
OXFORD REVIEW OF ECONOMIC POLICY	74	72	72	70
JOURNAL OF TRANSPORT ECONOMICS AND POLICY	106	103	104	102
ECONOMICS AND PHILOSOPHY	69	67	74	73
CANADIAN JOURNAL OF ECONOMICS-REVUE CANADIENNE D ECONOMIQUE	68	66	70	68
SCOTTISH JOURNAL OF POLITICAL ECONOMY	119	115	120	119
HEALTH ECONOMICS	87	84	84	85
SCANDINAVIAN JOURNAL OF ECONOMICS	58	56	58	57
JOURNAL OF DEVELOPMENT STUDIES	110	106	102	100
REVIEW OF INCOME AND WEALTH	76	73	63	63
ECONOMIC INQUIRY	50	48	48	48
OXFORD ECONOMIC PAPERS-NEW SERIES	61	58	60	59
JOURNAL OF LAW & ECONOMICS	20	19	21	19
ECONOMIC POLICY	18	17	14	14
JOURNAL OF INDUSTRIAL ECONOMICS	31	29	32	31
REVIEW OF ECONOMICS AND STATISTICS	15	14	16	16
JOURNAL OF INTERNATIONAL ECONOMICS	28	26	23	23
JOURNAL OF DEVELOPMENT ECONOMICS	51	47	47	46
JOURNAL OF COMPARATIVE ECONOMICS	89	82	81	76
JOURNAL OF HUMAN RESOURCES	23	21	18	18
JOURNAL OF ECONOMIC PERSPECTIVES	10	9	9	8
WORLD BANK ECONOMIC REVIEW	39	35	34	33
EUROPEAN ECONOMIC REVIEW	37	33	35	34
JOURNAL OF MONEY CREDIT AND BANKING	41	36	36	36

Table 3.6 illustrates the methodological influence of the VA-index compared to the Invariant method. Table 3.5 is reorganized such that the nominal and discounted versions of Inv* 24y and VA-index w can be compared to each other more easily.

Table 3.6: Invariant method vs. VA-index

Journal	Inv 24 nom	VA w nom	Inv 24 disc	VA w disc
RAND JOURNAL OF ECONOMICS	8	12	8	12
JOURNAL OF MATHEMATICAL ECONOMICS	34	42	40	43
JOURNAL OF ECONOMIC GROWTH	9	11	10	10
REVIEW OF ECONOMIC STUDIES	5	6	5	6
JOURNAL OF APPLIED ECONOMETRICS	26	31	28	32
THEORY AND DECISION	73	87	76	87
INTERNATIONAL ECONOMIC REVIEW	16	19	16	17
JOURNAL OF REGULATORY ECONOMICS	65	75	68	75

Journal	Inv 24 nom	VA w nom	Inv 24 disc	VA w disc
JAPANESE ECONOMIC REVIEW	102	117	102	116
ECONOMIC THEORY	35	40	37	41
REVIEW OF ECONOMIC DYNAMICS	21	24	22	24
JOURNAL OF FINANCIAL AND QUANTITATIVE ANALYSIS	22	25	23	26
OXFORD BULLETIN OF ECONOMICS AND STATISTICS	55	62	59	62
JOURNAL OF ECONOMICS & MANAGEMENT STRATEGY	33	37	34	37
INTERNATIONAL JOURNAL OF INDUSTRIAL ORGANIZATION	53	59	53	61
JOURNAL OF ECONOMIC SURVEYS	48	53	49	53
ECONOMICS LETTERS	77	85	77	84
INTERNATIONAL JOURNAL OF GAME THEORY	40	44	42	45
INTERNATIONAL JOURNAL OF FORECASTING	103	112	104	114
JOURNAL OF ECONOMIC THEORY	12	13	12	13
JOURNAL OF HOUSING ECONOMICS	84	91	86	93
ECONOMICA	62	67	63	67
ECONOMIC JOURNAL	25	27	25	25
ENVIRONMENTAL & RESOURCE ECONOMICS	100	108	100	108
INFORMATION ECONOMICS AND POLICY	108	116	113	117
ECONOMICS AND PHILOSOPHY	69	74	67	73
JOURNAL OF LABOR ECONOMICS	14	15	15	15
JOURNAL OF ECONOMICS	117	125	119	127
REVIEW OF ECONOMICS AND STATISTICS	15	16	14	16
JOURNAL OF REGIONAL SCIENCE	105	111	105	111
JOURNAL OF ECONOMETRICS	19	20	20	22
REVIEW OF INDUSTRIAL ORGANIZATION	96	101	98	104
JOURNAL OF LAW & ECONOMICS	20	21	19	19
REAL ESTATE ECONOMICS	86	90	89	90
FISCAL STUDIES	109	114	111	113
SOCIAL CHOICE AND WELFARE	47	49	51	51
MACROECONOMIC DYNAMICS	49	51	50	52
ECONOMIC DEVELOPMENT QUARTERLY	123	128	123	128
ECONOMIC MODELLING	130	135	129	132
RESOURCE AND ENERGY ECONOMICS	79	82	79	80
JOURNAL OF BUSINESS & ECONOMIC STATISTICS	27	28	27	28
JOURNAL OF REAL ESTATE FINANCE AND ECONOMICS	85	88	85	88
HITOTSUBASHI JOURNAL OF ECONOMICS	148	153	151	156
STUDIES IN NONLINEAR DYNAMICS AND ECONOMETRICS	90	93	92	94
WORK EMPLOYMENT AND SOCIETY	154	159	154	159
JOURNAL OF INDUSTRIAL ECONOMICS	31	32	29	31
ENERGY ECONOMICS	97	100	96	99
APPLIED ECONOMICS	135	139	135	139
CANADIAN JOURNAL OF ECONOMICS-REVUE CANADIENNE D ECONOMIQUE	68	70	66	68
JOURNAL OF POST KEYNESIAN ECONOMICS	153	157	153	158
JOURNAL OF AGRICULTURAL AND RESOURCE ECONOMICS	129	132	131	134
JOURNAL OF ECONOMIC DYNAMICS & CONTROL	45	46	46	47
SOUTHERN ECONOMIC JOURNAL	92	94	91	91
JOURNAL OF INSTITUTIONAL AND THEORETICAL ECONOMICS	94	96	95	98
JOURNAL OF AGRICULTURAL ECONOMICS	143	146	144	146
ECONOMIC HISTORY REVIEW	101	103	99	97
JOURNAL OF ECONOMIC ISSUES	158	161	159	162
LABOUR ECONOMICS	54	55	54	55
JOURNAL OF MACROECONOMICS	124	126	122	123
OPEN ECONOMIES REVIEW	125	127	125	126
JOURNAL OF PRODUCTIVITY ANALYSIS	67	68	69	71
FOOD POLICY	141	143	140	141
PUBLIC CHOICE	75	76	75	79
JAHRBUCHER FUR NATIONALOKONOMIE UND STATISTIK	161	163	162	163
LAND ECONOMICS	88	89	87	89
INDUSTRIAL AND CORPORATE CHANGE	104	105	107	106
AMERICAN JOURNAL OF AGRICULTURAL ECONOMICS	112	113	114	112
SCOTTISH JOURNAL OF POLITICAL ECONOMY	119	120	115	119
CAMBRIDGE JOURNAL OF ECONOMICS	128	129	128	129
JAPAN AND THE WORLD ECONOMY	132	133	132	137
SMALL BUSINESS ECONOMICS	140	141	141	144
DEFENCE AND PEACE ECONOMICS	144	145	143	143
CANADIAN JOURNAL OF AGRICULTURAL ECONOMICS-REVUE CANADIENNE D AGROECONOMIE	149	150	148	150
APPLIED ECONOMICS LETTERS	150	151	149	151
VALUE IN HEALTH	151	152	152	153
DEVELOPING ECONOMIES	155	156	155	157
POST-COMMUNIST ECONOMIES	159	160	160	160
EMERGING MARKETS FINANCE AND TRADE	163	164	163	164
POLITICKA EKONOMIE	169	170	169	170
AGRICULTURAL ECONOMICS	131	131	130	131
AMERICAN JOURNAL OF ECONOMICS AND SOCIOLOGY	147	147	147	147

Journal	Inv 24 nom	VA w nom	Inv 24 disc	VA w disc
AUSTRALIAN JOURNAL OF AGRICULTURAL AND RESOURCE ECONOMICS	138	138	137	138
BROOKINGS PAPERS ON ECONOMIC ACTIVITY	7	7	7	7
EASTERN EUROPEAN ECONOMICS	166	166	166	166
ECOLOGICAL ECONOMICS	136	136	136	136
ECONOMETRIC THEORY	38	38	38	39
ECONOMETRICA	2	2	2	3
ECONOMIC RECORD	134	134	134	133
EKONOMICKY CASOPIS	171	171	171	171
EKONOMISKA SAMFUNDETS TIDSKRIFT	172	172	172	172
FUTURES	165	165	165	165
GAMES AND ECONOMIC BEHAVIOR	17	17	18	20
INTERNATIONAL TAX AND PUBLIC FINANCE	78	78	78	78
JOURNAL OF AFRICAN ECONOMIES	122	122	124	121
JOURNAL OF ECONOMIC LITERATURE	4	4	4	4
JOURNAL OF EVOLUTIONARY ECONOMICS	98	98	101	105
JOURNAL OF MEDIA ECONOMICS	167	167	167	167
JOURNAL OF POLICY MODELING	142	142	142	142
JOURNAL OF POLITICAL ECONOMY	3	3	3	2
QUARTERLY JOURNAL OF ECONOMICS	1	1	1	1
REGIONAL SCIENCE AND URBAN ECONOMICS	64	64	64	65
SCANDINAVIAN JOURNAL OF ECONOMICS	58	58	56	57
TIJDSCHRIFT VOOR ECONOMISCHE EN SOCIALE GEOGRAFIE	162	162	161	161
TRIMESTRE ECONOMICO	168	168	168	168
REVUE D ETUDES COMPARATIVES EST-OUEST	170	169	170	169
JCMS-JOURNAL OF COMMON MARKET STUDIES	145	144	145	145
JOURNAL OF RISK AND INSURANCE	80	79	80	81
SOUTH AFRICAN JOURNAL OF ECONOMICS	157	155	157	154
JOURNAL OF BANKING & FINANCE	72	71	74	72
JOURNAL OF POPULATION ECONOMICS	70	69	70	69
ECONOMIST-NETHERLANDS	139	137	139	135
ENERGY JOURNAL	66	65	65	64
MANCHESTER SCHOOL	126	124	126	125
OXFORD ECONOMIC PAPERS-NEW SERIES	61	60	58	59
REVIEW OF INTERNATIONAL POLITICAL ECONOMY	121	119	121	120
CONTEMPORARY ECONOMIC POLICY	120	118	120	118
JOURNAL OF ENVIRONMENTAL ECONOMICS AND MANAGEMENT	57	56	57	56
JOURNAL OF TRANSPORT ECONOMICS AND POLICY	106	104	103	102
ECONOMIC GEOGRAPHY	99	97	97	96
JOURNAL OF URBAN ECONOMICS	46	45	45	44
EUROPEAN REVIEW OF AGRICULTURAL ECONOMICS	133	130	133	130
WORLD BANK RESEARCH OBSERVER	44	43	44	42
JOURNAL OF RISK AND UNCERTAINTY	42	41	41	40
INTERNATIONAL REVIEW OF LAW AND ECONOMICS	82	80	88	86
JOURNAL OF ECONOMIC PSYCHOLOGY	118	115	117	115
BULLETIN OF INDONESIAN ECONOMIC STUDIES	152	148	150	148
OXFORD REVIEW OF ECONOMIC POLICY	74	72	72	70
ECONOMICS OF TRANSITION	95	92	94	92
EXPLORATIONS IN ECONOMIC HISTORY	63	61	62	58
JOURNAL OF LAW ECONOMICS & ORGANIZATION	30	29	31	29
JOURNAL OF ECONOMIC BEHAVIOR & ORGANIZATION	59	57	61	60
HEALTH ECONOMICS	87	84	84	85
EURASIAN GEOGRAPHY AND ECONOMICS	164	158	164	155
EUROPE-ASIA STUDIES	160	154	158	152
JOURNAL OF ECONOMIC HISTORY	52	50	52	49
ECONOMIC INQUIRY	50	48	48	48
ECONOMY AND SOCIETY	146	140	146	140
CHINA ECONOMIC REVIEW	115	110	116	110
WORLD DEVELOPMENT	114	109	112	107
POST-SOVIET AFFAIRS	156	149	156	149
JOURNAL OF ECONOMIC EDUCATION	127	121	127	122
ECONOMICS OF EDUCATION REVIEW	81	77	81	77
EUROPEAN ECONOMIC REVIEW	37	35	33	34
ECONOMIC DEVELOPMENT AND CULTURAL CHANGE	91	86	90	83
WORLD ECONOMY	113	106	110	103
JOURNAL OF PUBLIC ECONOMICS	32	30	32	30
NATIONAL TAX JOURNAL	71	66	71	66
IMF STAFF PAPERS	56	52	55	50
JOURNAL OF DEVELOPMENT STUDIES	110	102	106	100
KYKLOS	107	99	109	101
INSURANCE MATHEMATICS & ECONOMICS	116	107	118	109
JOURNAL OF DEVELOPMENT ECONOMICS	51	47	47	46
AER Papers & Proceedings	24	22	24	21
MATHEMATICAL FINANCE	36	33	39	35
JOURNAL OF COMPARATIVE ECONOMICS	89	81	82	76
JOURNAL OF HEALTH ECONOMICS	43	39	43	38

Journal	Inv 24 nom	VA w nom	Inv 24 disc	VA w disc
JOURNAL OF ECONOMIC GEOGRAPHY	60	54	60	54
JOURNAL OF ECONOMIC PERSPECTIVES	10	9	9	8
FEMINIST ECONOMICS	137	123	138	124
JOURNAL OF ACCOUNTING & ECONOMICS	29	26	30	27
JOURNAL OF POLICY ANALYSIS AND MANAGEMENT	93	83	93	82
JOURNAL OF THE JAPANESE AND INTERNATIONAL ECONOMIES	83	73	83	74
JOURNAL OF MONEY CREDIT AND BANKING	41	36	36	36
WORLD BANK ECONOMIC REVIEW	39	34	35	33
REVIEW OF WORLD ECONOMICS	111	95	108	95
AMERICAN ECONOMIC REVIEW	6	5	6	5
REVIEW OF INCOME AND WEALTH	76	63	73	63
JOURNAL OF INTERNATIONAL ECONOMICS	28	23	26	23
JOURNAL OF HUMAN RESOURCES	23	18	21	18
ECONOMIC POLICY	18	14	17	14
JOURNAL OF MONETARY ECONOMICS	13	10	13	11
JOURNAL OF FINANCIAL ECONOMICS	11	8	11	9

Bibliography

Campbell, P. (2008): "Escape from the impact factor," *Ethics in Science and Environmental Politics*, 8, 5–7.

Ellison, G. (2002): "Evolving Standards for Academic Publishing: A q-r Theory," *Journal of Political Economy*, 110(5), 994–1034.

Garfield, E. (1972): "Citation analysis as a tool in journal evaluation," *Science*, 178(4060), 471–479.

Glasser, O. (1995): *Wilhelm Conrad Röntgen und die Geschichte der Röntgenstrahlen*. Springer, Berlin, 3rd edn.

Kalaitzidakis, P., T. P. Mamuneas, and T. Stengos (2003): "Rankings of Academic Journals and Institutions in Economics," *Journal of the European Economic Association*, 1(6), 1346–1366.

Kóczy, L. A., and M. Strobel (2010): "The World Cup of Economics Journals: A Ranking by a Tournament Method," Working Paper Series 1011, Óbuda University, Keleti Faculty of Economics.

Liebowitz, S. J., and J. P. Palmer (1984): "Assessing the Relative Impacts of Economic Journals," *Journal of Economic Literature*, 22(1), 77–88.

Nash, J. F. (1950): "Equilibrium Points in N-Person Games," *Proceedings of the National Academy of Sciences*, 36, 48–49.

Neary, J. P., J. A. Mirrlees, and J. Tirole (2003): "Evaluating Economics Research in Europe: An Introduction," *Journal of the European Economic Association*, 1(6), 1239–1249.

Nederhof, A. J., and R. F. Meijer (1995): "Development of bibliometric indicators for utility of research to users in society: Measurement of external knowledge transfer via publications in trade journals," *Scientometrics*, 32(1), 37–48.

Oswald, A. J. (2007): "An Examination of the Reliability of Prestigious Scholarly Journals: Evidence and Implications for Decision-Makers," *Economica*, 74(293), 21–31.

Palacios-Huerta, I., and O. Volij (2004): "The Measurement of Intellectual Influence," *Econometrica*, 72(3), 963–977.

Pinski, G., and F. Narin (1976): "Citation influence for journal aggregates of scientific publications: Theory, with application to the literature of physics," *Information Processing & Management*, 12(5), 297 – 312.

Ravallion, M., and A. Wagstaff (2010): "On measuring scientific influence," Policy Research Working Paper Series 5375, The World Bank.

Ritzberger, K. (2008): "A Ranking of Journals in Economics and Related Fields," *German Economic Review*, 9, 402–430.

Romer, P. M. (1990): "Endogenous Technological Change," *Journal of Political Economy*, 98(5), S71–102.

Serrano, R. (2004): "The Measurement of Intellectual Influence: the Views of a Sceptic," *Economics Bulletin*, 1(3), 1–6.

Shanley, P., and C. López (2009): "Out of the Loop: Why Research Rarely Reaches Policy Makers and the Public and What Can be Done," *Biotropica*, 41, 535–544.

Stigler, G. J., and C. Friedland (1975): "The Citation Practices of Doctorates in Economics," *Journal of Political Economy*, 83(3), 477–507.

Sunderland, T., J. Sunderland-Groves, P. Shanley, and B. Campbell (2009): "Bridging the Gap: How Can Information Access and Exchange Between Conservation Biologists and Field Practitioners be Improved for Better Conservation Outcomes?," *Biotropica*, 41, 549–554.

Ursprung, H. W., and M. Zimmer (2007): "Who is the "Platz-Hirsch" of the German Economics Profession? A Citation Analysis," *Journal of Economics and Statistics (Jahrbuecher fuer Nationaloekonomie und Statistik)*, 227(2), 187–208.

Chapter 4

How Do Editors Select Papers, and How Good are They at Doing It?

4.1 Introduction

The objective of this paper is to shed light on the editorial process of publishing journal articles in economics. To investigate the determinants of editorial decision making we use data on the B.E. Journals that rank articles into four quality tiers. In particular, we analyze whether the editorial process results in more frequently cited articles appearing in higher tiers.

Many studies have shown that professional success in academia largely depends on the number and quality of journal articles published (see e.g. Hamermesh and Pfann, 2011). The quality of a journal article can either be measured in terms of the number of received citations or in terms of the quality of the journal in which it appeared. Because journal weights are observable already at the time when an article is accepted for publication, whereas citations slowly accumulate over time, tenure and rank committees tend to rely on the signal provided by the quality of the journal. The rationale underlying the use of journal quality as a signal for article quality is that editorial and peer review guide articles to journals of appropriate quality. If editorial and peer review serve their purpose, journal quality adequately reflects article quality and research evaluation based on journal quality weights is justified.

A number of studies have investigated the peer reviewing process. Hamermesh (1994) finds that referees are generally of higher quality than the authors whose work they evaluate and they are often among the best scholars in their fields. Higher-ranked journals make use of referees who are better researchers in terms of citations and therefore arguably provide better advice to the editors and authors. Editors do, however, not assign lower-quality referees to less experienced or junior authors. According to Laband and Piette (1994), editors do not use their discretion to favor friends; they rather use it to attract good papers. Wilson (1978) investigates the success of manuscripts submitted to the *Journal of Clinical Investigation*. He finds that, on average, articles published by this journal receive significantly more citations than articles that were rejected and subsequently published elsewhere. Opthof, Furstner, van Geer, and Coronel (2000) find that papers accepted by *Cardiovascular Research* attract more citations than rejected ones even if a rejected paper later appears in a journal with a higher impact factor.

On the other hand, peer review is often perceived as intransparent and unfair. Geographic location, among others, may be a possible source of editorial bias. In the context

of the election of Fellows of the *Econometric Society*, Hamermesh and Schmidt (2003) find that, controlling for various measures of academic achievement, scholars from North America are more likely to be elected than others. Focusing on empirical studies, De Long and Lang (1992) suggest that manuscripts with statistically significant results are more likely to be accepted if the underlying hypotheses are counterintuitive and if their findings are likely due to statistical error. The *American Economic Review*'s choice of the best 20 articles published since its creation in 1911 (see Arrow, Bernheim, Feldstein, McFadden, Poterba, and Solow, 2011) illustrates the difficulties of identifying truly outstanding research. Interestingly, no article published later than in 1981 was included in the *American Economic Review*'s top 20 list. When asked why, Douglas Bernheim explained that each of the members of the committee had suggested "at least a couple of more recent papers [...] But as we move from older to younger papers, assessments vary more from person to person."[1] Just as the assessments made by *American Economic Review*'s committee, decisions reached by referees and editors often are not unanimous. Bornmann and Daniel (2008), for instance, find a surprisingly low level of agreement among referees' recommendations. Coupé (2010), finally, shows that papers that were awarded best article prizes rarely become the most cited articles published in the respective journal, although, in most cases, they get cited more often than the median paper.

Some editorial decisions may, therefore, turn out to have been wrong. Laband and Tollison (2003) find that not everything that gets published is valuable. 26 percent of the articles published in SSCI indexed journals do not get cited at all in the five years after publication. Similarly, Gans and Shepherd (1994) present a collection of famous articles that were first rejected, thereby demonstrating that excellence can be 'overlooked'. Both studies suggest that it is difficult to rate the scientific quality of a study immediately after its completion. Time-testing is important and the citation flows reflect this appreciation across time. However, citation flows are not available at early career stages when rank and tenure decisions need to be taken. Moreover, citations are only an imperfect indicator of scientific influence. Arrow, Bernheim, Feldstein, McFadden, Poterba, and Solow (2011), for example, have used citations as a starting point for their analysis, but have not entirely relied on the number of citations in selecting their top articles. Ursprung and Zimmer (2007) point out three shortcomings of citation-based evaluations. First, citations are

[1] Bernheim's remarks can be found in *The Economist*, 2011, vol. 398, iss. 8722, p. 72.

sensitive to the *halo effect*, i.e. some articles are cited because they have been cited before and not because of their scientific contributions. Second, too much weight is attributed to research that serves as an input for further research, and too little to final results. Third, they show that citation habits differ substantially across fields. This effect can not only be observed by comparing entire disciplines such as physics and philosophy but also, as the authors demonstrate using JEL-codes, when comparing subfields of economics. One may add self-citations, negative citations and citation cartels to the list of caveats.

We investigate the trade-off between timely information, i.e. which journal or, in our case, which tier an article appears in, and the quality of the information contained in citation flows, by analyzing the process of editing three journals published by *Berkeley Electronic Press*. We check how the editors' decisions reflect information that is available to them and how this information correlates with subsequent citation flows. We also draw conclusions on the validity of citation counts as an indicator of research quality. Section 4.2 describes the particular set-up used in our analysis. Section 4.3 presents the data and descriptive statistics, and Section 4.4 the econometric results. Section 4.5 concludes.

4.2 Editorial processes: The case of the BE journals

Berkeley Electronic Press publishes, among others, three economics journals, *The B.E. Journal of Economic Analysis & Policy (BEJEAP)*, *The B.E. Journal of Macroeconomics (BEJM)*, and *The B.E. Journal of Theoretical Economics (BEJTE)*. All three journals are divided into four quality-rated tiers. The editors ask the referees which tier they consider to be appropriate, but the ultimate decision is up to the editors. In the case of *The B.E. Journal of Macroeconomics* the four tiers are defined as follows: Publications which appear in the *Frontiers* are supposed to be suited for the top general interest journals in economics such as the *American Economic Review*, *Econometrica*, the *Journal of Political Economy*, or the *Quarterly Journal of Economics*. Less than 1 percent of all submissions are published in the *Frontiers*. Studies published in the *Advances* are supposed to be comparable in quality to a top field journal in macroeconomics such as the *Journal of Monetary Economics*. The publication rate in *Advances* is 6 percent. Publications in the *Contributions* tier are supposed to be suited for publication in the *European Economic Review* or the *Journal of Money, Credit, and Banking*, the publication rate is 16 percent. Publications in the *Topics* tier are, finally, supposed to be suited for

publication in *Economic Inquiry*, the publication rate is 22 percent. While articles in *Frontiers* should be of interest to anyone working in the field of macroeconomics, as one moves down the ladder, the articles are aimed at more narrow subcommunities.[2]

The publication strategy of *Berkeley Electronic Press* provides us with a unique setting: we can interpret quality tiers as separate journals. If an article is published in the *Contributions*, we interpret it as first having been rejected by the *Frontiers* and thereafter by the *Advances*. We argue that this is not at odds with the fact that higher and lower tiers are not only supposed to reflect differences in quality but also the difference between general interest journals and field journals. To be sure, some articles get published in field journals simply because they are of interest to a smaller set of scientists and not because they are of lower quality. However, especially for young scholars, it is generally preferable if their papers appear in top general interest journals rather than top field journals because top general interest journals have higher impact factors and are given more weight by tenure committees. We use citation data to investigate how well reviewers and editors fare at assigning articles to tiers and argue that these findings also apply to conventional journals.

Analyzing journals subdivided into quality tiers has several advantages over the traditional method of tracing rejected manuscripts: (1) there is only one refereeing process, so rejected papers are not "contaminated" by second opinions. (2) "Rejected" manuscripts do not suffer from an additional publication lag, but appear at the same time as higher ranked articles. Finally, (3) all articles are available to the same audience. We admit, however, that readers can use the quality ratings to filter their reading matter, which is liable to influence the forthcoming citation flow.

4.3 Data and descriptive statistics

In November 2010, we retrieved from the *IDEAS* data base, which is part of *RePEc*, publication and citation data of all articles that appeared in the *BEJEAP*, *BEJM*, and *BEJTE* in the years 2001 to 2006.[3] This provides us with 572 observations. Author ratings, ratings of working paper series and journals, and institutional ratings were collected from the same source. To this data set we added the authors' gender, the geographic

[2] See the journal website http://www.bepress.com/bejm/ratingsystem.html.

[3] See http://ideas.repec.org/. 2001 is the year in which the *B.E. Journals* were launched. *IDEAS* lists tiers only until 2006.

region of the authors' affiliations, and a dummy which indicates whether these affiliations are non-university institutions.

Table 4.1 presents the descriptive statistics of the publication variables. On average, publications in our sample were authored by 1.7 scholars. 14.95 percent of the authors of the average article are female. The share of female authors in the *BEJEAP* is markedly higher than in the other two journals. The top author variable, the affiliation variables, and the JEL code variables are dummies. "Top author" indicates whether at least one of the authors is among the top 5 percent in at least one of *IDEAS*'s 34 author rankings. Dummies for top institutions indicate whether the author affiliated with the highest-ranked institution is employed by one of the best 10, the top 11-30 or the remaining 31-124 institutions listed in *IDEAS*'s institution ranking.[4] JEL code dummies capture the top level categories of the JEL codes indicated in the article.

About one third of the articles were (co)authored by a top author and more than 50% of the articles were written by authors affiliated with top-listed institutions. Articles published in *BEJM* and *BEJTE* are significantly more likely to have been authored by scholars from Europe than articles published in *BEJEAP* (t-statistic = 5.7). For authors working in North America we observe the opposite. North American authors contribute to almost three out of four articles in *BEJEAP* compared to one half of the articles published in *BEJM* and *BEJTE*. The 'no affiliation' dummy indicates whether no affiliation was provided for at least one author of an article. A lower share of female economists in Europe is a potential explanation for the differences in gender composition at the three journals. The 'non-university' dummy comprises mainly think tanks, international organizations, such as the World Bank, and central banks. The research focus of these institutions appears to entail that their staff work more on topics that fit the aims and scope of the *BEJAEP* and the *BEJM* than the *BEJTE*. According to the JEL code dummies, more than a fifth of the articles in our sample are classified as 'Microeconomics' (JEL code D), 'Macroeconomics and Monetary Economics' (JEL code E), and 'Mathematical and Quantitative Methods' (JEL code C), respectively. For all other categories, the shares are lower. But there is much variation across the three journals. Finally, Table 4.1 shows that the number of articles published in all three journals has increased over time.

Table 4.2 presents descriptive statistics of the citation variables as found in the *IDEAS*

[4]These 124 institutions are the best 25% according to the *IDEAS* database's "Top Level Institutions Ranking", see http://ideas.repec.org/top/top.toplevel.html.

Table 4.1: Descriptive Statistics: Publication Variables

Journal	all	BEJEAP	BEJM	BEJTE
# coauthors	1.7045	1.8242	1.6612	1.5345
share female	0.1495	0.1905	0.1038	0.125
Africa	0.0035 (2)	0 (0)	0 (0)	0.0172 (2)
Asia	0.0682 (39)	0.0513 (14)	0.0929 (17)	0.069 (8)
Europe	0.3619 (207)	0.2454 (67)	0.4754 (87)	0.4569 (53)
Latin America	0.0122 (7)	0.011 (3)	0.0109 (2)	0.0172 (2)
North America	0.6259 (358)	0.7399 (202)	0.5355 (98)	0.5 (58)
Oceania	0.0245 (14)	0.011 (3)	0.0219 (4)	0.0603 (7)
no affiliation	0.042 (24)	0.0659 (18)	0.0164 (3)	0.0259 (3)
non-university	0.208 (119)	0.2198 (60)	0.2678 (49)	0.0862 (10)
Top 10 Inst	0.1556 (89)	0.1758 (48)	0.1202 (22)	0.1638 (19)
Top 11-30	0.1836 (105)	0.1685 (46)	0.2131 (39)	0.1724 (20)
Top 30 plus	0.222 (127)	0.2161 (59)	0.2186 (40)	0.2414 (28)
Top author	0.3444 (197)	0.3297 (90)	0.3825 (70)	0.319 (37)
JEL code A	0.0087 (5)	0.0073 (2)	0.0055 (1)	0.0172 (2)
JEL code B	0.0035 (2)	0 (0)	0.0109 (2)	0 (0)
JEL code C	0.215 (123)	0.1465 (40)	0.1639 (30)	0.4569 (53)
JEL code D	0.3479 (199)	0.2821 (77)	0.1803 (33)	0.7672 (89)
JEL code E	0.2255 (129)	0.044 (12)	0.6284 (115)	0.0172 (2)
JEL code F	0.1136 (65)	0.1502 (41)	0.1311 (24)	0 (0)
JEL code G	0.0664 (38)	0.0549 (15)	0.0546 (10)	0.1121 (13)
JEL code H	0.1241 (71)	0.1941 (53)	0.071 (13)	0.0431 (5)
JEL code I	0.0804 (46)	0.1575 (43)	0.0164 (3)	0 (0)
JEL code J	0.1311 (75)	0.1685 (46)	0.1475 (27)	0.0172 (2)
JEL code K	0.0472 (27)	0.0879 (24)	0 (0)	0.0259 (3)
JEL code L	0.1661 (95)	0.2198 (60)	0.0383 (7)	0.2414 (28)
JEL code M	0.021 (12)	0.044 (12)	0 (0)	0 (0)
JEL code N	0.0087 (5)	0.011 (3)	0.0109 (2)	0 (0)
JEL code O	0.1661 (95)	0.1136 (31)	0.3279 (60)	0.0345 (4)
JEL code P	0.0175 (10)	0.0183 (5)	0.0219 (4)	0.0086 (1)
JEL code Q	0.0647 (37)	0.1245 (34)	0.0055 (1)	0.0172 (2)
JEL code R	0.021 (12)	0.0293 (8)	0.0219 (4)	0 (0)
JEL code Z	0.014 (8)	0.0256 (7)	0.0055 (1)	0 (0)
JEL codes missing	0.0437 (25)	0.0659 (18)	0.0273 (5)	0.0172 (2)
2001	0.0734 (42)	0.0586 (16)	0.0874 (16)	0.0862 (10)
2002	0.0629 (36)	0.0476 (13)	0.0656 (12)	0.0948 (11)
2003	0.1451 (83)	0.1429 (39)	0.153 (28)	0.1379 (16)
2004	0.2273 (130)	0.2821 (77)	0.1749 (32)	0.181 (21)
2005	0.1801 (103)	0.1795 (49)	0.2186 (40)	0.1207 (14)
2006	0.3112 (178)	0.2894 (79)	0.3005 (55)	0.3793 (44)
# articles	572	273	183	116

Number of observations in parentheses next to relative frequencies.

Table 4.2: Descriptive Statistics: Citation Variables

criterion	(sub)set	obs	mean	S.D.	Min	Max
citations	all articles	572	2.451	3.5242	0	22
	BEJEAP	273	2.5568	3.3962	0	20
	BEJM	183	2.612	3.9373	0	22
	BEJTE	116	1.9483	3.0867	0	14
	Front/Adv	104	3.8173	3.916	0	16
	Contributions	212	2.8632	3.9768	0	22
	Topics	256	1.5547	2.6116	0	20
weighted citations	all articles	572	9.5231	17.9754	0	178.885
	BEJEAP	273	9.265	15.9703	0	94.298
	BEJM	183	11.6131	22.8806	0	178.885
	BEJTE	116	6.8335	12.5759	0	69.17
	Front/Adv	104	16.6366	23.2021	0	107.412
	Contributions	212	10.8076	17.6475	0	119.212
	Topics	256	5.5695	14.515	0	178.885

database. *IDEAS* scans the reference lists of all documents uploaded to the database. Since not all files can be read without mistakes, *IDEAS* tends to underreport citation counts compared to other sources. However, we do not think that this is likely to bias our results.[5] To maximize the information contained in our sample, we collected all citations referring to our sample articles up to November 2010, which implies that the citation period depends on the publication dates of the corresponding articles. The number of citations (upper panel) is in the range between zero and 22 per article with an average of 2.45 citations per article. Articles published in the *BEJM* are cited slightly more often than *BEJEAP* articles and attract almost one third more citations on average than *BEJTE* articles. The lower panel refers to citations weighted by simple impact factors as computed by *IDEAS*.[6] Here, articles that appeared in the *BEJM* also perform best. *BEJEAP* articles and *BEJTE* articles receive on average only 80% and 59% of the citations garnered by *BEJM* articles. These differences can reflect field specific citation habits, differences in the journals' quality within their fields, or a combination thereof. Articles published in the *Frontiers* or *Advances* receive significantly more citations than articles published in the lower tiers.[7] Articles published in the *Contributions* receive only 75% of the citations of *Frontiers* and *Advances* articles. *Topics*' articles receive only 41%. These differences are even more pronounced when weighted citations are taken into account, implying that a citation is more likely to come from a higher-quality journal as identified by *IDEAS* if the cited article appeared in one of the two top tiers.

[5] For details concerning RePEc's methodology see Zimmermann (2007).

[6] Impact factors computed by *IDEAS* were retrieved from http://ideas.repec.org/top/seriesfactors.txt.

[7] The difference between articles in *Frontiers* and *Advances* and the two lower tiers is significant at the 1% level with a t-statistic of 4.44.

The editorial system thus appears to perform rather well when it comes to assigning higher impact articles to higher tiers. Yet, a closer inspection reveals that the process of assigning articles to tiers does not work perfectly. The article which received most citations was published in the *Contributions* and the article with the highest score of weighted citations appeared in the *Topics*. Furthermore, uncited articles can be found in all tiers. Hence, not every article that appears in a higher tier performs better in terms of citations than lower-tiered articles. The same applies, of course, to traditional journals: It is well known that the excellent rating of top journals is due to a rather small number of articles which attract an extraordinary number of citations (see e.g. Wall, 2009). One possible explanation for this phenomenon is that papers submitted to higher-ranked journals tend to be characterized by a higher degree of originality. Since these studies have fewer links to the existing literature, their quality may be harder to judge.[8] This implies more citations on average in higher-ranked journals, but also more variation. The numbers shown in Table 4.2 appear to corroborate this view. As we move to higher tiers, standard deviations increase in three of the four cases. This picture is, however, reversed when coefficients of variation are taken into account, i.e. when standard deviations are divided by the respective means in order to control for higher averages in higher tiers. The coefficients of variation actually turn out to be higher in lower tiers.

4.4 Econometric Analysis

4.4.1 Determinants of Editorial Sorting

In this section, we analyze the determinants of editorial sorting. Table 4.3 shows the results of four ordered probit regressions that examine whether observed characteristics are correlated with the editors' quality assessment of the submitted manuscripts. Previous investigations have shown that co-authored studies tend to be of higher quality (see e.g. Laband and Tollison, 2000; Ursprung and Zimmer). We also find that the number of co-authors increases the likelihood of a paper being published in a higher tier. However, this effect becomes insignificant when we control for the authors' personal rankings, for the rankings of the institutions they are affiliated with, and for JEL codes. The share of

[8] Vandermeulen (1972) provides a list of six types of manuscripts that appear in journals of different quality. For instance, according to Vandermeulen, the staple ingredient of average journals is a type called *hunting trophies*. *Hunting trophies* are "gained by applying, purifying, extending or testing the visions" in another, more original category: *creative insights* reached by top scholars.

female authors is not significant in any specification.

Table 4.3: Ordered Probit Estimates

	(1)	(2)	(3)	(4)
# coauthors	0.111*	0.137**	0.0119	0.0298
	(0.0605)	(0.0610)	(0.0666)	(0.0672)
share female	-0.103	-0.0411	0.0638	0.0735
	(0.156)	(0.160)	(0.163)	(0.167)
Africa	-0.0537	-0.0510	-0.0265	-0.0839
	(0.826)	(0.830)	(0.816)	(0.821)
Asia	-0.0281	0.00500	0.0321	0.0706
	(0.212)	(0.213)	(0.221)	(0.222)
Europe	-0.0690	-0.109	-0.144	-0.160
	(0.168)	(0.169)	(0.177)	(0.178)
Latin America	0.192	0.0533	0.0841	-0.0239
	(0.428)	(0.431)	(0.439)	(0.443)
North America	0.434**	0.358**	0.145	0.101
	(0.171)	(0.172)	(0.186)	(0.187)
Oceania	-0.648*	-0.824**	-0.897**	-0.981**
	(0.362)	(0.371)	(0.388)	(0.392)
no affiliation	-0.306	-0.320	-0.0829	-0.144
	(0.258)	(0.261)	(0.292)	(0.295)
non-university			-0.175	-0.158
			(0.142)	(0.143)
Top 10 Institution			0.915***	0.914***
			(0.162)	(0.166)
Top 11-30			0.583***	0.555***
			(0.151)	(0.153)
Top 30 plus			0.520***	0.506***
			(0.134)	(0.135)
Top author			0.372***	0.372***
			(0.111)	(0.113)
JEL C			0.236*	0.235*
			(0.127)	(0.129)
JEL E			-0.207	-0.228*
			(0.137)	(0.138)
year dummies	no	yes	no	yes
additional controls	no	no	yes	yes
Pseudo-R^2	0.0307	0.0498	0.0988	0.114
Observations	572	572	572	572

Standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1; Dependent variable: Frontiers = 4, Advances = 3, Contributions = 2, Topics = 1; Note that continent dummies are not mutually exclusive, which is why no reference category is needed; Coefficient on JEL code reported only if significant in at least one specification.

Next, we check how the authors' geographic locations are correlated with sorting. It turns out that location has no significant effect, except for authors from Oceania and North America. However, the coefficient for North America appears to be positive only because North America is the home of many of the world's top institutions. If one of the authors is affiliated with such a top institution, sorting into higher tiers becomes more likely. When controlling for this effect, the North America dummy variable loses statistical significance. Top authors tend to publish in higher tiers. The last two findings may be due

to a causal relationship. But it may just as well be the case that highly ranked scholars and scholars from highly reputed institutions produce higher-quality output. Finally, JEL code C (Microeconomics) has a positive impact on editorial sorting, while the effect of JEL code E (Macroeconomics and Monetary Economics) is negative. Only the significance of JEL code E is affected by whether year dummies are included in our regressions. The coefficients on the remaining JEL codes are not significant.[9]

4.4.2 Determinants of Citation Success

We now turn to analyzing how author characteristics and editorial sorting are related to subsequent citation success. Even though we use "citations weighted by simple impact factors" as our dependent variable and not the raw number of citations, we treat "citations" as a count variable. Referring to the Pseudo Maximum Likelihood method pioneered by Gourieroux, Monfort, and Trognon (1984), Santos Silva and Tenreyro (2006) stress that all that is required for count data models to give consistent and robust estimates is the conditional mean to be correctly specified. We choose the Poisson model because it makes weaker distributional assumptions than the Negbin model, it is semiparametrically robust, and its estimates can be interpreted as semielasticities. As mentioned before, citation periods and the number of citations depend on the publication dates of the respective articles. We therefore also control for the year of publication. Table 4.4 shows the results. Robustness checks with raw citations and recursively weighted citations as dependent variables that can be found in Table 4.8 in the Appendix lead to similar results.

Specification (1) includes dummies for journals, tiers, the respective interaction terms, and publication year dummies. Articles published in the *BEJEAP* which is our baseline category and in the *BEJM* attract significantly more citations than papers published in the *BEJTE*, and, presumably in line with the editorial intention, articles published in higher tiers receive significantly more citations than articles published in lower tiers. Citation differences between the tiers are less pronounced in the *BEJM* than in the other two journals. The coefficients of the publication year dummies[10] show that articles that were published earlier were cited more often than articles published in the reference year 2006.

These findings admit two different interpretations: Articles in higher tiers may either attract more citations because they are better or because researchers are more likely to

[9] Please see Table 4.6 in the Appendix for the complete results of specification (4).
[10] Available from the authors upon request.

read articles allocated to higher tiers. However, as suggested by one of the referees, the tiered structure of the *B.E. Press*'s journals helps along these lines, too. It allows authors with papers in lower tiers to "free ride" off the articles in higher tiers: potential readers might initially be attracted by an *Advances* article but then also notice another paper lower down. Traditional journals without tiers, in contrast, are entirely separate and not indexed together. Someone looking for a particular paper in macroeconomics that was published in a top general interest journal such as the *American Economic Review* cannot at the same time see another paper published in a top field journal such as the *Journal of Monetary Economics*.

Specification (2) adds the number of authors and the share of female authors. In this specification, the share of female authors has no significant effect on the number of citations. Articles with a larger number of authors, however, appear, at a first glance, to get cited significantly more often. This result, alas, is not robust. In specification (3), we regress the weighted number of citations also on other author characteristics, but not on journal tiers. Qualitatively, the results are similar to our findings for editorial sorting (see Table 4.3): the estimates indicate that the number of authors is not significantly related to citation success. Being a top author and being affiliated with a top level institution according to *IDEAS*'s rankings both increase citation success. This effect is most pronounced for articles with authors from top-10 institutions. These articles are cited more than twice as often as articles with no top author and no author from a top institution. Articles by authors from non-university institutions attract significantly fewer citations. The coefficients of the geographic variables are somewhat different from those detailed in Table 4.3. Articles with authors from institutions in Europe, Latin America, and Oceania receive significantly fewer citations, while articles from authors with Asian affiliations perform significantly better. The unreported coefficients on Africa and North America are not significant.[11]

The coefficients on the JEL codes in specification (4) are shown in Table 4.7 in the Appendix. Just as Ursprung and Zimmer, we find that citation intensity varies significantly across fields. For instance, articles coded as *Law and Economics* (JEL code K) and *Industrial Organization* (JEL code L) get cited particularly often, whereas little attention is paid to articles coded as *Financial Economics* (JEL code G) and *Agricultural and*

[11] See Table 4.7 in the Appendix. Recall that no reference category for geographic location is required because the sum of their means is larger than one (see Table 4.1).

Table 4.4: Poisson Estimates

	(1)	(2)	(3)	(4)
BEJM	0.510***	0.516***	0.156**	0.482***
	(0.0760)	(0.0762)	(0.0709)	(0.0854)
BEJTE	-0.877***	-0.849***	-0.446***	-0.506***
	(0.126)	(0.126)	(0.0950)	(0.129)
Contributions	0.774***	0.761***		0.408***
	(0.0543)	(0.0544)		(0.0565)
Frontiers/Advances	1.229***	1.214***		0.803***
	(0.0560)	(0.0562)		(0.0594)
BEJM×Cont	-0.241***	-0.235***		-0.200***
	(0.0745)	(0.0745)		(0.0760)
BEJM×Front/Adv	-0.323***	-0.345***		-0.567***
	(0.0871)	(0.0869)		(0.0933)
BEJTE×Cont	0.195	0.215*		0.515***
	(0.122)	(0.122)		(0.128)
BEJTE×Front/Adv	0.0520	0.0580		-0.0135
	(0.128)	(0.128)		(0.134)
# coauthors		0.0703***	-0.00106	-0.0153
		(0.0153)	(0.0189)	(0.0192)
share female		-0.0258	0.138***	0.146***
		(0.0480)	(0.0501)	(0.0506)
Asia			0.243***	0.214***
			(0.0590)	(0.0599)
Europe			-0.169***	-0.110**
			(0.0503)	(0.0506)
Latin America			-1.852***	-1.715***
			(0.282)	(0.282)
Oceania			-1.212***	-1.044***
			(0.164)	(0.166)
no affiliation			0.265***	0.286***
			(0.0924)	(0.0935)
non-university			-0.356***	-0.289***
			(0.0432)	(0.0441)
Top 10 Institution			1.017***	0.881***
			(0.0471)	(0.0489)
Top 11-30			0.901***	0.818***
			(0.0460)	(0.0467)
Top 30 plus			0.567***	0.483***
			(0.0442)	(0.0449)
Top author			0.538***	0.510***
			(0.0328)	(0.0331)
additional controls	yes	yes	yes	yes
Pseudo-R^2	0.142	0.144	0.239	0.259
Observations	572	572	572	572

Standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1.

Natural Resource Economics / Environmental and Ecological Economics (JEL code Q). Moreover, after controlling for publication topics, the coefficient on the share of female authors becomes significantly positive: female authors appear to work on topics, which do not attract many citations (unreported results show that the coefficient on female share is not significantly different from zero if we control for all other author characteristics but not for JEL codes). Within given fields, however, articles written by women attract significantly more citations, be it through better quality or positive discrimination.

Articles by top-ranked authors and by authors affiliated with top institutions are thus not only more likely to appear in higher tiers, they also get cited more frequently. In specification (4), we include journal tiers as well as author characteristics. Articles with authors affiliated with top-10 institutions still attract 90 percent more citations than articles without top authors or authors from top institutions. In other words, the information that the authors' names and affiliations provide to editors and referees is not fully accounted for by editorial sorting. If the editors' objective had simply been to allocate articles that are more likely to attract many citations into higher tiers, they should have more heavily relied on who the authors are and which institutions the authors are affiliated with.

We offer three possible explanations for the finding that editors might ignore some relevant information. First, the editors may be biased against highly ranked authors, possibly because editors want to promote less experienced researchers. Second, the editors may overrate research which is of specific interest to them. Idiosyncratic interests can, for example, be triggered by a paper's topic or its main conclusion. Note that the editors might not even be conscious of having this kind of bias. Third, citations may be an inadequate measure of research quality, and highly skilled editors may be less impressed by an author's research prowess and affiliation than the citing profession at large. In other words, editors may well provide a more reliable measure of research quality than the authors' peers.

4.4.3 Performance of Editorial Sorting

After having provided some insights into the determinants of citation success, we now turn to assessing editorial performance assuming that the editors' objective had been to sort higher impact papers into higher tiers. As we have already seen, even after controlling for author characteristics, citations depend significantly and positively on a paper's ranking

by tier. Hence, on average, journal quality provides a useful signal of the impact a research article ultimately will have. In this section, we will take a look at what happens beyond the average. We, therefore, interpret peer reviewing as a testing procedure. An editor decides to accept a paper for publication if he considers it to be of interest to a certain subset of the scientific community. This subset may, for instance, consist of all game theorists worldwide. Under the null hypothesis, the paper is not relevant for this entire audience. From this population the editor picks two or three referees who provide a judgement on whether to reject the null hypothesis and to accept the paper, or whether to reject the paper.

To evaluate the quality of editorial sorting we rely on citation counts to measure article impact since a better measure is lacking. Our objective is to determine how many type I and type II errors the editors have made if their objective had been to sort highly cited articles (in relative terms, i.e. conditional on the topic as given by the JEL code) into higher tiers. We use our results from regression (3) in Table 4.4 to adjust each article's score of weighted citations to a reference level in order to make citation scores comparable across journals, topics, and years. More precisely, we calculate the expected citation scores of all articles, correcting for the three different journals (*BEJEAP*, *BEJM* and *BEJTE*) and publication years as well as JEL codes. For these expected scores, we assume the author to be one male scholar affiliated with an institution in North America. Neither the author nor the institution are top-ranked by *IDEAS*. We then subtract these estimates from the actual citation scores of the corresponding articles. Finally, we use these adjusted citation scores to sort articles into tiers. Table 4.5 shows the joint distribution of articles according to editorial and (ex post) citation-based sorting.

Recall that we interpret the editorial sorting procedure to be equivalent to sequential submissions to different journals. The only difference is that in this case the sorting decision is made in one step, i.e. without the article going through further rounds of revisions. We assume that all authors would prefer to have their articles published in the *Frontiers* and *Advances* tiers. If an article is published in the *Contributions* tier, we interpret this as the article having been rejected by *Frontiers* and *Advances* and then having been accepted for publication in *Contributions*. If an article has appeared in *Topics* this means that it has been rejected by all three higher-ranked tiers.

For the *Frontiers* and *Advances* "journals" we find that 68.3% of all published articles

Table 4.5: Editorial vs. Citation-based Sorting

		published as Frontiers/Advances	Contributions	Topics
citation-based	Frontiers/Advances	33	50	21
	Contributions	40	71	101
	Topics	31	91	134
	sum	104	212	256

should have been rejected (type I error). Moreover, 15.2% of the articles rejected by *Frontiers* and *Advances* actually should have been published in these top tiers (type II error). An alternative and maybe more intuitive measure is the factor by which editors outperform random assignment. This editorial performance index amounts to 1.75 which means that the share of correctly sorted articles is 75% higher if editors are in charge. Note that the size of the three measures (type I and type II errors, and editorial performance) is to some extent driven by sample size and the acceptance rate.[12] Assuming that the *Frontiers*, *Advances*, and *Contributions* form a single journal, and the *Topics* represented the set of rejected articles, we obtain a type I error of 38.6%, a type II error of 47.7%, and editorial performance of 1.11. Since comparable figures for other journals are not available, it is not possible to assess the relative performance of the review process at the *B.E. Journals*, and we refrain from any discussion.

One drawback of our analysis is that we do not observe the performance of articles which were rejected at all four tiers and which are therefore not included in our data set. Because of this lack of data, we cannot avoid underestimating the type I errors. And since rejected submissions are likely to perform worse than the accepted articles, we are likely to overestimate the type II errors and to underestimate the performance of the editorial system relative to random assignment. Our estimates of the importance of editorial mistakes may also be biased by missing data on cases in which authors withdrew their manuscripts when they were disappointed with an editorial decision to publish them in one of the lower tiers. Another drawback is the use of citations as a proxy for real impact. Adopting the view that citations favor top authors and authors affiliated with top institutions, we are likely to underestimate the performance of the review process. Finally, we note that in some cases the classification of articles into citation-based tiers depends on very small differences, which might not be significant given possible measurement errors.

[12]Suppose acceptance for publication was purely random. Then type I error would approach one and type two error zero as the number of published articles declines. On the other hand, the type I error converges to zero and the type two error to one as the share of accepted articles increases.

4.5 Conclusion

It is well known that assessing the overall impact of a piece of economic research at an early stage is not an easy assignment. Editing learned journals is therefore a tricky business. Our results suggest that, on average, peer-reviewing yields accurate estimates of which papers will have an impact and which will not. Given the severe shortcomings of other measures such as citation counts, it makes sense that quality-weighted journal publications are used as a research evaluation instrument. In times of increasing specialization and narrowing research interests (see e.g. Jones, 2009), peer-reviewing essentially represents division of labor in research evaluation. No individual economist can follow all new developments in the discipline, but journal editors can always attempt to pick referees whom they expect to best know the literature and methods related to a particular submission. Yet, we have also shown that a substantial number of errors occur in the editorial process.

To be sure, the aim of a journal editor is not confined to maximizing his journal's reputation as measured by the impact factor. Every economist has an idiosyncratic view of what is important in the discipline and in which direction future research should evolve. McAfee (2010) notes that although good editors are characterized by a lack of a personal agenda, they have "an opinion about everything". It is, of course, often hard to distinguish between the two. A personal agenda constitutes a bias and will lead to bad decisions because the editor may be led to reject excellent submissions that do not fit his agenda and to accept bad submissions that do. An editor's personal vision of economics, in contrast, encompasses all aspects and fields of economic research and simply summarizes the editor's conception of research quality.

Our findings give a first idea of the efficacy of journal editing. But we also note that our results suffer from the absence of data on rejections and from the fact that we do not know the objective function of the *B.E. Journals'* editors. Further research will be required to provide alternative measures of editorial performance.

4.6 Appendix

Table 4.6 shows results of specification (4) of the ordered probit regression above including all JEL codes with at least 20 observations. Table 4.7 presents results of specification (4) of the poisson regression above including all covariates except dummies for the publication years. Table 4.8 shows the robustness of our results with respect to the dependent variable. In addition to citations weighted by simple impact factors we also use the number of citations and citations weighted by recursive impact factor. We choose citations weighted by simple impact factors as baseline model, because it fits the data best. The reason why simple impact factors perform better than recursive ones is probably that the recursive weighting scheme is more convex and, therefore, closer to uniform weighting.

Table 4.6: Ordered Probit Estimates

# coauthors	0.0298	JEL C	0.235*
	(0.0672)		(0.129)
share female	0.0735	JEL D	0.0955
	(0.167)		(0.115)
Africa	-0.0839	JEL E	-0.228*
	(0.821)		(0.138)
Asia	0.0706	JEL F	-0.0702
	(0.222)		(0.170)
Europe	-0.160	JEL G	-0.326
	(0.178)		(0.210)
Latin America	-0.0239	JEL H	-0.0662
	(0.443)		(0.155)
North America	0.101	JEL I	0.0126
	(0.187)		(0.198)
Oceania	-0.981**	JEL J	-0.159
	(0.392)		(0.155)
no affiliation	-0.144	JEL K	-0.0585
	(0.295)		(0.235)
non-university	-0.158	JEL L	0.00605
	(0.143)		(0.143)
Top 10 Institution	0.914***	JEL O	-0.228
	(0.166)		(0.139)
Top 11-30	0.555***	JEL Q	0.133
	(0.153)		(0.205)
Top 30 plus	0.506***	other JEL	0.115
	(0.135)		(0.155)
Top author	0.372***	year dummies	yes
	(0.113)		
Pseudo-R^2	0.114	Observations	572

Standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1;
Dependent variable: Frontiers = 4, Advances = 3, Contributions = 2, Topics = 1.

Table 4.7: Poisson Estimates

BEJM	0.482***	Top 10 Institution	0.881***
	(0.0854)		(0.0489)
BEJTE	-0.506***	Top 11-30	0.818***
	(0.129)		(0.0467)
Contributions	0.408***	Top 30 plus	0.483***
	(0.0565)		(0.0449)
Frontiers/Advances	0.803***	Top author	0.510***
	(0.0594)		(0.0331)
BEJM×Cont	-0.200***	JEL C	0.0807**
	(0.0760)		(0.0388)
BEJM×Front/Adv	-0.567***	JEL D	-0.207***
	(0.0933)		(0.0357)
BEJTE×Cont	0.515***	JEL E	-0.0267
	(0.128)		(0.0464)
BEJTE×Front/Adv	-0.0135	JEL F	0.149***
	(0.134)		(0.0456)
# coauthors	-0.0153	JEL G	-0.950***
	(0.0192)		(0.0801)
share female	0.146***	JEL H	0.00875
	(0.0506)		(0.0456)
Africa	-0.846	JEL I	-0.124*
	(0.688)		(0.0651)
Asia	0.214***	JEL J	0.197***
	(0.0599)		(0.0417)
Europe	-0.110**	JEL K	0.466***
	(0.0506)		(0.0638)
Latin America	-1.715***	JEL L	0.340***
	(0.282)		(0.0404)
North America	-0.0381	JEL O	0.120***
	(0.0553)		(0.0375)
Oceania	-1.044***	JEL Q	-0.498***
	(0.166)		(0.0743)
no affiliation	0.286***	JEL other	0.0492
	(0.0935)		(0.0436)
non-university	-0.289***	Constant	1.103***
	(0.0441)		(0.0863)
year dummies	yes		
Pseudo-R^2	0.259	Observations	572

Standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1.

Table 4.8: Robustness of Poisson Estimates

	cites	w cites simple	w cites recursive
BEJM	0.142	0.482***	0.670**
	(0.169)	(0.0854)	(0.305)
BEJTE	-0.671***	-0.506***	-0.448
	(0.242)	(0.129)	(0.463)
Contributions	0.412***	0.408***	0.554***
	(0.103)	(0.0565)	(0.202)
Frontiers/Advances	0.659***	0.803***	0.879***
	(0.112)	(0.0594)	(0.214)
BEJM×Cont	-0.245*	-0.200***	-0.255
	(0.147)	(0.0760)	(0.266)
BEJM×Front/Adv	-0.565***	-0.567***	-0.494
	(0.191)	(0.0933)	(0.319)
BEJTE×Cont	0.523**	0.515***	0.408
	(0.229)	(0.128)	(0.448)
BEJTE×Front/Adv	0.180	-0.0135	-0.0309
	(0.244)	(0.134)	(0.469)
# coauthors	0.0630*	-0.0153	-0.0178
	(0.0346)	(0.0192)	(0.0679)
share female	0.0308	0.146***	0.0800
	(0.0993)	(0.0506)	(0.177)
Asia	-0.179	0.214***	0.247
	(0.130)	(0.0599)	(0.206)
Europe	-0.140	-0.110**	-0.0781
	(0.102)	(0.0506)	(0.171)
Latin America	-0.559*	-1.715***	-1.858*
	(0.303)	(0.282)	(0.962)
Oceania	-0.769***	-1.044***	-1.018*
	(0.279)	(0.166)	(0.574)
no affiliation	-0.250	0.286***	0.561*
	(0.196)	(0.0935)	(0.304)
non-university	-0.136	-0.289***	-0.329**
	(0.0832)	(0.0441)	(0.151)
Top 10 Institution	0.551***	0.881***	0.806***
	(0.0929)	(0.0489)	(0.170)
Top 11-30	0.476***	0.818***	0.839***
	(0.0877)	(0.0467)	(0.160)
Top 30 plus	0.254***	0.483***	0.469***
	(0.0836)	(0.0449)	(0.155)
Top author	0.489***	0.510***	0.436***
	(0.0642)	(0.0331)	(0.114)
additional controls	yes	yes	yes
Pseudo-R^2	0.173	0.259	0.181
Observations	572	572	572

Standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1.

Bibliography

Arrow, K. J., B. D. Bernheim, M. S. Feldstein, D. L. McFadden, J. M. Poterba, and R. M. Solow (2011): "100 Years of the *American Economic Review*: The Top 20 Articles," *American Economic Review*, 101(1), 1–8.

Bornmann, L., and H.-D. Daniel (2008): "The Effectiveness of the Peer Review Process: Inter-Referee Agreement and Predictive Validity of Manuscript Refereeing at Angewandte Chemie," *Angewandte Chemie International Edition*, 47(38), 7173–7178.

Coupé, T. (2010): "Peer Review versus Citations - An Analysis of Best Paper Prizes," Discussion Papers 35, Kyiv School of Economics.

De Long, J. B., and K. Lang (1992): "Are All Economic Hypotheses False?," *Journal of Political Economy*, 100(6), 1257–72.

Gans, J. S., and G. B. Shepherd (1994): "How Are the Mighty Fallen: Rejected Classic Articles by Leading Economists," *Journal of Economic Perspectives*, 8(1), 165–79.

Gourieroux, C., A. Monfort, and A. Trognon (1984): "Pseudo Maximum Likelihood Methods: Applications to Poisson Models," *Econometrica*, 52(3), 701–20.

Hamermesh, D. S. (1994): "Facts and Myths about Refereeing," *Journal of Economic Perspectives*, 8(1), 153–63.

Hamermesh, D. S., and G. A. Pfann (2011): "Markets for Reputation: Evidence on Quality and Quantity in Academe," *Economic Inquiry*, forthcoming.

Hamermesh, D. S., and P. Schmidt (2003): "The Determinants of Econometric Society Fellows Elections," *Econometrica*, 71(1), 399–407.

Jones, B. F. (2009): "The Burden of Knowledge and the 'Death of the Renaissance Man': Is Innovation Getting Harder?," *Review of Economic Studies*, 76(1), 283–317.

Laband, D. N., and M. J. Piette (1994): "Favoritism versus Search for Good Papers: Empirical Evidence Regarding the Behavior of Journal Editors," *Journal of Political Economy*, 102(1), 194–203.

Laband, D. N., and R. D. Tollison (2000): "Intellectual Collaboration," *Journal of Political Economy*, 108(3), 632–661.

——— (2003): "Dry Holes in Economic Research," *Kyklos*, 56(2), 161–173.

McAfee, R. P. (2010): "Edifying Editing," *The American Economist*, 55(1), 1–8.

Opthof, T., F. Furstner, M. van Geer, and R. Coronel (2000): "Regrets or no regrets? No regrets! The fate of rejected manuscripts," *Cardiovascular Research*, 45(1), 255–258.

Santos Silva, J. M. C., and S. Tenreyro (2006): "The Log of Gravity," *The Review of Economics and Statistics*, 88(4), 641–658.

Ursprung, H. W., and M. Zimmer (2007): "Who is the "Platz-Hirsch" of the German Economics Profession? A Citation Analysis," *Journal of Economics and Statistics (Jahrbuecher fuer Nationaloekonomie und Statistik)*, 227(2), 187–208.

Vandermeulen, A. (1972): "Manuscripts in the maelstrom: A theory of the editorial process," *Public Choice*, 13(1), 107–111.

Wall, H. J. (2009): "Don't Get Skewed Over by Journal Rankings," *The B.E. Journal of Economic Analysis & Policy*, 9(1), 34.

Wilson, J. D. (1978): "Peer Review and Publication. Presidential address before the 70th annual meeting of the American Society for Clinical Investigation, San Francisco, California, 30 April 1978.," *Journal of Clinical Investigation*, 61(6), 1697–1701.

Zimmermann, C. (2007): "Academic Rankings with RePEc," Working papers 2007-36, University of Connecticut, Department of Economics.

Chapter 5

Standardizing Grades

5.1 Introduction

Grades and grade point averages (GPA) play an important role in signaling students' intellectual abilities. The economic value of this signal, i.e. the extent to which employers, managers of graduate programs, etc. can rely on summary statistics such as the GPA, depends crucially on how well the GPA reflects the true ability of the student. There are many reasons why the GPA might not properly reflect abilities: exams usually cover only a selection of the relevant material, student performance is subject to shocks, such as health or personal problems, and grading can be idiosyncratic or even erroneous. These factors of influence are very hard to eliminate. We identify however two additional factors that can be controlled for: the difficulty of the exam and the composition of the students taking the exam.

We define the exam's difficulty as the grade that a student endowed with average ability would obtain when performing at his or her average. Variations in the difficulty of exams can arise, for example, from the selection of topics, the specific design of the problems, and - across courses - from differences in the complexity of the material taught, or differences in the aspiration level of the instructor. The composition of the student body attending a course has also a bearing on grading because it is likely to influence the lecturer's appraisal of the students' ability to cope with the material taught. An instructor who is used to teaching a selection of high performing students might consider the students to be less competent than they actually are. If, on the other hand, mostly sub-prime students select themselves into a course, the instructor may well confuse weak student performance with an excessively demanding course material or an excessively difficult exam.

Relative grading schemes (also known as grading on a curve), such as the ECTS grading scheme,[1] correct for variations in exam difficulty. Relative grades do, however, not correct for the self selection of students into courses and thereby into exams. We therefore suggest to use a regression-based approach. The proposed method assigns standardized grades that take into account the ability distribution of the students who take the various exams. The method works as follows. The first step consists in estimating the relationship between grades awarded in each course and the students' ability. Based on this estimate, students who have not taken some exam are, in a second step, assigned a grade for this exam,

[1]The European Credit Transfer System (ECTS) stipulates that the best 10% of students in an exam receive grade A, the next best 25% grade B, and so on.

given their ability. In the third step, taking into account the grades of the participating students and the stipulated grades of the students who have not actually taken the exam, all students are assigned relative grades for each course. In the forth step the average relative grades are computed for each student whose ability is then identified with this average relative grade. Since the students' abilities serve as an input for the first step, the method consists of a feedback control system that can analytically be represented as a system of equations whose solution yields the standardized grades.

The basic idea of the proposed method thus consists in simulating a fictitious study program in which all students take all courses. In such an environment in which no self-selection into courses takes place, the problems associated with relative grading disappear and we can easily assign each student standardized grades that much better reflects his or her actual performance than any traditional system.[2]

In the education economics literature grading is a neglected issue. To be sure, many studies use grades as an indicator of academic achievement, but questions relating to the determinants and effects of grading have not yet been dealt with at great depth in the economics literature. Dubey and Geanakoplos (2005) provide an interesting game theoretic analysis of the incentives deriving from the coarseness of the grading scale and Bar and Zussman (forthcoming) find that the political orientation of university professors affect the variance of the employed grade distribution, republican professors being associated with a less egalitarian grade distributions. Somewhat better represented in the economics literature are studies on grade inflation. Beginning with the study by Sabot and Wakeman-Linn (1991) this issue has attracted considerable attention - especially in the Journal of Economic Perspectives. Two recent contributions are Achen and Courant (2009) and Bar, Kadiyali, and Zussman (2009). Our class-rank based method to standardizing grades may contribute to taking the sting out of grade inflation.

Our method is related to the method proposed by Johnson (1997) that also relies on regression analysis. The main difference between the two methods is that we standardize the grades of each single exam whereas Johnson (1997) standardizes the overall achievement of students. Our approach thus appears to be better suited to conveying the full picture of the academic achievements of individual students. Further related methods that also focus on standardizing a student's overall achievement are, for example, Young (1990)

[2]Notice, that we assume that all courses have the same number of credit points. Applying our method to course programs in which courses differ with respect to credit points is a trivial endeavour.

and Caulkins, Larkey, and Wei (1996).

5.2 Theory

Consider a set of I students. Each student takes J_i courses (exams) from a set J of courses (exams). The performance $p_{i,j}$ of student i in exam j depends on i's ability a_i and random noise $\epsilon_{i,j}$: $p_{i,j} = a_i + \epsilon_{i,j}$, where $\epsilon_{i,j} \sim F_j(.)$ is a zero-mean random variable. We assume that grades are a function of student performance, with better performance translating into better grades. Without loss of generality, and in line with grading in Germany, grades with low numerical values are associated with good performance:

$$g_{i,j} = f_j(a_{i,j} + \epsilon_{i,j}) \quad \left| \frac{\partial f_j(.)}{\partial a_{i,j}} = \frac{\partial f_j(.)}{\partial \epsilon_{i,j}} > 0 \right. . \tag{5.1}$$

The objective is to assess i's performance in one of the exams i has taken relative to all $I - 1$ peers even though some of these peers have not taken this exams.

If students' abilities were known, we could estimate the grading function f_j by regressing grades on abilities. Moreover, using $\hat{f}_j$ we could estimate how students who have not taken a certain exam would have performed if they had taken it. In particular we could calculate the probabilities with which each of these non-participants would have received any available grade and derive the corresponding probability distribution. Aggregating the actual distribution over grades of participants and the inferred distribution over grades of non-participants, we would then obtain a grade distribution of the whole student body. Such a comprehensive distribution can then easily be translated into *standardized* grades $g^*_{i,j}$ which can then be assigned to the actual participants.

The problem of our approach is that we do not know the students' abilities and thus can neither run the regression nor perform the imputation of grades to non-participants. If standardized grades were however known to start with, we could use these grades to predict student ability, because the expected random component of a student is zero. Formally:

$$\frac{1}{J_i} E\left[\sum_{j \in J_i} f^{-1}(g^*_{i,j})\right] = \frac{1}{J_i} \sum_{j \in J_i} E[f^{-1}(f(a_i + \epsilon_{i,j}))] = a_i. \tag{5.2}$$

Standardized grades can thus be estimated using abilities, and abilities can be estimated using standardized grades. Our problem can thus be represented as a solvable

system of equations. In mathematical terms we have $g^*_{i,j} = g^*_{i,j}(\{g_{i,j}\}_{i\in I_j}, \{a_i\}_{i\in I})$ and $a_i = \frac{1}{n_i}\sum_{j\in J_i} g^*_{i,j} = a_i(\left\{g^*_{i,j}\right\}_{j\in J_i})$. Combining the two formulas (and abusing notation) we obtain $g^* = g^*(g, a(g^*)) = g^*(g)$. The standardized grade vector g^* turns out to be an implicit function of the vector of raw grades g; we thus have to deal with a system of equations. Since we rely on standard regression methods, we minimize the sum of squared residuals resulting from the system.

This procedure implies that we can determine which set of abilities is most likely to give rise to the raw grades that we actually observe. And we can then use these abilities to compute standardized grades.

5.3 Implementation

5.3.1 Estimation strategy

The estimation strategy depends on the grading scheme applied by the respective educational establishment and the sample to be used. In particular, one has (1) to account for the type of grades that can contain cardinal or ordinal information, (2) to specify whose students' exam data will be used to estimate the grade function f and whose students' exam portfolio will be augmented by imputation, and (3) to specify which exams will be standardized.

In our showcase application we have 11 ordinal grades running from 1.0 (best), 1.3, 1.7, 2.0, ... , 4.0 (marginal pass grade) to 5.0 (worst = failed). The natural estimation method would be ordered probit. However, we chose not to use ordered probit because estimating threshold levels is not feasible in many exams due to the small number of participants. Instead we rely on interval regression which is essentially equivalent to an ordered probit regression with fixed threshold levels. We set the ten threshold levels to 1, 2, ... , 9, and 11, where the bigger interval for grade 4.0 reflects the fact that this grade is more frequently awarded than the other grades 1.0 to 3.7.

The regression equation for the latent variable is $\hat{g}_{i,j} = \alpha_j + \beta_j \hat{a}_i + e_{i,j}$, where the estimate for $\hat{a}_i$ is not based on the current exam to avoid endogeneity of the results, i.e. $\hat{a}_i = \sum_{k\neq j} g^*_{i,j}$. Since some students in our sample took part in only a few exams, we exclusively rely in our regressions and imputations on grades of students who have taken at least six exams. As the precision of the estimated ability of the students depends on

the number of exams taken, we account for heteroscedasticity in our regression. Finally, we standardize only grades of exams taken by at least 8 students who can be relied on for the regression and imputation.

5.3.2 Computational implementation

We use an iterative method to standardize grades, i.e. to solve the system $g^* = g^*(g)$. First, we choose some vector of initial values of student abilities. The choice of the initial ability values appears not to be relevant for the final result.[3] We then run iterations to update our estimates of the students' abilities until convergence is achieved.

For each exam j, each iteration consists of four steps::

1. calculate students' abilities based on all exams except j,
2. regress grades on ability,
3. impute for the students who did not take exam j the probability distribution over grades in exam j (given their ability levels and the regression results) and calculate the joint grade distribution of participants and non-participants, and
4. assign standardized grades to participants.

In step iii, in which the distribution of the expected grades of non-participants is merged with the actual distribution of grades of the participants, both distributions are weighted by the number of non-participants and participants respectively. Based on the joint distribution we then assign a quantile to each grade which we interpret as standardized grades with respect to current ability.

After standardizing all exams, we use the standardized GPA to update ability, check for convergence and repeat the process, if necessary.

5.4 Showcase application

5.4.1 Data

We use examination data of economics students at the University of Konstanz. The set of exams is restricted to 231 exams that were taken by at least 8 students in the years

[3] Choosing the average grade of all relevant exams as starting vector leads to the same results as choosing the average grade of all exams, a constant (i.e. dropping the ability for the first iteration), and seven random vectors drawn from a uniform distribution.

2004 to 2008. All exams of these were taken by junior and senior (third and forth year) students. The grades awarded in these exams thus counted for the graduation grade point average which is the most important signal in the job market.[4] These exams were therefore highly incentivized. Our sample comprises in principle the whole population of exams and consists of the exam grades of 1158 different students who have taken at least one of these exams. Thereof 709 students have taken at least 6 relevant exams and form the basis for our regressions and imputations.

An important aspect of our data is that it is highly connected. The students in our dataset could, subject to certain rules, rather freely choose to enroll in the courses offered for junior and senior students. Different curricular preferences, different success strategies (for example optimization of the graduation GPA), and also differences in the chosen duration of study (in Germany the regular duration of study is usually exceeded) gave rise to a fairly mixed student body (according to cohort, ability, etc.) attending the courses and taking the respective exams. This overlapping pattern of students improves the convergence properties of our method and renders the results more accurate.

5.4.2 Results

Applying the proposed method to our data set, we obtain the following distribution of abilities as measured by standardized GPAs. Figure 5.1 reveals that measured abilities in our sample are not distributed smoothly which may be due the small number of only 709 observations.

Figure 5.2 shows the distribution of grades actually assigned to participants and the distribution of imputed grades for non-participants based on all exams in our sample. Except for the grades 1.0, 4.0, and 5.0 the two distributions are quite similar. The differences that we observe for grades 4.0 and 5.0 may result from imposing an inappropriate threshold level between the two grades. A first analysis of the log likelihoods that we obtain when varying the threshold levels in the interval regressions suggests that a lower threshold fits the data better.

The big difference between the actual and the imputed frequency of the grade 1.0 can be explained by analyzing individual exams such as exam 23. The actual grade distribution in exam 23 looks rather inconspicuous, although the students taking this exam appear to

[4] In the 2004-2008 period, the grades awarded to freshman and sophomore students did not count for the graduation grade point average.

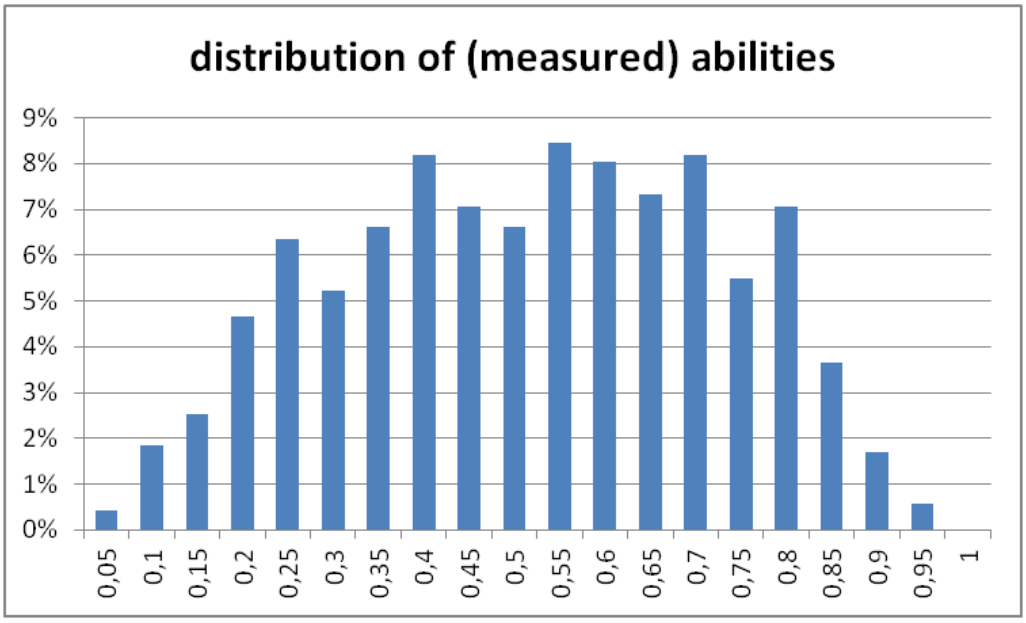


Figure 5.1: Distribution of abilities

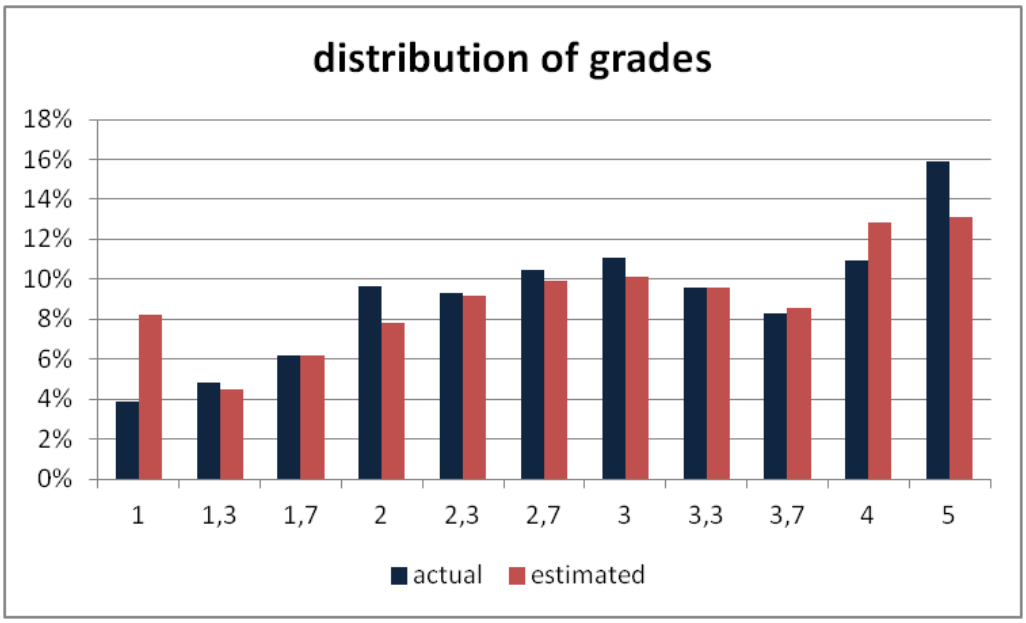


Figure 5.2: Distribution of grades

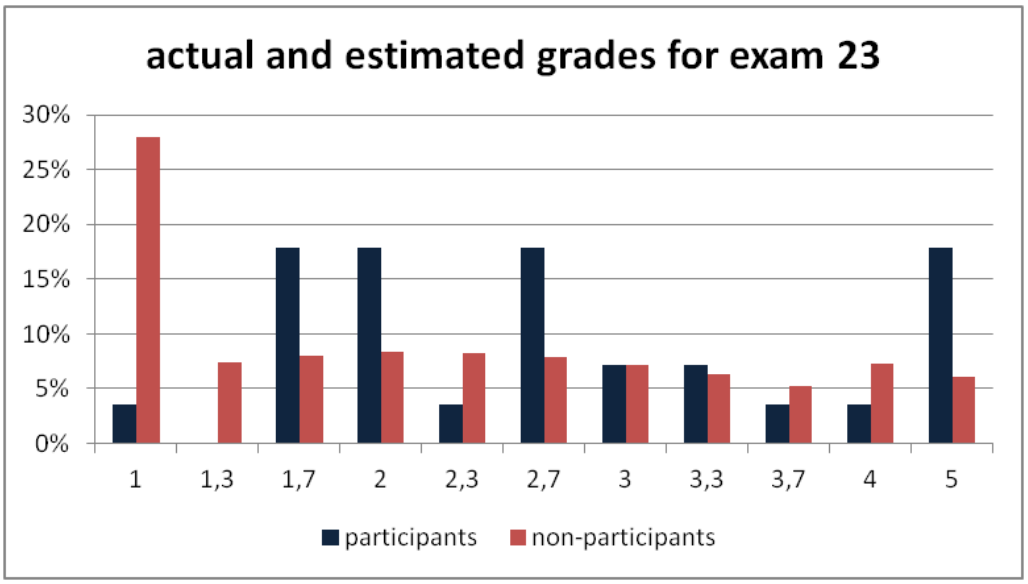


Figure 5.3: Actual vs. estimated grades for exam 23

be mostly low-ability students. Comparing the actual with the imputed grades suggests that more than one quarter of all non-participants would have obtained the top-grade 1.0 if they had taken the exam.

Comparing ECTS style grades with grades standardized in the way we suggest reveals that not controlling for student quality results in grades which are, on average, about 10% better than they should actually be.

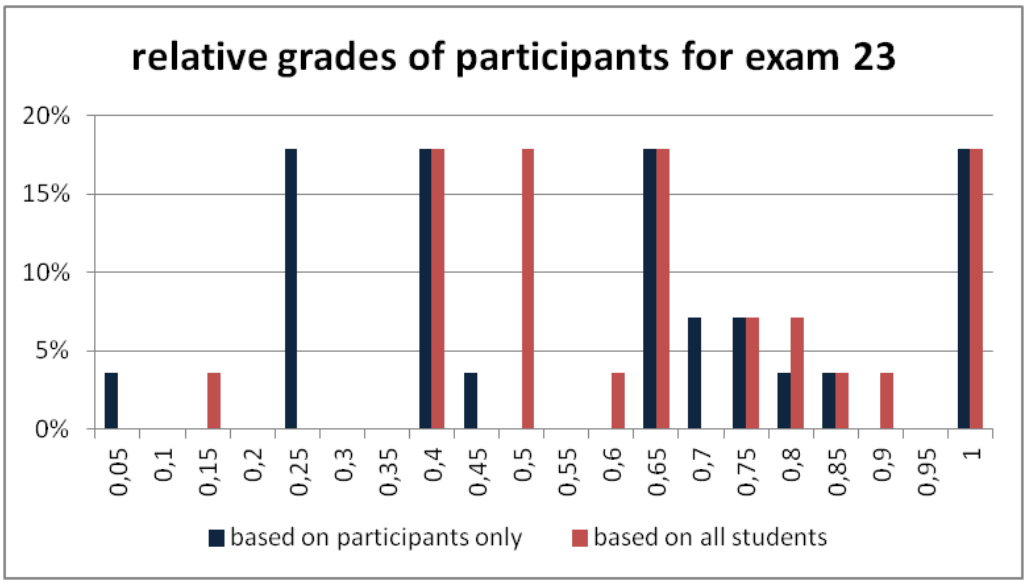


Figure 5.4: Relative grades based on participants only and on whole student body

5.4.3 Monte Carlo Simulation

In order to investigate the performance of our method we would like to compare a student's standardized GPA and her ECTS average to her true ability. In particular we would like to know if our method provides more accurate estimates of student ability than ECTS grading. Suppose that in addition to grades also student ability was known. Then one could estimate abilities based on ECTS and standardized GPA and compare the estimates to the actual abilities. Since we do not know a student's true ability this is, however, not possible.

Yet we can shed some light on the performance of our method by running a Monte Carlo Simulation, i.e. by generating random samples of grades based on a known data generating process. To this end we assume that (i) a student's true ability is equal to the one measured by standardized GPA in section 5.4.2, (ii) the true grading functions are the ones measured in section 5.4.2, and (iii) the allocation of students to exams is identical to the one in our original data. Assumption (iii) implies that a student receives a grade in the Monte Carlo Simulation if and only if she received a grade in our original data. Note that assumptions (i) to (iii) are rather soft but, at the same time, ensure that the random samples we generate closely match our real data. Moreover, when it comes to self-selection of students into courses we do not impose any artificial assumptions but just rely on the actual allocation patterns of our students. Finally, we assume that (iv) the random element of exam performance, i.e. the error term, is normally distributed.

Provided we know (i) a student's ability ai, (ii) the grading function f_j, and (iv) the distribution of the random element $\epsilon_{i,j}$ we can compute the probability of i to obtain any of the available grades in exam j. In the Monte Carlo Simulation we randomly draw a grade for each exam a student attends and we do so for every student. Thereby we create a new set of grades for which we know not only a student's grades but also her ability.

Based on the randomly generated grades we calculate ECTS averages and standardized GPAs for each random sample as described in Section 5.3. Finally, we can compare these estimates of ability with the true ability, i.e. the ability that we used in the sampling process.

To get a first impression of the performance of our method we plot the ECTS averages and standardized GPAs that we obtain from the first Monte Carlo trial against the true

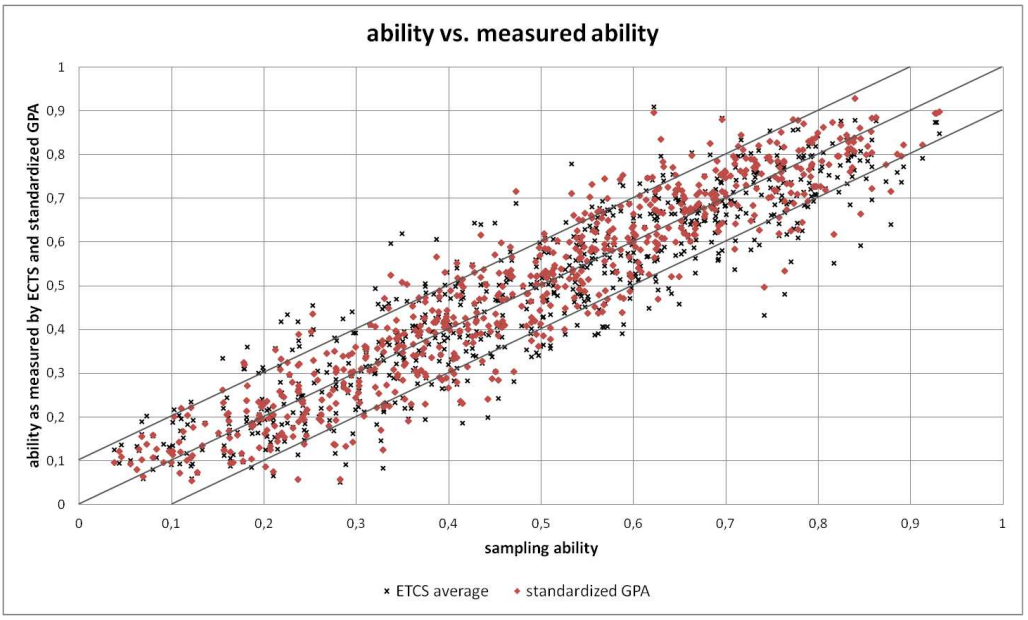


Figure 5.5: Scatter plot of actual and measured ability

ability in Figure 5.5.[5] The vertical dimension illustrates which grades student of a given ability receive. The horizontal dimension illustrates the spectrum of ability levels that are contained in grades as measured by ECTS and standardized GPA. A closer inspection of Figure 5.5 reveals that ECTS averages are more represented among outliers than standardized GPAs.

In order to gain more general insights we take a closer look at students whose true ability is close to the first, second, and third quartile. Based on all 1000 Monte Carlo trials, we compute the measurement error of ECTS averages and standardized GPAs for those students who are within a four percent interval centered at the respective quartile.

Figure 5.6 shows the measurement error for students close to the median.[6] We observe that, at the median, both measures are biased: standardized GPAs have more mass at positive measurement errors and ECTS averages more mass at negative measurement errors. The bias of standardizes GPAs[7] amounts to 1.3% more than twice the bias of ECTS averages (-0.6%). However, when looking at measurement errors equal to or larger than 0.18 in absolute terms, it can be noted that ECTS averages have more mass at the

[5]We plot only the first of the 1000 trials in order to limit the number of observations in the graph thereby keeping the informational value.

[6]This includes students with ability levels between 0.48 and 0.52. We also use a 0.04 interval for the first and third quartile, i.e. the first quartile corresponds to abilities between 0.23 and 0.27 and the third quartile corresponds to abilities between 0.73 and 0.77.

[7]We are not sure what is causing the bias of standardized GPAs and will investigate this in an updated version of this paper. One potential reason is an inappropriate choice of threshold levels in our regression model.

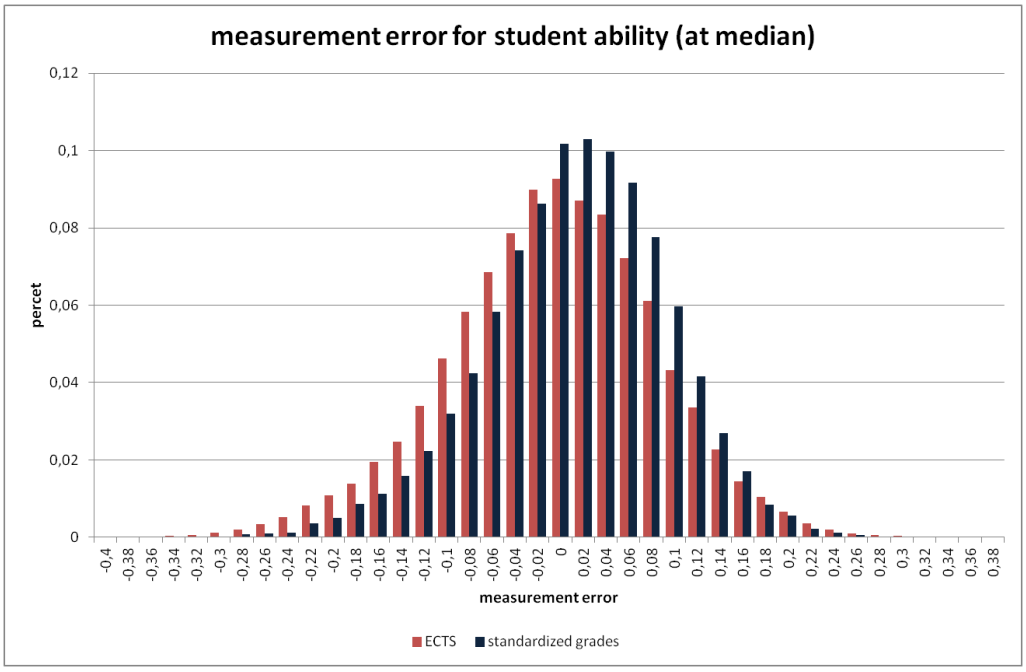


Figure 5.6: Evaluating measurement errors at the median

extremes.

When looking at Figure 5.7 we see a similar picture for the third quartile: standardized GPAs are upwards biased while ECTS averages are downwards biased and have more mass at the extremes. A markable difference is that ECTS averages exhibit a much larger bias (-3%) than our method (0.7%).

Finally, Figure 5.8 plots the measurement errors of the two methods at the first quartile. ECTS averages are biased by 1% towards median ability while standardized GPAs yield abilities that are on average 0.3% too low. Just like in the previous examples, ECTS averages have more mass at the extremes.

Having analyzed the performance of standardized GPAs and of ECTS averages at three distinct points of the ability distribution we are interested in the performance of the two methods when considering the whole continuum of abilities. Therefore we calculate the variance of both methods based on the whole sample of students and all Monte Carlo trials. We find that ECTS averages have a variance of 0.0077 while the variance of standardized GPAs amounts to 0.0054 a decrease of 29% compared to the variance of ECTS averages. From this finding we conclude that our method constitutes a significant improvement in the quality of grades being a signal of student ability.

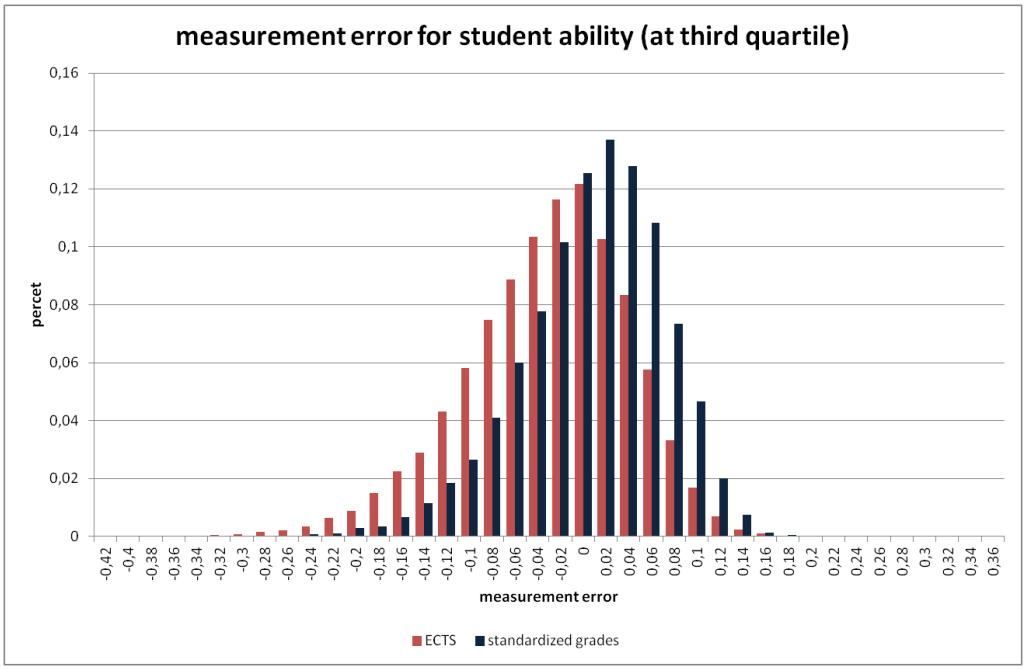


Figure 5.7: Evaluating measurement errors at the third quartile

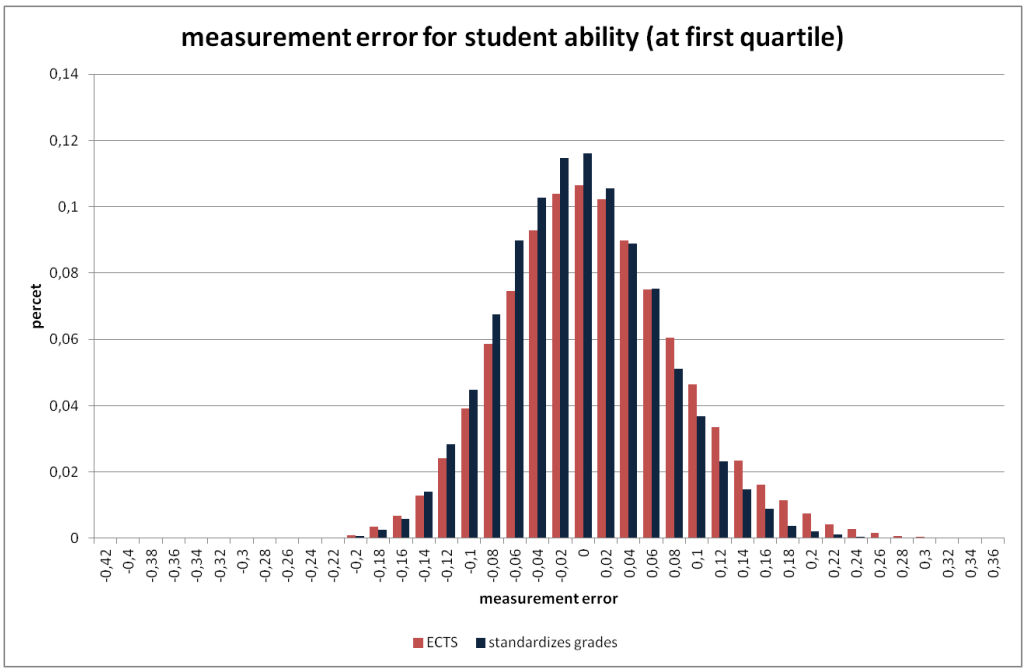


Figure 5.8: Evaluating measurement errors at the first quartile

5.4.4 Justification of the one-dimensional ability assumption

One aspect of our model is that we assume that ability is a one-dimensional concept. We thus do not differentiate between cognitive and non-congnitive skills, between mathematical, verbal, and memory-related skills. Although a one-dimensional ability concept does not reflect the current state of research applying a one-dimensional concept does not hamper our analysis if (i) all relevant skills are highly correlated, as suggested by Charles Spearman's g-factor theory, or (ii) all exams in our sample require the same composition of abilities.

To test the validity of our one-dimensionality assumption we conduct a principal component analysis on the raw grades from 13 compulsory subjects that all students have to take as freshmen and sophomores (first and second year). We find one factor explaining more than 50% of all variation. All 13 subjects load heavily on this factor and it is the only factor with an eigenvalue larger than one (see Table 5.1).[8]

Table 5.1: Principal Component Factors

Factor	Eigenvalue	Difference	Proportion	Cumulative
Factor 1	6.58350	5.69832	0.5064	0.5064
Factor 2	0.88518	0.08152	0.0681	0.5745
Factor 3	0.80366	0.08226	0.0618	0.6363
Factor 4	0.72140	0.10873	0.0555	0.6918
Factor 5	0.61267	0.05388	0.0471	0.7390
Factor 6	0.55879	0.01651	0.0430	0.7819
Factor 7	0.54228	0.09488	0.0417	0.8237
Factor 8	0.44740	0.01046	0.0344	0.8581
Factor 9	0.43694	0.02431	0.0336	0.8917
Factor 10	0.41263	0.06413	0.0317	0.9234
Factor 11	0.34850	0.00960	0.0268	0.9502
Factor 12	0.33890	0.03077	0.0261	0.9763
Factor 13	0.30813	.	0.0237	1.0000

We maintain that all other factors represent unsystematic noise because, first, the

[8] Boldt (1973) already found that the analyzed system of law school grades was essentially one factor in nature.

fraction of variation explained by the remaining factors decreases gradually without any substantial drops from 6.8% for factor 2 and 6.2% for factor 3 to 2.4% for factor 13. Second, none of the loadings of factors 2 to 13 have the same sign for the courses "econometrics", "mathematics 1", "mathematics 2", "statistics 1", and "statistics 2". 10 out of 12 factors even provide different signs of factor loadings for either "mathematics 1" and "mathematics 2", or "statistics 1" and "statistics 2", or both. The matrix of factor loadings can be found in the appendix. Given these results, we are confident that, for our purposes, a one-dimensional ability concept captures all relevant information.

5.5 Conclusion

Our method improves the quality of grades as signals of ability and thereby provides direct benefits to potential employers, managers of graduate programs, etc.

At a first glance, the effect of standardized grades on students is undetermined since standardizing grades improves some students' grades whereas it downgrades others. But there is another effect which benefits all students. If grades are not standardized in a sensible way, students are likely to choose soft courses. Students thus are tempted (and will more often than not succumb to this temptation) to trade their thirst for education against better grades. Standardizing grades eliminates such inefficiencies and allows students to select courses that are more helpful in completing their human capital.

5.6 Appendix

The figure below shows the distribution of betas. We interpret beta as the reliability of an exam: a low value of beta indicates that high ability is less associated with good grades. We were surprised to find a small number of exams for which grades and abilities are not or even negatively correlated.[9] For most of the exams, however, beta assumes reasonable levels. A beta of 10, for example, suggests that having ability 0.2 instead of 0.3 is associated with receiving the next best grade in an exam. Very high levels of beta can be obtained if all students attending an exam have similar ability but the resulting grades still use full spectrum of grades.

Figure 5.10 illustrates the distribution of alpha which can be regarded as the difficulty of an exam. An alpha of 1, for example, indicates that the best possible student (ability close to 0) would just miss the best grade for any positive value of beta. Similarly, an alpha of 2 indicates that the best student would just miss the second best grade.

Table 5.2 presents factor loadings from the factor analysis of 13 compulsory subjects for freshman and sophomores.

[9]These very low levels of beta were, however, not statistically significant. Thus the grades awarded in these exams appear to be uncorrelated with ability.

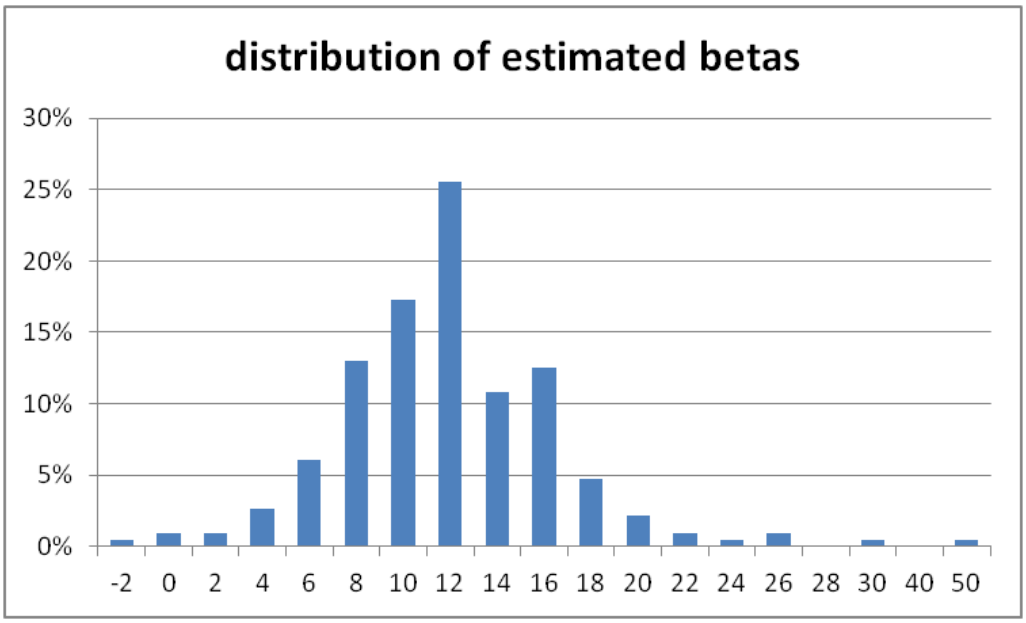


Figure 5.9: Distribution of beta hat in different exams

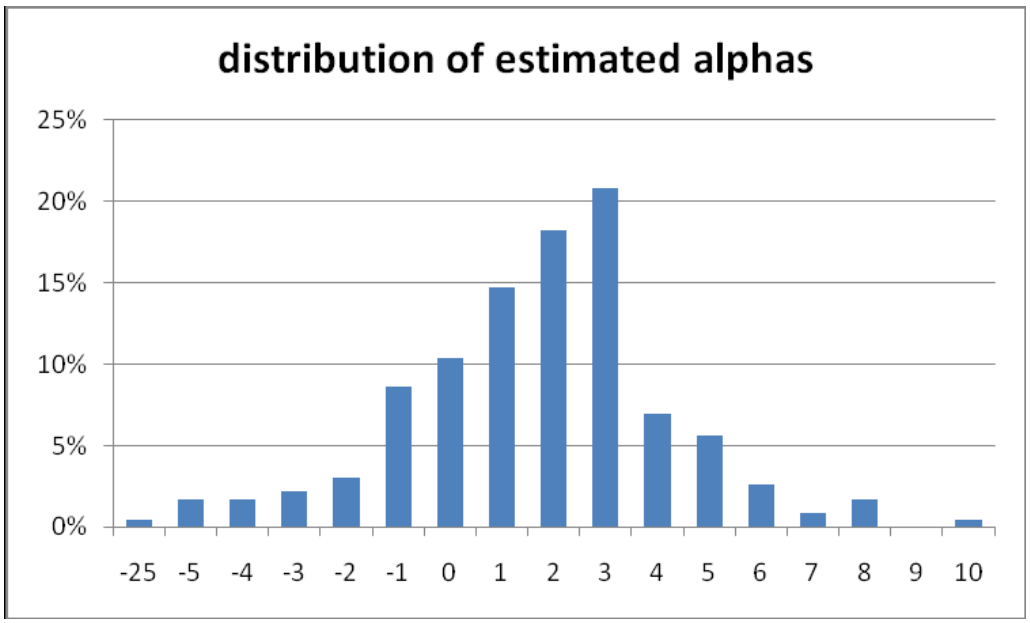


Figure 5.10: Distribution of alpha hat in different exams

Table 5.2: Matrix of factor loadings

exam	Factor1	Factor2	Factor3	Factor4	Factor5	Factor6	Factor7
Civil Law	0.6597	-0.4223	0.2246	0.1914	0.2863	-0.1257	-0.0961
Accounting	0.6673	-0.2368	0.3443	-0.2109	-0.1000	0.2672	0.4208
Business Administration I	0.7368	0.2540	-0.0182	0.1433	0.0259	0.1410	-0.3516
Business Administration II	0.7237	-0.0345	0.2417	0.0902	0.2886	0.3525	-0.1383
Econ 1 on 1	0.7253	-0.1862	0.2356	-0.1064	0.0257	-0.4267	-0.0822
Macro I	0.7532	0.0414	0.0731	-0.0324	-0.2925	0.0551	-0.1980
Micro I	0.8068	-0.0077	-0.0860	-0.0514	-0.2049	-0.2658	0.0551
Micro II	0.6780	0.4536	0.1486	-0.3437	-0.0384	0.0141	-0.1031
Econometrics	0.6386	0.5020	0.1504	0.2860	0.1386	-0.2059	0.3119
Statistics I	0.7441	-0.2406	-0.1465	-0.0262	-0.3728	0.0667	-0.0572
Statistics II	0.6949	0.0109	-0.2527	0.4977	-0.1650	0.1271	0.1766
Mathematics I	0.7107	-0.1681	-0.4783	-0.0320	0.2443	-0.0553	-0.0083
Mathematics II	0.6949	0.0680	-0.3859	-0.3837	0.2489	0.0739	0.1476

exam	Factor8	Factor9	Factor10	Factor11	Factor12	Factor13
Civil Law	0.4172	0.0613	0.0399	0.0299	-0.1099	-0.0050
Accounting	-0.0596	0.0311	0.2613	0.0058	0.0433	-0.0499
Business Administration I	-0.1330	0.2219	0.3769	-0.0778	-0.1087	0.0273
Business Administration II	-0.2163	-0.0501	-0.3417	-0.0851	0.0393	-0.0840
Econ 1 on 1	-0.3502	-0.0366	-0.0070	0.1782	0.0061	0.1649
Macro I	0.1289	-0.5149	0.0605	-0.0909	-0.0041	0.0549
Micro I	-0.0543	0.0762	-0.0940	-0.1782	-0.1761	-0.3780
Micro II	0.2198	0.1182	-0.0932	0.2967	0.1191	-0.0939
Econometrics	0.0656	-0.0096	-0.0194	-0.2051	0.1123	0.1321
Statistics I	0.0849	0.2875	-0.1925	-0.1411	0.1466	0.2264
Statistics II	-0.0462	-0.0401	-0.0548	0.3338	-0.1120	-0.0098
Mathematics I	-0.0248	-0.0861	0.1390	0.0042	0.3632	-0.1212
Mathematics II	0.0222	-0.0495	-0.0542	-0.0315	-0.2963	0.1801

Bibliography

Achen, A. C., and P. N. Courant (2009): "What Are Grades Made Of?," *Journal of Economic Perspectives*, 23(3), 77–92.

Bar, T., V. Kadiyali, and A. Zussman (2009): "Grade Information and Grade Inflation: The Cornell Experiment," *Journal of Economic Perspectives*, 23(3), 93–108.

Bar, T., and A. Zussman (forthcoming): "Partisan Grading," *American Economic Journal - Applied Economics*.

Boldt, R. F. (1973): "Factor Analysis of Law School Grades," Research bulletin, Educational Testing Service, Princeton.

Caulkins, J. P., P. D. Larkey, and J. Wei (1996): "Adjusting GPA to Reflect Course Difficulty," Working papers, Heinz School of Public Policy and Management, Carnegie Mellon University.

Dubey, P., and J. Geanakoplos (2005): "Grading in Games of Status: Marking Exams and Setting Wages," Cowles Foundation Discussion Papers 1544, Cowles Foundation for Research in Economics, Yale University.

Johnson, V. E. (1997): "An Alternative to Traditional GPA for Evaluating Student Performance," *Statistical Science*, 12(4), 251–69.

Sabot, R., and J. Wakeman-Linn (1991): "Grade Inflation and Course Choice," *Journal of Economic Perspectives*, 5(1), 159–70.

Spearman, C. (1904): ""General intelligence" objectively determined and measured," *American Journal of Psychology*, 15, 201–93.

Young, J. W. (1990): "Adjusting the Cumulative GPA Using Item Response Theory," *Journal of Educational Measurement*, 27(2), 175–86.

General Bibliography

Achen, A. C., and P. N. Courant (2009): "What Are Grades Made Of?," *Journal of Economic Perspectives*, 23(3), 77–92.

Arora, A., P. David, and A. Gambardella (1998): "Reputation and Competence in Publicly Funded Science: Estimating the Effects on Research Group Productivity," *Annales d'Economie et de Statistique*, 49-50, 163–198.

Arrow, K. J., B. D. Bernheim, M. S. Feldstein, D. L. McFadden, J. M. Poterba, and R. M. Solow (2011): "100 Years of the *American Economic Review*: The Top 20 Articles," *American Economic Review*, 101(1), 1–8.

Bar, T., V. Kadiyali, and A. Zussman (2009): "Grade Information and Grade Inflation: The Cornell Experiment," *Journal of Economic Perspectives*, 23(3), 93–108.

Bar, T., and A. Zussman (forthcoming): "Partisan Grading," *American Economic Journal - Applied Economics*.

Berghoff, S., G. Federkeil, P. Giebisch, C.-D. Hachmeister, M. Hennings, and D. Müller-Böling (2006): "Das CHE ForschungsRanking deutscher Universitäten 2006," Working Papers 79, Centrum für Hochschulentwicklung, Gütersloh.

BMFSFJ (2005): *Gender Datenreport 2005*. Bundesministerium für Familie, Senioren, Frauen und Jugend, section 1.7.

Boldt, R. F. (1973): "Factor Analysis of Law School Grades," Research bulletin, Educational Testing Service, Princeton.

Bornmann, L., and H.-D. Daniel (2008): "The Effectiveness of the Peer Review Process: Inter-Referee Agreement and Predictive Validity of Manuscript Refereeing at Angewandte Chemie," *Angewandte Chemie International Edition*, 47(38), 7173–7178.

Bräuninger, M., and J. Haucap (2001): "Was Ökonomen lesen und schätzen: Ergebnisse einer Umfrage," *Perspektiven der Wirtschaftspolitik*, 2(2), 185–210.

Cainelli, G., A. de Felice, M. Lamonarca, and R. Zoboli (2006): "The Publications of Italian Economists in ECONLIT. Quantitative Assessment and Implications for Research Evaluation," *Economia Politica*, 0(3), 385–423.

Campbell, P. (2008): "Escape from the impact factor," *Ethics in Science and Environmental Politics*, 8, 5–7.

Caulkins, J. P., P. D. Larkey, and J. Wei (1996): "Adjusting GPA to Reflect Course Difficulty," Working papers, Heinz School of Public Policy and Management, Carnegie Mellon University.

Clemenz, G., and K. Neusser (1991): "International impact of Austrian economists," *Empirica*, 18, 201–219, 10.1007/BF00924912.

Combes, P.-P., and L. Linnemer (2001): "La publication d'articles de recherche en économie en France," *Annales d'Economie et de Statistique*, 62, 5–47.

——— (2003): "Where are the Economists Who Publish? Publication Concentration and Rankings in Europe Based on Cumulative Publications," *Journal of the European Economic Association*, 1(6), 1250–1308.

Coupé, T. (2010): "Peer Review versus Citations - An Analysis of Best Paper Prizes," Discussion Papers 35, Kyiv School of Economics.

Davies, J. B., M. G. Kocher, and M. Sutter (2008): "Economics research in Canada: a long-run assessment of journal publications," *Canadian Journal of Economics*, 41(1), 22–45.

De Long, J. B., and K. Lang (1992): "Are All Economic Hypotheses False?," *Journal of Political Economy*, 100(6), 1257–72.

Dolado, J. J., A. García-Romero, and G. Zamarro (2003): "Publishing performance in economics: Spanish rankings (1990-1999)," *Spanish Economic Review*, 5, 85–100.

Dubey, P., and J. Geanakoplos (2005): "Grading in Games of Status: Marking Exams and Setting Wages," Cowles Foundation Discussion Papers 1544, Cowles Foundation for Research in Economics, Yale University.

Eichenberger, R., U. Meier, and R. Arpagaus (2000): "Ökonomen, Publikationen und Zitationen: Ein europäischer Vergleich," *Perspektiven der Wirtschaftspolitik*, 1(2), 143–160.

Ellison, G. (2002): "Evolving Standards for Academic Publishing: A q-r Theory," *Journal of Political Economy*, 110(5), 994–1034.

Fabel, O., and F. Hesse (1999): "Befragungsstudie vs. Publikationsanalyse: Zur Interpretation von Ranglisten der Forschungsaktivitäten deutscher betriebswirtschaftlicher Fachbereiche," *Die Betriebswirtschaft*, 59, 196–204.

Frey, B. S. (2007): "Evaluierungen, Evaluierungen H Evaluitis," *Perspektiven der Wirtschaftspolitik*, 8(3), 207–220.

Gans, J. S., and G. B. Shepherd (1994): "How Are the Mighty Fallen: Rejected Classic Articles by Leading Economists," *Journal of Economic Perspectives*, 8(1), 165–79.

Garfield, E. (1972): "Citation analysis as a tool in journal evaluation," *Science*, 178(4060), 471–479.

Glasser, O. (1995): *Wilhelm Conrad Röntgen und die Geschichte der Röntgenstrahlen*. Springer, Berlin, 3rd edn.

Gourieroux, C., A. Monfort, and A. Trognon (1984): "Pseudo Maximum Likelihood Methods: Applications to Poisson Models," *Econometrica*, 52(3), 701–20.

Guimarães, P. (2002): "The State of Portuguese Research in Economics: An Analysis based on Publications in International Journals," *Portuguese Economic Journal*, 1, 3–25.

Hamermesh, D. S. (1994): "Facts and Myths about Refereeing," *Journal of Economic Perspectives*, 8(1), 153–63.

Hamermesh, D. S., and G. A. Pfann (2011): "Markets for Reputation: Evidence on Quality and Quantity in Academe," *Economic Inquiry*, forthcoming.

Hamermesh, D. S., and P. Schmidt (2003): "The Determinants of Econometric Society Fellows Elections," *Econometrica*, 71(1), 399–407.

Hein, M. (2006): "How Did Economic Research in Switzerland Develop since the Beginning 90s?," *Swiss Journal of Economics and Statistics (SJES)*, 142(II), 285–306.

Hodgson, G. M., and H. Rothman (1999): "The Editors and Authors of Economics Journals: A Case of Institutional Oligopoly?," *Economic Journal*, 109(453), F165–86.

Jacob, B., and L. Lefgren (2007): "The Impact of Research Grant Funding on Scientific Productivity," NBER Working Papers 13519, National Bureau of Economic Research, Inc.

Johnson, V. E. (1997): "An Alternative to Traditional GPA for Evaluating Student Performance," *Statistical Science*, 12(4), 251–69.

Jones, B. F. (2009): "The Burden of Knowledge and the 'Death of the Renaissance Man': Is Innovation Getting Harder?," *Review of Economic Studies*, 76(1), 283–317.

Kalaitzidakis, P., T. P. Mamuneas, and T. Stengos (2003): "Rankings of Academic Journals and Institutions in Economics," *Journal of the European Economic Association*, 1(6), 1346–1366.

Kocher, M. G., and M. Sutter (2001): "The Institutional Concentration of Authors in Top Journals of Economics during the Last Two Decades," *Economic Journal*, 111(472), F405–21.

Kóczy, L. A., and M. Strobel (2010): "The World Cup of Economics Journals: A Ranking by a Tournament Method," Working Paper Series 1011, Óbuda University, Keleti Faculty of Economics.

Laband, D. N., and M. J. Piette (1994): "Favoritism versus Search for Good Papers: Empirical Evidence Regarding the Behavior of Journal Editors," *Journal of Political Economy*, 102(1), 194–203.

Laband, D. N., and R. D. Tollison (2000): "Intellectual Collaboration," *Journal of Political Economy*, 108(3), 632–661.

——— (2003): "Dry Holes in Economic Research," *Kyklos*, 56(2), 161–173.

Liebowitz, S. J., and J. P. Palmer (1984): "Assessing the Relative Impacts of Economic Journals," *Journal of Economic Literature*, 22(1), 77–88.

Maske, K. L., G. C. Durden, and P. E. Gaynor (2003): "Determinants of Scholarly Productivity among Male and Female Economists," *Economic Inquiry*, 41(4), 555–564.

McAfee, R. P. (2010): "Edifying Editing," *The American Economist*, 55(1), 1–8.

Nash, J. F. (1950): "Equilibrium Points in N-Person Games," *Proceedings of the National Academy of Sciences*, 36, 48–49.

Neary, J. P., J. A. Mirrlees, and J. Tirole (2003): "Evaluating Economics Research in Europe: An Introduction," *Journal of the European Economic Association*, 1(6), 1239–1249.

Nederhof, A. J., and R. F. Meijer (1995): "Development of bibliometric indicators for utility of research to users in society: Measurement of external knowledge transfer via publications in trade journals," *Scientometrics*, 32(1), 37–48.

Neri, F., and J. R. Rodgers (2006): "Ranking Australian Economics Departments by Research Productivity," *The Economic Record*, 82(s1), S74–S84.

Opthof, T., F. Furstner, M. van Geer, and R. Coronel (2000): "Regrets or no regrets? No regrets! The fate of rejected manuscripts," *Cardiovascular Research*, 45(1), 255–258.

Oswald, A. J. (2007): "An Examination of the Reliability of Prestigious Scholarly Journals: Evidence and Implications for Decision-Makers," *Economica*, 74(293), 21–31.

Palacios-Huerta, I., and O. Volij (2004): "The Measurement of Intellectual Influence," *Econometrica*, 72(3), 963–977.

Pinski, G., and F. Narin (1976): "Citation influence for journal aggregates of scientific publications: Theory, with application to the literature of physics," *Information Processing & Management*, 12(5), 297 – 312.

Rauber, M., and H. W. Ursprung (2008a): "Evaluation of Researchers: A Life Cycle Analysis of German Academic Economists," in *Conferences on New Political Economy*, ed. by M. Albert, S. Voigt, and D. Schmidtchen, vol. 25 of *Conferences on New Political Economy*, pp. 101–122(2). Mohr Siebeck, Tübingen.

——— (2008b): "Life Cycle and Cohort Productivity in Economic Research: The Case of Germany," *German Economic Review*, 9, 431–456.

Ravallion, M., and A. Wagstaff (2010): "On measuring scientific influence," Policy Research Working Paper Series 5375, The World Bank.

Ritzberger, K. (2008a): "Eine invariante Bewertung wirtschaftswissenschaftlicher Fachzeitschriften," *Perspektiven der Wirtschaftspolitik*, 9(3), 267–285.

——— (2008b): "A Ranking of Journals in Economics and Related Fields," *German Economic Review*, 9, 402–430.

Rodríguez, D. (2006): "Publishing Performance of Spanish Academics: 1970-2004," *Spanish Economic Review*, 8, 255–270.

Romer, P. M. (1990): "Endogenous Technological Change," *Journal of Political Economy*, 98(5), S71–102.

Sabot, R., and J. Wakeman-Linn (1991): "Grade Inflation and Course Choice," *Journal of Economic Perspectives*, 5(1), 159–70.

Santos Silva, J. M. C., and S. Tenreyro (2006): "The Log of Gravity," *The Review of Economics and Statistics*, 88(4), 641–658.

Schulze, G. G., S. Warning, and C. Wiermann (2008): "Zeitschriftenrankings für die Wirtschaftswissenschaften - Konstruktion eines umfassenden Metaindexes," *Perspektiven der Wirtschaftspolitik*, 9(3), 286–305.

Serrano, R. (2004): "The Measurement of Intellectual Influence: the Views of a Sceptic," *Economics Bulletin*, 1(3), 1–6.

Shanley, P., and C. López (2009): "Out of the Loop: Why Research Rarely Reaches Policy Makers and the Public and What Can be Done," *Biotropica*, 41, 535–544.

Spearman, C. (1904): ""General intelligence" objectively determined and measured," *American Journal of Psychology*, 15, 201–93.

Stigler, G. J., and C. Friedland (1975): "The Citation Practices of Doctorates in Economics," *Journal of Political Economy*, 83(3), 477–507.

Sunderland, T., J. Sunderland-Groves, P. Shanley, and B. Campbell (2009): "Bridging the Gap: How Can Information Access and Exchange Between Conservation Biologists and Field Practitioners be Improved for Better Conservation Outcomes?," *Biotropica*, 41, 549–554.

Taylor, S. W., B. F. Fender, and K. G. Burke (2006): "Unraveling the Academic Productivity of Economists: The Opportunity Costs of Teaching and Service," *Southern Economic Journal*, 72(4), 846–859.

Turnovec, F. (2007): "Publication Portfolio of the Czech Economists and Problems of Rankings," *Ekonomický Casopis/Journal of Economics*, 55, 623–645.

Ursprung, H. W. (2003): "Schneewittchen im Land der Klapperschlangen: Evaluation eines Evaluators," *Perspektiven der Wirtschaftspolitik*, 4(2), 177–190.

Ursprung, H. W., and M. Zimmer (2007): "Who is the "Platz-Hirsch" of the German Economics Profession? A Citation Analysis," *Journal of Economics and Statistics (Jahrbuecher fuer Nationaloekonomie und Statistik)*, 227(2), 187–208.

Vandermeulen, A. (1972): "Manuscripts in the maelstrom: A theory of the editorial process," *Public Choice*, 13(1), 107–111.

Wall, H. J. (2009): "Don't Get Skewed Over by Journal Rankings," *The B.E. Journal of Economic Analysis & Policy*, 9(1), 34.

Wilson, J. D. (1978): "Peer Review and Publication. Presidential address before the 70th annual meeting of the American Society for Clinical Investigation, San Francisco, California, 30 April 1978.," *Journal of Clinical Investigation*, 61(6), 1697–1701.

Wooldridge, J. M. (2007): *Econometric Analysis of Cross Section and Panel Data*. MIT Press, Cambridge, MA.

Young, J. W. (1990): "Adjusting the Cumulative GPA Using Item Response Theory," *Journal of Educational Measurement*, 27(2), 175–86.

Zimmermann, C. (2007): "Academic Rankings with RePEc," Working papers 2007-36, University of Connecticut, Department of Economics.

Zusammenfassung

Die vorliegende Dissertation ist eine Sammlung von fünf Forschungspapieren die während meines Promotionsstudiums an der Universität Konstanz von Oktober 2007 bis Dezember 2010 entstanden. Die ersten vier Studien beschäftigen sich mit bibliometischen Themen wohingegen die letzte Studie von der Benotung von Studenten handelt. Alle fünf Studien enthalten empirische Analysen, allerdings ist sind das dritte und fünfte Papier überwiegend theoretischer Ausrichtung. Der erste Artikel untersucht unterschiedliche methodische Aspekte des *Handelsblatt Rankings* aus dem Jahr 2007. Der zweite Artikel bewertet die Forschungsproduktivität im Bereich Betriebswirtschaftslehre von Universitäten in Deutschland, Österreich und der deutschsprachigen Schweiz. Der dritte Artikel präsentiert eine neue Methode zur Bewertung von Forschungsergebnissen, welche eine gemeinsame Grundlage zur Bewertung von Forschung aus unterschiedlichen Bereichen der Wissenschaft darstellt. Der vierte Artikel untersucht Determinanten der Herausgabe wissenschaftlicher Zeitschriften. Das fünfte Papier präsentiert schließlich eine Methode die Aussagekraft von Noten als Schätzer für die Fähigkeit von Studenten zu verbessern. Das Papier enthält auch ein Fallbeispiel. Im Folgenden fasse ich die Hauptergebnisse kurz zusammen:

Bei Kapitel 1 handelt es sich um den Nachdruck eines gemeinsamen Artikels mit Prof. Dr. Heinrich Urpsrung von der Universität Konstanz. Der Artikel *Das Handelsblatt Ökonomen-Ranking 2007: Eine kritische Beurteilung* wurde in den *Perspektiven der Wirtschaftspolitik*, Vol. 9, No. 3, pp. 254-266, 2008 veröffentlicht. Wir untersuchen die Methode des Handelsblatt Rankings das gegenwärtig das anerkannteste Ranking akademischer Forschung im Bereich Volkswirtschaftslehre in Deutschland, Österreich, und der deutschsprachigen Schweiz darstellt und jährlich veröffentlicht wird. Gerade weil es oft als Forschungsindikator genutzt wird muss das Handelsblatt Ranking anreizkompatibel sein. Wir argumentieren, dass (i) die Auswahl einer kleinen Teilmenge an Zeitschriften

anstelle einer Vollerhebung, (ii) die Gewährung eines Bonus für Koautorenschaft, und (iii) das Vernachlässigen der Artikellänge Fehlanreize im Forschungssektor schafft und die Forschungsleistung nicht richtig widergibt. Wir analysieren dies für jeden der genannten Aspekte separat und zeigen, dass das Ranking einiger Forscher maßgeblich von der verwendeten Methode abhängt. Weiter zeigen wir, dass diese Unterschiede auch auf der aggregierten Ebene von Fachbereichen zu finden sind. Zu Guter Letzt plädieren wir dafür, dass das Handelsblatt Ranking sich stärker mit der Forschungsleistung pro Professor als mit dem Gesamtoutput eines Fachbereichs beschäftigen sollte, da letzterer stark von der Fachbereichsgröße abhängt und nicht notwendigerweise mit hoher Durchschnittsproduktivität zusammenhängt.

Kapitel 2 ist eine gemeinsame Arbeit mit Prof. Dr. Oliver Fable (Universität Wien) und Miriam Henseler, geb. Hein (DFG). Unser Papier *Research Productivity in Business Economics: An Investigation of Austrian, German and Swiss Universities* wurde 2008 in der *German Economic Review*, Vol. 9, No. 4, pp. 506-531 veröffentlicht. Wir benutzen einen neuen, umfassenden Datensatz, in dem der Forschungsoutput von Betriebswirtinnen und Betriebswirten erfasst ist, die im Frühjahr des Jahres 2008 an einer Universität in Deutschland, Österreich oder der Schweiz beschäftigt waren. Neben den Publikationsdaten enthält der Datensatz auch persönliche Angaben zu etwa 1.800 Forscherinnen und Forschern. Wir berechnen Forschungsrankings der Fachbereiche und erstellen Listen der fünf besten Fachbereiche in einigen ausgewählten Teilbereichen der Betriebswirtschaftlehre. Insgesamt sind die Produktivitätsunterschiede zwischen den Universitäten gering. Auch die Konzentration des Forschungsoutputs zwischen den verschiedenen Universitäten ist niedrig. Mit Hilfe von Tobit und Hurdle Regressionen untersuchen wir außerdem, inwieweit institutionelle Faktoren und individuelle Charakteristika die Forschungsproduktivität beeinflussen. Die Größe des Fachbereichs - gemessen an der Zahl der Fachbereichsmitglieder - und die Anzahl der Professoren, die publizieren, haben einen signifikanten Einfluss auf die Forschungsproduktivität. Die Forschungsproduktivität ist zudem höher in Fachbereichen, in denen auch der Studiengang Volkswirtschaftslehre angeboten wird. Entsprechend der Lebenszyklus-Hypothese nimmt die Forschungsproduktivität von publizierenden Forschern mit zunehmendem Karrierealter ab. Betriebswirtinnen scheinen weniger produktiv zu sein als ihre männlichen Kollegen. Es sollte erwähnt werden, dass das Papier nennenswerte Aufmerksamkeit in der wissenschaftlichen Gemeinschaft erfahren hat und zu zwei Kom-

mentare führte. In zwei Antworten zeigen wir, dass unsere Ergebnisse bezüglich der in den Kommentaren hervorgebrachten Kritikpunkte robust sind.

Kapitel 3 entwickelt eine neue Methode Forschung zu bewerten und ist als Arbeitspapier der Universität Konstanz verfügbar. In diesem Papier nehme ich den Standpunkt ein, dass das eigentliche Ziel von Forschung nicht intellektuelle Erkenntnis sondern der Beitrag zum Wohlergehen der Menschheit sei. Daher schlage ich eine generationsbasierten Ansatz zur Bewertung von Forschung vor. Auf dem Fluss wissenschaftlicher Ergebnisse basierend, entwickle ich einen Wertschöpfungs (VA) Index der es im Prinzip erlaubt jedem Forschungsergebnis und, durch Aggregation, ganzen Wissenschaftsgebieten einen Geldwert zuzuweisen. Die grundsätzliche Idee des VA-Index ist es den Wert aller Anwendungen die Forschungsergebnisse enthalten auf die zugrundeliegenden Forschungsarbeiten zu verteilen und davon ausgehend weiter zu Forschungsarbeiten früherer Generationen auf die die Autoren der unmittelbar zugrundeliegenden Arbeiten direkt oder indirekt aufbauen. Dadurch wird einem Forschungsergebnis der Wert seines finalen Beitrags zum Nutzen der Menschheit zugewiesen. Der wesentliche Beitrag des VA-Index ist es ein Maß für den Wert von Forschung bereitzustellen, das über wissenschaftliche Disziplinen hinweg vergleichbar ist. Um zu zeigen wie dieser generationenbasierte Ansatz funktioniert, präsentiere ich ein VX-basiertes Zeitschriftenranking und ein Ranking der einflussreichsten Artikel für den Bereich der Wirtschaftswissenschaft.

Kapitel 4 ist ein gemeinsam mit Matthias Krapf (Universität Wien) geschriebener Artikel. Der Artikel *How Do Editors Select Papers, and How Good are They at Doing It?* erschien 2011 in *The B.E. Journal of Economic Analysis & Policy*, Vol. 11, Iss. 1 (Topics), Article 64. Wir benutzen Daten über die B.E. Journals welche Artikel in vier Qualitätsklassen unterscheiden um die Genauigkeit der Forschungsevaluation im Bereich Wirtschaftswissenschaften zu untersuchen. Wir zeigen, dass Einreichungen von Autoren mit starkem Forschungsprofil und von Autoren die einer gut gerankten Institution angehören mit signifikant größerer Wahrscheinlichkeit in höheren Qualitätsklassen publiziert werden. Zitationserfolg, gemessen nach RePEc Statistiken, hängt ebenfalls stark vom Forschungsprofil der Autoren ab. Weiterhin finden wir, dass Frauen signifikant mehr Zitate bekommen, wenn man für das Forschungsgebiet gemäß JEL codes kontrolliert. Schließlich, und am wesentlichsten, messen wir wie erfolgreich die Herausgeber der B.E. Journals und ihre Gutachter dabei sind Artikel in Qualitätsklassen einzuordnen. Während

sie im Schnitt bessere von schlechteren Arbeiten unterscheiden können, beobachten wir auch dass viele Einstufungen nicht mit der Ansicht vereinbar sind, dass Zitate die Qualität einer Forschungsarbeit widerspiegeln.

Kapitel 5 ist die aktuelle Fassung eines Papiers mit Prof. Dr. Heinrich Ursprung (Universität Konstanz) in welchem wir eine neue Methode zur Standardisierung von Noten entwickeln wenn nicht alle Studenten dieselben Prüfungen besuchen. Unsere Methode nutzt Regressionsanalyse um für Unterschiede in der Schwierigkeit einzelner Prüfungen zu kontrollieren und so zu ausgewogeneren Noten zu kommen. Unser Ansatz beruht dabei auf der Idee dass der Unterschied zwischen der Fähigkeit eines Studenten und seiner Note die Schwierigkeit der Prüfung reflektiert. Wir schätzen die Fähigkeit eines anhand eines Gleichungssystems, d.h. wir leiten Fähigkeiten endogen ab. Ein Kernaspekt unserer Methode im Vergleich zu anderen Methoden ist, dass wir nicht nur einen standardisierten Notendurchschnitt berechnen sondern standardisierte Noten für die einzelnen Prüfungen erstellen. Auf Grundlage echter Prüfungsdaten zeigen wir, dass unser Ansatz zu besseren Ergebnissen führt als ECTS Noten. Ferner zeigen wir, dass die Benotungsmaßstäbe unterschiedlicher Prüfungen, und wohl auch der zugrundeliegenden Fächer, signifikant variieren. Sollten Noten nicht sinnvoll standardisiert werden, ist es wahrscheinlich dass Studenten zu leichte Fächer besuchen. Das Standardisieren von Noten schafft derartige Verzerrungen ab und führt dadurch zu einer effizienteren Wahl der Fächer durch die Studenten.

Danksagung

Zurückblickend möchte ich mich bei einigen Menschen bedanken, die mich und diese Arbeit in den letzten Jahren begleitet und unterstützt haben.

bei meinem Doktorvater Heinrich Ursprung für sein Verständnis, insbesondere wenn ich Dinge nicht ausreichend erklärt habe, für seine Liebe zum Schreiben, seine Denkanstöße und die gemeinsame Arbeit und Zeit.

bei Oliver Fabel für seine offene Art, Wien, unsere Zusammenarbeit, und die Begutachtung dieser Arbeit.

bei Miriam Henseler für das gemeinsame Papier und für ihre Doktorarbeit die hier weitgehend als Formatvorlage diente.

bei Olaf Storbeck, Ulrich Wacker und den HiWis am Lehrstuhl Ursprung für die Arbeit an den Handelsblatt Rankings.

bei Michael Rauber ebenfalls für die Arbeit am Ranking, London und ein Zimmer mit Ausblick als die Alpen noch nicht verbaut waren.

bei Matthias Krapf für etliche Diskussionen, gelegentliche Nachhilfe in Ökonometrie, Looping Louie und unser Papier.

bei Simeon Schudy für viel(e) Unterhaltung(en) und kritisches Denken.

bei George Emen und (Wolfgang) Nik(olaus) Probst für viel Unsinn, Rexle und Herrn Blache und für handfeste Arbeit.

bei meinen Eltern und Großeltern für ihre Unterstützung. Insbesondere bei meiner Mutter Siggi für ihr Verständnis.

bei Tina, dafür dass sie für mich da ist.

Erklärung

Ich versichere hiermit, dass ich die vorliegende Arbeit mit dem Thema:

Five Essays on the Quantification and Measurement of Intellectual Achievements

ohne unzulässige Hilfe Dritter und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe. Die aus anderen Quellen direkt oder indirekt übernommenen Daten und Konzepte sind unter Angabe der Quelle gekennzeichnet. Weitere Personen, insbesondere Promotionsberater, waren an der inhaltlich materiellen Erstellung dieser Arbeit nicht beteiligt.[1] Die Arbeit wurde bisher weder im In- noch im Ausland in gleicher oder ähnlicher Form einer anderen Prüfungsbehörde vorgelegt.

Konstanz, den 05.12.2011

(Robert Hofmeister)

[1] Siehe hierzu die Abgrenzung auf der folgenden Seite.

Abgrenzung

Kapitel 1 entstammt einer gemeinsamen Arbeit mit Herrn Prof. Heinrich Ursprung (Universität Konstanz). Idee und Text stammen von Herrn Urspung, ich habe die Datenauswertung beigesteuert.

Kapitel 2 entstammt einer gemeinsamen Arbeit mit Herrn Prof. Oliver Fabel (Universität Wien) und Frau Dr. Miriam Henseler, née Hein (DFG). Die Idee stammt von Herrn Fabel. Frau Henseler hat die Regressionsanalysen in den Abschnitten 2.4 und 2.5 durchgeführt. Ich habe die Daten aufbereitet und die Rankings in Abschnitt 2.3 erstellt. Die Ergebnisse wurden von allen Autoren gemeinsam interpretiert.

Ich versichere hiermit, dass ich Kapitel 3 der vorliegenden Arbeit ohne Hilfe Dritter und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe.

Kapitel 4 entstammt einer gemeinsamen Arbeit mit Herrn Matthias Krapf (Universität Wien). Herr Krapf und ich haben die Idee gemeinsam entwickelt und die Ergebnisse gemeinsam interpretiert.

Kapitel 5 entstammt einer gemeinsamen Arbeit mit Herrn Prof. Heinrich Ursprung (Universität Konstanz). Die Idee stammt von Herrn Ursprung, die Methode haben wir gemeinsam entwickelt, wobei mein Anteil vermutlich etwas überwiegt. Die Resultate wurden gemeinsam interpretiert, die Programmierung stammt von mir.

Konstanz, den 05.12.2011

_____

(Robert Hofmeister)