



University of Freiburg
Department of International Economic Policy
Discussion Paper Series
Nr. 3

Zeitschriftenrankings für die Wirtschaftswissenschaften

Konstruktion eines umfassenden Metaindexes

Günther G. Schulze, Susanne Warning und
Christian Wiermann

March 2008

ISSN 1866-4113

University of Freiburg
Department of International Economic Policy
Discussion Paper Series

The Discussion Papers are edited by:
Department of International Economic Policy
Institute for Economic Research
University of Freiburg
D-79085 Freiburg, Germany
Platz der Alten Synagoge 1

Tel: +49 761 203 2342
Fax: +49 761 203 2414
Email: iep@vwl.uni-freiburg.de

Editor:
Prof. Dr. Günther G. Schulze

ISSN: 1866-4113
Electronically published: 27.3.2008

Zeitschriftenrankings für die Wirtschaftswissenschaften – Konstruktion eines umfassenden Metaindexes*

Günther G. Schulze^a

Susanne Warning^b

Christian Wiermann^c

Universität Freiburg

IAAEG Trier

Roland Berger Strategy
Consultants GmbH

Version: 26. März 2008

Zusammenfassung in Englisch:

This paper constructs several meta-indexes of journal rankings from existing journal rankings which all have limited coverage. It is applicable both for researchers in the field of economics and business administration (including their respective subfields) and includes also journals which are published in German. We discuss the relative merits of meta-indexes based on peer assessment and on citations.

Stichwörter: Zeitschriftenranking, Evaluation von Forschung, Meta-Index

JEL Klassifikation: A11

* Wir danken Klaus Ritzberger für die Bereitstellung der Daten seines Rankings und Bernd Fitzenberger, Oliver Fabel, Mihai Paunescu, Thomas Plümper, Heinrich Ursprung und einem anonymen Gutachter für konstruktive Hinweise.

^a Korrespondierender Autor: Albert-Ludwigs-Universität Freiburg im Breisgau, Institut für Allgemeine Wirtschaftsforschung, Abteilung Internationale Wirtschaftspolitik, Platz der Alten Synagoge 1, D-79085 Freiburg, Email: guenther.schulze@vwl.uni-freiburg.de

^b Institut für Arbeitsrecht und Arbeitsbeziehungen in der Europäischen Gemeinschaft (IAAEG), Universität Trier, Email: warning@iaaeg.de

^c Email: Christian_Wiermann@de.rolandberger.com

1 Einleitung

Die Messung von individuellen Publikationsleistungen gewinnt zunehmend an Bedeutung für die ökonomische Profession. Zunächst ist sie notwendig für das Verständnis des akademischen Marktes für Wirtschaftswissenschaftler; so lassen sich Ranglisten für die produktivsten Forscher oder Fachbereiche erstellen (u.a. Bommer und Ursprung 1998, Coupé 2003, Kalaitzidakis et al. 2003, Ursprung und Zimmer 2008), die Produktivitätsentwicklung über den Lebenszyklus und zwischen Kohorten abbilden (Rauber und Ursprung 2007, 2008) und die Determinanten von Berufsentscheidungen analysieren (Heining et al. 2007, Schulze et al. 2007). Verstärkt wird die gemessene Forschungsleistung aber auch als Steuerungsinstrument für Ressourcenverteilung verwendet – so haben messbare Publikationserfolge in Zielvereinbarungen für Professoren Eingang gefunden und Fachbereiche verteilen Mittel u.a. nach Forschungsleistung auf die Lehrstühle. Auch die Verteilung von Ressourcen nach Forschungsleistung zwischen Universitäten nach dem Vorbild der britischen „Research Assessment Exercise“ (Geary et al. 2004) scheint in Deutschland denkbar.

Ein solcher teilgebietsübergreifender Vergleich von Forschungsleistungen erfordert jedoch umfassende Zeitschriftenranglisten, die alle relevanten Journale aus allen Teilbereichen der Wirtschaftswissenschaften (der BWL und der VWL) umfassen und mit der gleichen Skala bewerten. Solche Ranglisten gibt es jedoch nicht. Selbst die sog. Wiener Liste¹ hat mit ihren über 1800 erfassten Zeitschriften deutliche Lücken etwa im Bereich der Volkswirtschaftslehre und der Statistik; die längste eher auf Volkswirtschaft fokussierte Liste von Combes und Linnemer (2003) hat gegenwärtig „nur“ 1141 Einträge. Das mag für die Erfassung relevanter Forschung im Sinne der Generierung neuer Erkenntnisse in der Volkswirtschaftslehre genügen; zum Vergleich der Forschungsleistungen innerhalb und zwischen Fakultäten oder als Steuerungsinstrument für Fakultäten ist sie ungeeignet. Streit über die mangelnde Reichweite der zu verwendenden Liste ist vorprogrammiert und damit ist die Akzeptanz jeder leistungsbezogenen Mittelverteilung von vornherein begrenzt.²

In diesem Papier stellen wir eine Methode zur Konstruktion eines Metarankings auf der Grundlage einer Basisliste vor: Die Bewertungen der in der Basisliste fehlenden Zeitschriften werden mit Hilfe anderer Rankings imputiert, so dass die entstehende Liste alle Zeitschriften enthält, die in mindestens einem der verwendeten Zeitschriftenrankings auftauchen. In unserem Fall sind dies 2822 Zeitschriften. Damit erfassen wir praktisch alle nennenswerten Zeitschriften im Bereich der VWL und BWL, jedoch nicht alle der angrenzenden Disziplinen (Psychologie, Politologie, Soziologie etc.).

Ein typischer Streitpunkt bei der Leistungsmessung besteht in der Frage, welche Methodik einer zu verwendenden Liste zugrunde liegen sollte. Grundsätzlich können die Listen auf der Basis von Expertenurteilen erstellt werden, wie dies bei der Wiener Liste, der Liste des Verbandes der

¹ <http://bach.wu-wien.ac.at/bachapp/cgi-bin/fides/fides.aspx/fides.aspx?journal=true;lang=DE>.

² Lücken in bestehenden Rankings ergeben sich in besonderem Maße für kleine Teilfächer, aber auch weniger prominente Journale im Bereich größerer Teildisziplinen fehlen in vielen Rankings.

Hochschullehrer für Betriebswirtschaft (Hennig-Thurau et al. 2004) oder der auf einer Emailbefragung basierenden Liste von Bräuninger und Haucap (2001) der Fall ist. Alternativ können die Zeitschriften auf der Basis der Häufigkeit bewertet werden, mit der ihre Artikel in anderen Journalen zitiert werden (wobei diese Zitationen wiederum qualitätsgewichtet eingehen können). Beide Verfahren haben ihre Vor- und Nachteile – damit hängt die Wahl der Liste davon ab, wie diese gegeneinander gewichtet werden.

In unserem Papier beschreiben wir die relativen Vorzüge beider Verfahren und erstellen mehrere Metarankings auf der Basis unterschiedlicher Listen – der Wiener Liste, der VHB Liste, der Liste von Ritzberger (2008) und der Liste von Combes und Linnemer (2003) –, um dem Leser die Wahl zu erlauben, statt sie ihm zu oktroyieren. Dabei werden die Unterschiede zwischen diesen Listen deutlich. Wir beschreiben die Eigenschaften der generierten Listen; aus Platzgründen stehen die vollständigen Listen im Internet zum Herunterladen unter www.vwl.uni-freiburg.de/iwipol/rankings.html zur Verfügung.

Im nächsten Abschnitt beschreiben wir unterschiedliche Methoden zum Erstellen von Zeitschriftenrankings und ihre relativen Vorzüge, im dritten Abschnitt diskutieren wir unsere Methode des Meta-Rankings, der vierte Abschnitt diskutiert die Ergebnisse der verschiedenen Meta-Rankings. Abschließende Bemerkungen folgen im fünften Abschnitt.

2. Ein Überblick über die Methodik von Zeitschriftenrankings

Expertenbefragung: Zeitschriften lassen sich bewerten, indem verschiedene Experten nach deren Reputation befragt und diese Bewertungen aggregiert werden.³ Die Zeitschriften werden kategorisiert nach einem vorgegebenen Schema, etwa A+, A, B, C, D. Beispiele für solche Rankings sind Dusansky und Vernon (1998), die Wiener Liste (vgl. Fn. 1), die Liste des Verbandes der Hochschullehrer für Betriebswirtschaft (VHB, Hennig-Thurau et al. 2004), Bräuninger und Haucap (2001) oder die Liste der Vereinigung holländischer Universitäten (VSNU) aus dem Jahre 1994.

Offensichtlich hat diese Methode den gravierenden Nachteil, dass Perzeptionsdefizite und –verzerrungen nicht ausgeschlossen werden können. Es kann nicht davon ausgegangen werden, dass die Experten alle Journale gleichermaßen kennen und beurteilen können; auch kann nicht ausgeschlossen werden, dass sie bestimmte Journale, etwa in ihrem eigenen Teilgebiet oder solche in denen sie zu publizieren hoffen oder publiziert haben, bewusst oder unbewusst bevorzugen. Ein bewusstes strategisches Verhalten ist insbesondere dann zu erwarten, wenn das Ziel einer Zeitschriftenbewertung die leistungsbezogene Mittelvergabe ist. Damit ist das Ergebnis dieser subjektiven Methode nicht nur stark abhängig von der Auswahl der Experten; die Methode ist darüber hinaus in vielen Situationen nicht anreizkompatibel. Verzerrte

³ Das bekannteste Verfahren ist die Delphi-Methode, bei der jeder Experte sein Votum unabhängig abgibt, jedoch seine Bewertung in einem zweiten Schritt korrigieren kann, nachdem er die Bewertungen der anderen erfahren hat.

Bewertungen sind die Folge. Ein zusätzliches Problem ergibt sich daraus, dass sie bei zu grober Kategorienbildung nicht hinreichend differenziert.⁴

Bewertung nach Impact-Faktoren: Eine interpersonell klar nachvollziehbare Bewertungsmethode liegt in der Berechnung der Impact-Faktoren. Diese werden durch Iteration auf der Basis der durchschnittlichen Häufigkeit, mit der Artikel des untersuchten Journals zitiert werden, berechnet, wobei die Zitate wiederum mit der Qualität der zitierenden Zeitschrift gewichtet werden.

Der Impact-Faktor I des Journals mit Ordnungskennzahl i nach der t -ten Iteration ergibt sich als

$$I_{i,t} = \frac{\sum_{j=1}^n Z_{ij} I_{j,t-1}}{U_i} \quad \text{mit} \quad I_{i,0} = \frac{\sum_{j=1}^n Z_{ij}}{U_i},$$

wobei Z_{ij} die Zahl der Zitierungen von Beiträgen der Zeitschrift i in der Zeitschrift j innerhalb eines vorgegebenen Zeitraums (z.B. bis zwei Jahre nach Erscheinen) bezeichnet und U_i den Umfang der Zeitschrift i angibt. Dieser kann durch die Zahl der publizierten Zeichen, Standardseiten oder Artikel gemessen werden. Selbstzitationen werden ausgeschlossen. Palacios-Huerta und Volij (2004) schlagen in einem axiomatischen Ansatz vor, Z_{ij} durch die Zitationsintensität der Zeitschrift j zu dividieren, also der Anzahl der Zitationen je Artikel, um so unterschiedlichen Zitationsgewohnheiten in unterschiedlichen Zeitschriften (und Teildisziplinen) Rechnung zu tragen. Sie nennen das die ‚invariante Methode‘ (vgl. auch Ritzberger (2008)). n ist die Anzahl der Zeitschriften, deren Zitationen erhoben werden. Diese werden im *Journal Citation Report* (JCR) erfasst, der vom *Institute for Scientific Information* herausgegeben wird und über 1700 sozialwissenschaftliche Zeitschriften enthält, davon waren in der Ausgabe 2003 160 in Economics. Typischerweise werden nur diese zur Analyse der Zitationsverflechtung verwendet.⁵

Diese Art der Journal-Bewertung wurde von Liebowitz und Palmer (1984) eingeführt, weitere Arbeiten sind u.a. von Laband und Piette (1994), Kalaitzidakis et al. (2003) und Theoharakis und Axarloglou (2003). Die invariante Methode verwenden Palacios-Huerta und Volij (2004),

⁴ Die Wiener Liste ist ein „gutes“ Beispiel: So haben dort etliche Zeitschriften den Rang A, aber die impact-basierten Rankings von Kalaitzidakis et al. (KMS), Ritzberger (Rb) und Combes und Linnemer (CL) weisen große Rang- und Bewertungsunterschiede auf (Ränge und Werte in Klammern). Beispiele sind *Games and Economic Behavior* [CL Wert 0,67, Rb Rang 16 von 262 (Wert 21,24), KMS Rang 11 von 159 (Wert 35,49)], *Journal of International Economics* [CL Wert 0,67, Rb Rang 15 (Wert 22,84), KMS Rang 30 (Wert 7,84)], aber auch *Open Economies Review* [CL Wert 0,33, Rb Rang 211 (Wert 0,21), KMS Rang 110 (Wert 0,34)], *Kyklos* [CL Wert 0,33, Rb Rang 198 (Wert 0,4), KMS Rang 81 (Wert 0,91)] und sogar *Food Policy* [CL 0,17, Rb Rang 213 (Wert 0,26), KMS Rang 119 (Wert 0,23)]. Dies sind keine Einzelfälle. Es wird offensichtlich, dass die Wiener Liste in ihrer Zuordnung zu Kategorie A nicht stark differenziert.

⁵ Etwa bei Kalaitzidakis et al. (2003). Das liegt v.a. daran, dass bislang die auf Impact-Faktoren basierenden Ranglisten v.a. für den Bereich Volkswirtschaftslehre erstellt wurden. Dies muss allerdings nicht so bleiben: Der Social Science Citation Index enthielt 2007 213 Zeitschriften in der Kategorie „Economics“ und insgesamt 216 in den Kategorien „Business“ (78), „Business, Finance“ (50) und „Management“ (88) (<http://www.thomsonscientific.com/cgi-bin/jmlst/jlsubcatg.cgi?PC=J>). Vgl. auch Fn 7.

Ritzberger (2008) und Kodrzycki und Yu (2006), die Journal Rankings für unterschiedliche Gruppen von zitierenden Zeitschriften erstellen.⁶

Diese Verfahren haben den großen Vorteil objektiv in dem Sinne zu sein, dass sie nicht von subjektiven Bewertungen Einzelner abhängen und sich für Manipulationen der Bewertenden nicht eignen, da die Ergebnisse unabhängig validiert werden können. Dies mag den großen Erfolg dieser Analysen erklären. Dennoch hat auch diese Methode nicht unerhebliche Nachteile, die nachfolgend kurz genannt seien.

- Das Ergebnis hängt wesentlich von der Wahl der Journale ab, deren Zitate zur Berechnung der Impact-Faktoren herangezogen werden. Ist diese Auswahl nicht repräsentativ für die zu evaluierende Gruppe an Zeitschriften, kommt es zu Verzerrungen zu Lasten der unterrepräsentierten Journale. Dies ist beispielsweise dann der Fall, wenn die im JCR berücksichtigten Journale nicht repräsentativ für bestimmte Teilgebiete sind oder nicht-englischsprachige Zeitschriften weniger stark berücksichtigen. Beides mag beispielsweise deutsche betriebswirtschaftliche Zeitschriften dann benachteiligen, wenn nur die Journale der Kategorie „Economics“ des JCR berücksichtigt werden.⁷
- Zitationskartelle können die Bedeutung von Zitaten als Maß für die wissenschaftliche Bedeutung verzerren, wenn die Zitationskartelle in bestimmten Zeitschriften stärker vertreten sind.
- Bewirken Gutachterkartelle oder nationale oder andere Präferenzen der Herausgeber ungleichen Zugang zu Zeitschriften und müssen die Benachteiligten auf schlechtere Zeitschriften ausweichen, so ist die zitationsbasierte individuelle Leistungsbewertung verzerrt. Blank (1991) zeigt anhand des *American Economic Review*, dass Autoren an Spitzenuniversitäten bevorzugt werden.
- Die Zahl der im JCR erfassten Journale ist begrenzt, so dass viele eher zweit- und drittklassige Journale so nicht gerankt werden können.
- Neue Journale kann dieses Verfahren nicht oder nicht adäquat berücksichtigen, da sie entweder noch nicht im JCR enthalten sind oder noch keine entsprechende Zitationsbasis haben. Manche Gründungen von Zeitschriften aber reflektieren ein sich schnell entwickelndes Forschungsgebiet, das erheblich zum Erkenntnisgewinn in den Wirtschaftswissenschaften beiträgt.
- Kleinere Teilfächer werden tendentiell benachteiligt, da Beiträge (gleicher Qualität) größenbedingt weniger zitiert werden. Dies gilt auch für sehr technische Zeitschriften, die eine geringere Leserschaft haben.
- Journale, die viele Überblicksartikel oder Sonderbände publizieren, erhalten cet.par. ein höheres Ranking, da diese öfter zitiert werden.⁸ Die Anzahl der Zitationen ist eben nur ein ungenügendes Maß für den Beitrag eines Artikels zum Erkenntnisgewinn der Profession.

⁶ Daneben gibt es auch einfache Impact-Berechnungen, die den Wert der zitierenden Zeitschrift nicht berücksichtigen, sondern nur Zitationen (oder nicht-zitierte Papiere) zählen – der JCR weist einen solchen Impact-Faktor aus. Da solche Berechnungen die Logik des Zeitschriftenranking nicht konsequent anwenden – die zitierten Zeitschriften werden nach Qualität unterschieden, die zitierenden nicht – werden sie nicht weiter betrachtet.

⁷ Ritzberger (2008) verringert eine solche Verzerrung erheblich, da er auch die Kategorien „Business“, „Business, Finance“, „Industrial Relations and Labor“ (nicht aber „Management“) und weitere Statistik-Journale einbezieht.

⁸ Das hohe Ranking des *Journal of Economic Literature (JEL)* zeigt dies sehr deutlich: KMS ermitteln Rang 20 von 159, Rb nimmt das *JEL* aus seinem Ranking.

- Die Impact-Faktoren werden auf der Basis aller (internationaler) Zitationen berechnet; oftmals sind aber Artikel mit starkem Länderbezug besser in „nationalen“ Zeitschriften aufgehoben, weil sie die relevante Leserschaft besser erreichen. So mag die Bedeutung einer Publikation im *German Economic Review* für einen deutschen Forscher höher sein als eine im *Scottish Journal of Political Economy*, obwohl beide gleich gerankt sind [CL 0,17, Wien 4]. Das zitationsbasierte Verfahren kann das naturgemäß nicht abbilden. Bräuninger und Haucap (2001) zeigen, dass die Einschätzung der deutschsprachigen Ökonomen systematisch von Impact-Rankings abweicht – Zeitschriften aus dem deutschsprachigen Raum werden höher bewertet als in den Impact-basierten Ranglisten.⁹
- Besonders im unteren Bereich sind Impact-Faktoren ziemlich volatil, so dass das Ranking dort stark vom Untersuchungszeitraum abhängen kann.

Viele dieser Nachteile sind dann potentiell Vorteile der auf Expertenbefragungen basierenden und der hybriden Verfahren, da sich dort diese Nachteile grundsätzlich vermeiden lassen.

Ein weiterer Nachteil mag in der extrem schiefen Verteilung der Impact-Faktoren gesehen werden. In dem bevorzugten Ranking von Kalaitzidakis et al. (2003) beispielsweise hat die Zeitschrift mit Rang 1 den Wert 100, Rang 5 den Wert 58, Rang 10 bereits 36, und die Ränge 20, 30, 40, 50, 60 die Werte 19, 8, 5, 4 und 2. Alle Journale mit den Rängen 78 – 159 liegen unter 1. Demnach würde eine Veröffentlichung im *American Economic Review* genauso gewertet wie 12 Publikationen im *Journal of International Economics* oder 25 im *Journal of Law, Economics and Organization*, der jeweils führenden Zeitschrift in ihrem Gebiet. Eine Veröffentlichung in *Kyklos* oder dem *Weltwirtschaftlichen Archiv* wöge weniger als ein Hundertstel. Ähnliches gilt für die anderen Impact-basierten Rankings. Es stellt sich allgemein die Frage, ob ein untransformierter Impact-Faktor ein adäquater Bewertungsmaßstab für die individuelle Forschungsleistung ist.

Umgekehrt lässt sich natürlich fragen, ob die Gewichtung, die implizit der Wiener Liste oder anderer kategorialer Listen zugrunde liegt, angemessen ist. Obwohl die Liste ein ordinales Ranking vornimmt, wird man nicht umhin können bei der Bewertung individueller Forschungsleistungen verschiedene Kategorien zu aggregieren. Wählt man dabei die Zahlen 5 bis 1 für die Kategorien A+, A, B, C, D, so bedeutete dies, dass eine Veröffentlichung in der *Zeitschrift für Betriebswirtschaft* oder dem *German Economic Review* 80 Prozent einer Veröffentlichung in *Management Science* oder dem *American Economic Review* zählte – eine Einschätzung, die viele Wissenschaftler nicht teilen dürften.¹⁰

Beide Verfahren mögen die relativen Bedeutungen der einzelnen Publikationen nicht adäquat widerspiegeln, deshalb haben einige Autoren hybride Verfahren entwickelt.

⁹ Ähnliche Evidenz findet sich in Malouin und Outreville (1987) für Forscher aus dem Vereinigten Königreich, Frankreich und Quebec, die europäische Journale höher bewerten als ihre Kollegen in den USA.

¹⁰ Natürlich ließen sich grundsätzlich andere Abstufungen zwischen den Kategorien wählen, etwa wie dies Combes und Linnemer (2003) in ihrer hybriden Methode verwendet haben (s.u.).

Hybride Methoden

Combes und Linnemer (2003) bewerten die EconLit Zeitschriften von 1 bis 1/12, wobei die ersten fünf Zeitschriften den Wert 1, die nächsten sechzehn den Wert 8/12, weitere 39 den Wert 6/12 und die folgenden 68 den Wert 4/12 haben. Den nächsten 138 Journalen weisen die Autoren den Wert 2/12 zu, allen anderen den Wert 1/12.¹¹ In ihrer Kategorisierung sind sie weitgehend konsistent mit den Werten der Impact-Faktoren, soweit diese vorhanden sind. Außerdem geben sie den führenden Zeitschriften eines Teilfeldes mindestens den Wert 6/12 und umgekehrt sehr spezialisierten Zeitschriften nicht mehr als 6/12.

Bauwens (1998) gibt allen im JCR enthaltenen Ökonomiezeitschriften einen Wert zwischen 1 und 5 auf der Basis der Impact-Faktoren und den Zitationen in einem Jahr und weist allen nicht im JCR enthaltenen Zeitschriften den Wert 1 zu.

Lubrano et al. (2003) verwenden folgendes Verfahren: Einer der Koautoren (Alan Kirman) überprüft als Experte ein früheres Ranking von Combes und Linnemer, dehnt es von 307 auf 505 Zeitschriften aus und weist den Zeitschriften die Werte 1,2,4,6,8,10 zu. Parallel dazu wird der Impact-Faktor multipliziert mit der Gesamtzahl der Zitationen und dieser Wert wird dann nach einem vorgegebenen Schema einem der obigen Werte von 1 bis 10 zugeordnet. Diese Information wird in einem zweiten Schritt dem Experten zugänglich gemacht, der die Gelegenheit erhält, sein früheres Urteil zu revidieren.

Natürlich kombinieren die hybriden Verfahren Vor- und Nachteile der beiden anderen Verfahren; allerdings scheint die Abstufung der Journalgewichte bei Combes und Linnemer und Lubrano et al., wiewohl subjektiv, weniger extrem als die der anderen. Außerdem erlauben sie einige der Schwächen der Impact-Rankings zu korrigieren – etwa die vergangenheitsbezogene schlechtere Bewertung neuer Journale oder die Höherbewertung von Journalen mit erheblichem Teil an Überblicksartikeln und Sonderheften oder von semi-professionellen Zeitschriften. Die Orientierung an den Impact-Faktoren setzt willkürlichen Bewertungen gewisse Grenzen.

3 Methode und Datensatz

3.1 Verwendete Daten

Wir konstruieren mehrere Meta-Rankings auf der Basis jeweils einer Liste, wobei wir die fehlenden Werte mit Hilfe anderer Rankings imputieren. Wir verfolgen damit das Ziel, Rankings zu erstellen, die sowohl der Betriebswirtschaftslehre als auch der Volkswirtschaftslehre einschließlich ihrer Teilgebiete gerecht werden. Dies bestimmt die Wahl unserer Grundrankings. Diese müssen relativ aktuell, relativ umfassend und thematisch nicht zu speziell sein, damit es zu hinreichenden Überschneidungen zwischen den Rankings kommt, die erst das Imputieren

¹¹ Die Liste von Combes und Linemer bewertet also alle in EconLit erfassten Zeitschriften. Da sich deren Zahl ständig erhöht hat, ist eine Liste im Sinne von Combes und Linnemer (2003) „mitgewachsen“. Werden alle bis 2002 in EconLit erfassten Zeitschriften berücksichtigt, so umfasst die Liste 951 Zeitschriften; gegenwärtig enthält EconLit 1141 Zeitschriften (vgl. Fn 14).

erlauben. Gleichzeitig muss die Vereinigungsmenge der in ihnen erfassten Zeitschriften die Bandbreite der Wirtschaftswissenschaften abdecken, wobei in unserer Studie namentlich auch die Zeitschriftenaus dem deutschsprachigen Raum erfasst sein müssen.

Wir verwenden¹²

1. Die Wiener Liste **[Wien]**: Expertenbefragung, Wertebereich 1-5, n=1870, Schwerpunkt eher betriebswirtschaftlich.¹³
2. Combes und Linnemer (2003) **[CL]**: Hybride Liste, Wertebereich 1/12 – 1, n=1141, Schwerpunkt eher volkswirtschaftlich.¹⁴
3. VHB Liste **[VHBR]**: Expertenbefragung, Wertebereich 1-6, n=681, Schwerpunkt betriebswirtschaftlich. Die VHBR Liste ist eine transformierte Liste einer ursprünglichen Liste, die eine feinere Abstimmung erlaubte **[VHB]**. Als erklärende Liste zum Imputieren verwenden wir die zweite Liste, da sie mehr Informationen enthält; als Basisliste verwenden wir die vom VHB zum Ranking verwendete Liste mit ganzzahligen Werten 1 – 6.¹⁵
4. Ritzberger**[Rb]**: Impact-Bewertung (invariante Methode), Wertebereich 0 – 100, n=261, Schwerpunkt volkswirtschaftlich. Die **RbR** Liste ist eine ordinalisierte Liste der ursprünglichen Liste.
5. Khalaitzidakis et al. (2003) **[KMS]**: Impact-Bewertung, Wertebereich 0 – 100, n=159, Schwerpunkt volkswirtschaftlich.
6. Bräuninger und Haucap (2001) **[BH]**: Internetbefragung der Mitglieder des Vereins für Socialpolitik, Wertebereich 1,02 – 4,83, n=149, Schwerpunkt stärker deutsche Zeitschriften.¹⁶
7. Theoharakis und Axaroglou (2003) **[TA]**: Impact-Bewertung, Wertebereich 1,2 - 91,64, n=100, Schwerpunkt volkswirtschaftlich.

¹² Deskriptive Statistiken finden sich in Tabelle A1 im Anhang. Die Zahlen der Zeitschriften berücksichtigen dabei immer Korrekturen von Doppelzählungen, Umbenennungen u. dgl., so dass die Werte vereinzelt von denen in den Originallisten abweichen können. In den seltenen Fällen, in denen für unterschiedliche Namen des gleichen Journals unterschiedliche Bewertungen in mindestens einer Liste angegeben wurden, haben wir beide Namen aufgenommen. Bei der Wiener Liste und der VHB Liste haben wir die ordinalen Werte von A+ bis D bzw. E in kardinale Werte 5-1 bzw. 6-1 transformiert.

¹³ Zu finden unter <http://bach.wu-wien.ac.at/bachapp/cgi-bin/~des/~des.aspx/~des.aspx?journal=true;lang=DE>, Stand: 26.2.2007

¹⁴ Die Bewertungen in Combes und Linnemer (2003) beziehen sich auf die Zeit bis 2002 und die bis 2002 in EconLit erfassten Zeitschriften. Das von uns verwendete Ranking von CL umfasst alle jetzt in EconLit erfassten Zeitschriften, wobei wir allen Zeitschriften, die von CL keinen höheren Wert zugewiesen bekommen haben, mit 0,08 bewerten (wie das CL mit den „restlichen“ Journalen gemacht haben). Potentiell machen wir dann einen Fehler, wenn einige der neu hinzugekommenen Zeitschriften nach der CL-Methode in eine andere Kategorie fallen würden. Allerdings scheint dieser Fehler kleiner als wenn wir die neuen Zeitschriften vernachlässigt hätten.

¹⁵ Vgl. Hennig-Thurau et al. (2004). Die Liste findet sich unter <http://pbwi2www.uni-paderborn.de/WWW/VHB/VHB-Online.nsf/id/EB24EF9AF51F72D2C125709600494116>, Stand: 26.2.2007. Wir beziehen uns auf die längste Liste, die alphabetisch *alle* bewerteten Zeitschriften enthält. Während ursprünglich nur die Zeitschriften mit mehr als zehn Nennungen eine Kategorie zugewiesen wurde, übernehmen wir die Kategorisierung für alle aufgeführten Zeitschriften mit mindestens fünf Bewertungen: (A+: Wert größer 9; A: Wert zwischen 8 und 9; B: zwischen 7 und 8; C: zwischen 6 und 7; D: zwischen 5 und 6; E: kleiner 5).

¹⁶ Bräuninger und Haucap stellen zwei Rankings auf: eines nach der Reputation der Zeitschrift, eines nach der Bedeutung der Zeitschrift für die eigene Arbeit. Wir verwenden Erstgenanntes.

Tabelle 1 gibt die paarweisen Kreuzkorrelationen sowie den Umfang der Schnittmenge an.

Tabelle 1: Korrelationen der Originalrankings

	Rb	RbR	Wien	VHB	VHBR	CL	KMS	TA	BH
Rb	1,0000 261								
RbR	0,7791 261	1,0000 261							
Wien	0,3901 230	0,4412 230	1,0000 1870						
VHB	0,5047 126	0,6872 126	0,7440 510	1,0000 677					
VHBR	0,4914 126	0,6609 126	0,7105 510	0,9635 677	1,0000 677				
CL	0,6990 202	0,6715 202	0,4144 359	0,5892 154	0,5649 154	1,0000 1141			
KMS	0,8526 133	0,6868 133	0,4664 143	0,5662 60	0,5195 60	0,7719 153	1,0000 159		
TA	0,7892 76	0,5892 76	0,4583 89	0,5218 49	0,4935 49	0,7228 99	0,8541 71	1,0000 100	
BH	0,7108 93	0,6945 93	0,5691 124	0,6661 71	0,6812 71	0,7480 136	0,7504 80	0,8092 68	1,0000 149

Die Zahlen unter den Korrelationskoeffizienten geben den Umfang der Schnittmenge beider Rankings an.

Es ist augenfällig, dass die Impact-basierten Listen [KMS, TA, Rb, aber auch CL] hoch miteinander korreliert sind, aber auch die Wiener Liste und die VHB Liste. Zwischen diesen Gruppen ist die Korrelation eher niedrig, was die unterschiedliche Methodik deutlich reflektiert. Die Liste von Bräuninger und Haucap nimmt eine Mittelstellung ein. Die Überschneidungen der kleinen Listen [KMS, TA, Rb, BH] mit den großen Listen [Wien, CL] sind relativ hoch, da sie v.a. die international führenden Zeitschriften beinhalten, die die größeren Listen natürlich auch haben. VHB und Wiener Liste haben ebenfalls große Überschneidungen; erstaunlich aber ist die geringe relative Überschneidung der großen Listen [Wien und CL] – sie sind in Methode und Abdeckung sehr unterschiedlich.

3.2 Verwendete Methode

Um alle Zeitschriften in gemeinsamen Rankings vergleichbar zu machen, erstellen wir vier Meta-Rankings, die aus vier verschiedenen Basislisten durch Imputation jeweils eine gesamte Liste ermitteln. Diese Imputationen verlängern die Basislisten um die fehlenden Zeitschrifteneinträge mit Hilfe der jeweils anderen sechs Listen.

Wir gehen von der Grundidee ähnlich, in der methodischen Umsetzung aber deutlich anders als Mingers und Harzing (2007) vor, die vor dem Hintergrund der regelmäßigen Evaluation von britischen Fakultäten durch das Research Assessment Exercise (RAE) ein Zeitschriftenranking für 859 Zeitschriften aus den Bereichen „Business“ und „Management“ erstellen. Auf der Basis von elf Rankings erzeugen Mingers und Harzing (2007) über eine Clusteranalyse vier Gruppen von

Zeitschriften jeweils gleicher Qualität, die den durch das RAE vorgegebenen Bewertungsstufen entsprechen. Um dieses Verfahren anwenden zu können, verlängern sie die Basislisten durch Imputation.¹⁷ Auf diese Weise werden alle Listen verlängert, jedoch nicht vollständig aufgefüllt, so dass sich die Listen in ihrer Länge unterscheiden. Das eigentliche Meta-Ranking findet dann durch eine Clusteranalyse auf der Basis der verlängerten Listen statt.

Das Metaranking in diesem Beitrag folgt einer anderen Logik. Unser Ziel ist es, bereits existierende Listen soweit wie möglich zu verlängern und dabei den Grundcharakter dieser Listen zu erhalten, anstatt durch Clusteranalyse eine „mittlere“ Liste zu generieren. Wir wollen damit den Fakultäten ermöglichen, diejenigen Listen zu verwenden, die ihren Präferenzen am besten entsprechen, und die Robustheit der so erzielten Ergebnisse durch andere Listen zu überprüfen.

Bei der Wahl unserer vier Basislisten spielen mehr inhaltliche, denn technische Gründe eine Rolle: Wir verwenden zwei eher volkswirtschaftlich geprägte Listen, die von Combes und Linnemer (2003) und die von Ritzberger (2008), und zwei betriebswirtschaftlich fokussierte Listen, die Wiener Liste und die VHB-Liste.

Als imputierte Ergebnislisten berechnen wir nur kategoriale Listen, weil wir glauben, dass die extrem schiefe Verteilung der Listen, die nur Impact-Faktoren berechnen [KMS, Rb, TA], sich als Bewertungs- und Steuerungsinstrument nicht eignen: Die Unterschiede in den Impactfaktoren überschätzen erheblich die unterschiedliche Forschungsleistung bei den Top-Zeitschriften, aber auch den sehr guten Zeitschriften; die Zeitschriften am unteren Rand haben vielfach den Wert Null oder nahe Null (vgl. Abschnitt 2).¹⁸ Dabei müssen diesen Kategorien nicht die kardinalen Werte 1-6 oder 1-5 zugewiesen werden, andere Abstufungen, wie etwa bei Combes und Linnemer sind auch für die anderen imputierten Listen denkbar.

Wien und VHBR sind ordinale Listen, CL ist kategorial mit sechs Kategorien, wobei diese kardinale Werte annehmen (s.o.). Die Liste von Ritzberger kategorisieren wir nach einem Verfahren, das sich sehr eng an die von Ritzberger selbst gebildeten sechs Kategorien anlehnt. Allerdings benutzt Ritzberger (2008, Abschnitt 4) als Kriterium nicht nur sein eigenes Ranking sondern auch das von KMS und fordert, dass die gerankten Zeitschriften in *beiden* Listen vorkommen und dort jeweils eine bestimmte Positionierung haben. Das führt dazu, dass in KMS nicht gerankte

¹⁷ Hierfür wählen sie die beiden längsten Listen in einem ersten Schritt als unabhängige um die drittlängste Liste darauf zu regressieren und anschließend Werte für die fehlenden Zeitschriften in der abhängigen durch die Werte der in beiden unabhängigen Variablen vertretenden Zeitschriften zu imputieren (logistische Regression). In einem zweiten Schritt nimmt die nächst längste Liste die Funktion der abhängigen Variablen ein und die im ersten Schritt imputierte wie auch die beiden längsten Listen gehen auf die rechte Seite. Für die Imputation werden nur die Variablen herangezogen, die auf dem Ein-Prozent-Niveau signifikant sind.

¹⁸ Nach KMS zählte danach ein Artikel im *Quarterly Journal of Economics* nur gut halb soviel wie ein Artikel im *American Economic Review*, ein Artikel im *Economic Journal* zählte nur ein Fünftel. Die führenden Zeitschriften in ihrem Gebiet *Journal of Development Economics*, *Journal of International Economics*, *Journal of Labor Economics* und *Journal of Public Economics* werden bei KMS im Verhältnis 1: 1,4 : 2,3 : 3.6 gewichtet. Andererseits zählt eine Veröffentlichung in den *Jahrbüchern für Nationalökonomie und Statistik* oder *Betriebswirtschaftliche Forschung und Praxis* nach Impactfaktor in Rb nichts. Ein weiterer Grund besteht in der starken Volatilität von Impactfaktoren über die Zeit.

Zeitschriften in Ritzbergers ordinalem Ranking fehlen und manche Zeitschriften anders gerankt sind, als wenn nur das Ritzberger Ranking verwendet worden wäre. Im Ergebnis ordnet Ritzberger die Zeitschriften in die Gruppen A+ (enthält 10 Journals), A (13), B+ (15), B (19), C+ (30) und C (30).¹⁹

Wir verwenden in unserer Kategorisierung der Ritzbergerliste nur Informationen aus dieser Liste: In einem ersten Schritt weisen wir den Journalen, die in Ritzbergers ordinalem Ranking auftauchen, die Kategorien A+, A, B+, B, C+ und C zu, allerdings nun unabhängig von Kalaitzidakis et al. (2003). Dabei behalten wir die Größe der Kategorien bei, so dass sich die Zuordnung der Journale auf die Gruppen leicht verändert – vor allem am Rand der Gruppen. Konkret werden die Zeitschriften von Rang 1 bis 10 in die Kategorie A+, Rang 11 bis 23 in Kategorie A, Rang 24 bis 38 in Kategorie B+, Rang 39 bis 57 in Kategorie B, Rang 58 bis 87 in Kategorie C+ und ab Rang 88 in Kategorie C eingeordnet. Auf diese Weise erhalten wir wiederum sechs Gruppen. Unsere Liste ähnelt stark der ursprünglichen ordinalen Ritzberger Liste.

In einem zweiten Schritt sortieren wir dann in diese sechs Gruppen die bislang im ordinalen Ranking von Ritzberger nicht kategorisierten Zeitschriften entsprechend ihrer ermittelten Impactwerte ein, so dass sich einige Kategorien in ihrer Besetzung erhöhen, aber ihre Intervallgrenzen ihre Impactfaktoren beibehalten. In einem dritten Schritt schließlich weisen wir den Kategorien A+ bis C die numerischen Werte 6 bis 1 zu, was dem Vorgehen bei der Wiener Liste und der VHB Liste entspricht. Diese neue Liste bezeichnen wir mit **RbR**.

Diese vier kategorialen Listen – VHBR, Wien, CL und RbR – dienen im Folgenden als Basislisten für die Imputation, so dass am Ende vier imputierte Listen vorliegen. Zur Imputation ziehen wir nacheinander die jeweils anderen sechs Listen aus KMS, TA, BH, Wien, VHB, Rb und CL heran.²⁰ Wir verwenden zwei verschiedene Verfahren zur Konstruktion unserer Metaindexes, ein iteratives und ein randomisiertes Verfahren, die allerdings – trotz ihrer technischen Unterschiede – zu sehr ähnlichen Ergebnissen führen. Da das randomisierte Verfahren deutlich weniger anfällig für Ausreißer und damit robuster ist, konzentrieren wir uns auf dieses und skizzieren die Idee des iterativen Verfahrens kurz im Anhang.

Die Grundidee des randomisierten Imputationsverfahrens besteht darin, die Basisliste zunächst auf *eine* zufällig ausgewählte andere Liste zu regressieren und dann die Werte aufgrund dieser Schätzung in die Basisliste zu imputieren. Dieses Verfahren wird mit allen anderen Listen wiederholt, bis die imputierte Liste Werte für alle Zeitschriften enthält, die in mindestens einer Liste aufscheinen. Dient beispielsweise CL als Basisliste und KMS als erste Liste zur Imputation, so regressieren wir CL auf KMS in einem geordneten Probitmodell. Auf Basis der geschätzten Koeffizienten werden dann den Zeitschriften, die in KMS, aber nicht der Basisliste CL enthalten sind, die Werte der Basisliste zugeordnet, die die höchste Prognosewahrscheinlichkeit haben. Die imputierte Liste enthält nun alle Zeitschriften, die in der Basisliste CL oder in KMS enthalten

¹⁹ Die übrigen 144 Zeitschriften kategorisiert Ritzberger nicht, sondern führt diese mit ihren kardinalen Bewertungen im Anhang auf. Es sind dies v.a. jene Zeitschriften, die nicht in KMS enthalten sind.

²⁰ Als Basislisten verwenden wir RbR und VHBR, zur Imputation die Listen Rb und VHB.

sind. Anschließend wird die teilweise imputierte Basisliste schrittweise auf jede andere Liste regressiert, bis sie vollständig imputiert ist.

Da das Ergebnis dieses Verfahren bei einmaliger Durchführung stark von der Reihenfolge abhängen würde, mit der die Listen zur Imputation herangezogen werden, haben wir diesen Prozess randomisiert und wiederholt. Eine gleichverteilte Zufallsvariable entscheidet über die Reihenfolge der Regressionen und damit der Imputationsschritte. Dieser Prozess der Imputation in zufällig gereihten Imputationsschritten wird 1000 Mal wiederholt und anschließend der Mittelwert der 1000 imputierten Listen gebildet und auf ganze Zahlen gerundet. So erhalten wir vier robuste Metarankings auf Basis von vier deutlich unterschiedlich konstruierten Basislisten. Es zeigt sich, dass das iterative Imputationsverfahren sehr ähnliche Ergebnisse liefert wie das wiederholte, randomisierte Verfahren. Die Korrelationen der nach unterschiedlichen Verfahren imputierten Listen derselben Basisliste sind alle höher als 0,95.

4 Empirische Ergebnisse: Meta-Rankings

Zunächst scheinen die deskriptiven Statistiken zu zeigen, dass die imputierten den Originallisten bei allen vier Rankings recht ähnlich sind (Tabelle 2). Allerdings sind die Unterschiede in den Mittelwerten jeweils hoch signifikant (t-test). Auffällig ist, dass die Mittelwerte bei den beiden BWL orientierten Listen, die auf Expertenurteilen beruhen, nach der Imputation höher, bei dem VWL orientierten Ranking auf Basis der Zitationsanalyse deutlich geringer sind. Während die erstgenannten Listen Lücken im oberen Bereich haben, fehlen CL im unteren Bereich Zeitschriften. Zum Teil liegt das Ergebnis auch an der Konstruktion der Originallisten – die oberen Kategorien bei CL sind relativ schwach besetzt, während bei VHB und Wien die Kategorie A (5 bzw. 4) sehr stark besetzt ist.

Tabelle 2: Deskriptive Statistiken der Original- und der imputierten Listen nach dem randomisierten Imputationsverfahren

<i>Variable</i>	<i>Anz. Beob.</i>	<i>Mittelwert</i>	<i>Std. Abw.</i>	<i>Min</i>	<i>Max</i>
CL	1141	0,1324	0,1288	0,08	1
CL_IMP	2822	0,1058	0,0916	0,08	1
RbR	261	2,3257	1,5004	1	6
RbR_IMP	2822	1,1743	0,6684	1	6
VHBR	677	3,4476	1,3800	1	6
VHBR_IMP	2822	4,0039	1,3783	1	6
Wien	1870	3,1348	0,9407	1	5
Wien_IMP	2822	3,3816	0,8775	1	5

Die Unterschiede zwischen der imputierten und der jeweiligen Original-Liste werden durch die Verteilung und Besetzung der Kategorien noch deutlicher. Tabelle 3 zeigt die Häufigkeitsverteilung für die drei Meta- und Original-Rankings.

Tabelle 3: Häufigkeitsverteilung der Original- und der imputierten Listen nach dem randomisierten Imputationsverfahren

<i>CL</i> ²¹	<i>CL</i>	<i>CL_IMP</i>	<i>VHB/Wien/RbR1</i>	<i>VHBR</i>	<i>VHBR_IMP</i>	<i>RbR</i>	<i>RbR_IMP</i>	<i>Wien</i>	<i>Wien_IMP</i>
0,08	876 (76,77)	2.502 (88,66)	1	82 (12,11)	369 (13,08)	107 (41,00)	2.574 (91,21)	142 (7,59)	150 (5,32)
0,17	137 (12,01)	145 (5,14)	2	87 (12,85)	87 (3,08)	62 (23,75)	122 (4,32)	250 (13,37)	261 (9,52)
0,33	68 (5,99)	112 (3,97)	3	149 (22,01)	167 (5,92)	36 (13,79)	58 (2,06)	734 (39,25)	815 (28,88)
0,50	39 (3,42)	41 (1,45)	4	196 (28,95)	772 (27,36)	25 (9,58)	33 (1,17)	702 (37,54)	1.554 (55,07)
0,67	16 (1,40)	17 (0,60)	5	131 (19,35)	1.395 (49,43)	18 (6,90)	20 (0,71)	42 (2,25)	42 (1,49)
1,00	5 (0,44)	5 (0,18)	6	32 (4,73)	32 (1,13)	13 (4,98)	15 (0,53)	--	--

Absolutwerte sind in der jeweils oberen Zeile, Prozentwerte in Klammern in der unteren. *_IMP* kennzeichnet die imputierten Listen.

Hier lässt sich klar erkennen, wie verschieden die Imputation auf die einzelnen Listen wirkt, was die unterschiedliche Veränderung der Mittelwerte erklärt. Während bei der VHBR-Liste die Kategorie 5, bei der Wien-Liste die Kategorie 4 überproportional aufgefüllt wurde – beides Gruppen mit Zeitschriften höherer Qualität –, so wurde bei der CL-Liste vor allem die unterste Kategorie (0,08) aufgefüllt, was die Häufigkeitsverteilung entsprechend verschob. Bei der kategorialen Ritzbergerliste (RbR) ist dies sogar noch ausgeprägter der Fall.

Der Zusammenhang zwischen den imputierten und den Original-Werten zeigt sich auch anhand der Korrelation: Während der Zusammenhang zwischen der Wiener und der VHB Liste (VHBR) durch die Imputation enger wird (von 0,7105 auf 0,8646 steigt), verringert sich die Korrelation dieser beiden zu der imputierten Combes und Linnemer Liste (0,2186 statt 0,4144 bzw. 0,2165 statt 0,5649) und zu der kategorialen Ritzbergerliste (RbR von 0,2202 statt 0,4412 bzw. von 0,2132 statt 0,6872). Die erste Beobachtung ist durch eine relative große Überschneidung, die zweite durch eine relativ geringe Überschneidung der Listen zu erklären. Die Korrelation zwischen Combes und Linnemer und Ritzberger nimmt durch die Imputation leicht zu (von 0,6715 auf 0,7867).

Tabelle 4: Paarweise Korrelationen der imputierten mit den Original-Listen nach dem randomisierten Imputationsverfahren

	<i>Wien_IMP</i>	<i>VHBR_IMP</i>	<i>CL_IMP</i>	<i>RbR_IMP</i>
<i>Wien_IMP</i>	1,0000			
<i>VHBR_IMP</i>	0,8646	1,0000		
<i>CL_IMP</i>	0,2186	0,2165	1,0000	
<i>RbR_IMP</i>	0,2202	0,2132	0,7867	1,0000
<i>Wien</i>	1,0000	0,8499	0,3111	0,3266
<i>VHBR</i>	0,7230	1,0000	0,4993	0,5815
<i>CL</i>	0,1121	0,1501	1,0000	0,7692
<i>RbR</i>	0,3932	0,4698	0,7029	1,0000

²¹ Manchester School (früher: Manchester School of Economic and Social Studies) sind bei CL getrennt aufgeführt. Wir fassen zusammen und haben daher eine Beobachtung in der Kategorie 0,17 weniger.

Es wird deutlich, dass sich die CL Liste und die RbR Liste in der Methode (Zitationsorientierung), in der Besetzung der einzelnen Kategorien (relativ geringere Besetzung der höheren Klassen) wie in der Abstufung der einzelnen Kategorien (nichtlinear von 0,08 bis 1 im Gegensatz zu der linearen von VHB und Wien) deutlich von den anderen beiden Basislisten unterscheiden. Dieser Gegensatz wird durch die Imputation nicht nur nicht aufgehoben, sondern sogar noch etwas verstärkt. Dies liegt an den jeweiligen Lücken der Listen – bei CL und RbR liegen die in ihrem unteren Bereich, bei VHB und Wien in deren oberen Bereich.

Die vollständigen Listen mit 2822 Zeitschriften stehen im Internet zum Herunterladen unter www.vwl.uni-freiburg.de/iwipol/rankings.html zur Verfügung. Bei diesen Listen ist zu beachten, dass dort alle Zeitschriften aufgeführt sind, die in irgendeiner der verwendeten Listen genannt sind. Da einige der Ausgangs-Rankings sehr umfassend sind und auch nicht primär wirtschaftswissenschaftlich orientierte Zeitschriften enthalten, finden sich diese auch in unserem Zeitschriften-Ranking wieder.

5 Abschließende Bemerkungen

Die Messung individueller Forschungsleistungen ist schon allein deshalb ein inhärent schwieriges Geschäft, weil zum Zeitpunkt der Veröffentlichung oft noch nicht abzusehen ist, welchen Einfluss ein Artikel auf die Entwicklung der Profession haben wird. Allerdings wird eine systematische Erfassung von individuellen oder kollektiven Forschungsleistungen sich nicht den Luxus erlauben können, jeden einzelnen Artikel inhaltlich zu bewerten. Deshalb hat sich aus pragmatischen Gründen die Bewertung anhand der Qualität der Zeitschrift, in der der Artikel veröffentlicht worden ist, durchgesetzt. Dies ist offensichtlich eine Verkürzung in doppelter Hinsicht – zum einen findet Forschung auch in anderen Medien ihren Niederschlag, etwa in Sammelbänden, Büchern oder im Internet, zum anderen haben gute Zeitschriften auch mittelmäßige Beiträge und eher zweitrangig eingestufte Zeitschriften exzellente Artikel.²² Der Gutachterprozess ist eben alles andere als perfekt (Frey 2004) und die Wahl der Zeitschrift nicht nur durch ihr Ranking beeinflusst. Unter pragmatischen Gesichtspunkten scheint es jedoch der einzig gangbare Weg zur umfassenden Messung von Forschungsleistungen, zumal eine solche Bewertungsmethode die Anreize so setzt, dass die Forscher tatsächlich versuchen werden, ihre Forschungsergebnisse in möglichst guten Zeitschriften zu veröffentlichen.

Dies macht Zeitschriftenrankings zum zentralen Bewertungsmaßstab und damit zum Gegenstand fortwährender Diskussion. Eine typische Reaktion auf ein neues Ranking ist die Suche nach den Zeitschriften, in denen man veröffentlicht hat oder zu veröffentlichen hofft, um dann das Ranking als unsinnig zu verwerfen, wenn die Zeitschriften entweder nicht zu finden sind oder aus der Sicht des Betroffenen zu schlecht bewertet sind. In diesem Artikel haben wir versucht, das erste – berechnete – Einfallstor für Kritik zu schließen, indem wir Metarankings erstellt haben, die eine möglichst große Zahl von Journalen umfassen, so dass sie Volks- und Betriebswirten in möglichst allen Teildisziplinen gleichermaßen gerecht werden. Den zweiten Ansatzpunkt der Kritik können wir nicht beseitigen – alle Ranglisten haben notwendigerweise

²² Die Veröffentlichungen des Nobelpreisträgers Reinhard Selten sind ein gutes Beispiel für den letztgenannten Fall.

Schwächen – aber wir bieten mehrere Metarankings an, bei denen die Stärken und Schwächen in anderen Bereichen liegen, und wir machen diese Stärken und Schwächen transparent. So kann eine Diskussion um das „richtige“ Ranking auf eine rationalere Grundlage gestellt werden.

Bei aller berechtigter Kritik an ausufernden Evaluationen (Frey 2007) sind wir von der Notwendigkeit eines umfassenden (Meta-)Rankings überzeugt, um der Leistungsmessung über die Grenzen von Teildisziplinen hinweg die notwendige Akzeptanz zu verleihen. Wir stellen zwei Imputierungsmechanismen in unserem Papier vor, von denen wir glauben, dass sie methodisch wohlüberlegt sind. Dennoch sind andere Verfahren denkbar. Meta-Rankings können nur so gut sein, wie die ihnen zugrunde liegenden Ranglisten; wir haben die aus unserer Sicht geeignetsten Ranglisten verwendet, die jedoch alle Schwächen haben. Die Entwicklung umfassender Metarankings ist ein laufender Prozess – so ist nicht nur die Einbeziehung weiterer Ranglisten denkbar, um so die Zahl der erfassten Journale weiter zu erhöhen, sondern manche Ranglisten werden neuen Gegebenheiten angepasst – wie etwa die VHB Liste, die gegenwärtig revidiert wird. Damit entsteht die Notwendigkeit, auch die Metarankings zu aktualisieren.

Auch lassen sich die relativen Gewichte der verschiedenen Zeitschriftenkategorien (etwa A+ bis E) verändern, die bei der Aggregation von Zeitschriftenaufsätzen verschiedener Kategorien notwendigerweise angewendet werden. Statt Werten von 1,2, ...,6 für die Kategorien E bis A+ ließen sich auch bei Wiener, VHB oder der kategorisierten Ritzberger Liste den Kategorien Werte zuordnen, die die Unterschiede der Klassen anders – besser? – widerspiegeln. So sind Abstufungen wie von Combes und Linnemer verwendet auch dort denkbar. In jedem Fall macht die in diesem Aufsatz vorgestellte Methode umfassende ordinale Rankings möglich; wir bieten für vier verschiedene Basislisten solche Rankings an, die das Klassifizierungssystem der Basislisten beibehalten, aber Änderungen der Klassengewichte erlauben. Die aktuelle Version findet sich online zum Herunterladen unter www.vwl.uni-freiburg.de/iwipol/rankings.html.

Aus unserer Sicht führt an der Bewertung von Forschungsleistungen durch Ranglisten kein Weg vorbei, dennoch wollen wir davor warnen, die ermittelten Werte als exakte Messung der individuellen Forschungsleistung zu überinterpretieren. Wir schließen uns Palacios-Huerta und Volij (2004:964) an: "Citation analysis, however sophisticated it may be, cannot be a substitute for critical reading and expert judgment."

Literaturverzeichnis

- Bauwens, L. (1998): A New Method to Rank University Research and Researchers in Economics in Belgium, mimeo, CORE, Université Catholique de Louvain, Belgien.
- Blank, R. (1991): The Effects of Double-Blind versus Single-Blind Reviewing: Experimental Evidence from the American Economic Review, *American Economic Review*, 81: 1041-67.
- Bommer, R. und H.W. Ursprung (1998): Spieglein, Spieglein an der Wand: Eine publikationsanalytische Erfassung der Forschungsleistungen volkswirtschaftlicher Fachbereiche in Deutschland, Österreich und der Schweiz, *Zeitschrift für Wirtschafts- und Sozialwissenschaften*, 118: 1-28.
- Bräuning, M. und J. Haucap (2001): Was Ökonomen lesen und schätzen: Ergebnisse einer Umfrage, *Perspektiven der Wirtschaftspolitik*, 2: 185-210.
- Combes, P. und 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: 1250-1308.
- Coupé, T. (2003): Revealed Performances: Worldwide Rankings of Economists and Economic Departments 1969 – 2000, *Journal of the European Economic Association*, 1: 1309-1345.
- Dusansky, R. und C. Vernon (1998): Rankings of U.S. Economics Departments, *Journal of Economic Perspectives*, 12: 59-64.
- Frey, B. (2004): Publizieren als Prostitution? *Perspektiven der Wirtschaftspolitik*, 5: 333–336.
- Frey, B. (2007): Evaluierungen, Evaluierungen ... Evaluitis, *Perspektiven der Wirtschaftspolitik*, 8: 207-220.
- Garfield, E. (1972): Citation Analysis as a Tool in Journal Evaluation, *Science*, 178: 471-479.
- Geary, J., Marriott, L. und M. Rowlinson (2004): Journal rankings in business and management and the 2001 research assessment exercise in the UK, *British Journal of Management*, 15: 95-141.
- Hennig-Thurau, T., Walsh, G., und U. Schrader (2004): VHB-JOURQUAL: Ein Ranking von betriebswirtschaftlich-relevanten Zeitschriften auf der Grundlage von Expertenurteilen, *Zeitschrift für betriebswirtschaftliche Forschung*, 56: 520–545.
- Heining, J., Jerger, J. und J. Lingens (2007): Success in the Academic Labor Market for Economists – The German Experience, Universität Regensburg, mimeo.
- Kalaitzidakis, P., Mamuneas, T. und T. Stengos (2003): Ranking of Academic Journals and Institutions in Economics, *Journal of the European Economic Association*, 1: 1346-1366.
- Kodrzycki, Y und P. Yu (2006): New Approaches to Ranking Economics Journals, *Contributions to Economic Analysis & Policy*, Berkeley Electronic Press, 5, Article 24. <http://www.bepress.com/bejeap/contributions/vol5/iss1/art24>
- Laband, D. und M. Piette (1994): The Relative Impact of Economics Journals: 1970-1990, *Journal of Economic Literature*, 32: 640-666.
- Liebowitz, S. und J. Palmer (1984): Assessing the Relative Impacts of Economic Journals, *Journal of Economic Literature*, 22: 77-88.
- Lubrano, M., Bauwens, L., Kirman, A. und C. Protopopescu (2003): Ranking Economics Departments in Europe: A Statistical Approach, *Journal of the European Economic Association*, 1: 1367-1401.

- Mingers, J. und A.-W. Harzing (2007): Ranking Journals in Business and Management: A Statistical Analysis of the Harzing Dataset, Working Paper, University of Melbourne, forthcoming European Journal of Informations Systems, 16(4).
- Malouin, J.-L. und J.-F. Outreville (1987): The Relative Impact of Economics Journals: A Cross-Country Survey and Comparison, *Journal of Economics and Business*, 39: 267-77.
- Palacios-Huerta, I. und O. Volij (2004): The Measurement of Intellectual Influence, *Econometrica*, 72: 963-977.
- Rauber, M., und H.W. Ursprung (2008): Evaluation of researchers: a life cycle analysis of German Academic Economists, in M. Albert, D. Schmidtchen, S. Voigt (Hg.): *Scientific Competition*, Tübingen: Mohr Siebeck, forthcoming.
- Rauber, M., und H.W. Ursprung (2007): Life Cycle and Cohort Productivity in Economic Research: The Case of Germany, CESIFO working paper no. 2093, at http://www.cesifo.de/DocCIDL/cesifo1_wp2093.pdf .
- Ritzberger, K. (2008): Eine invariante Bewertung wirtschaftswissenschaftlicher Fachzeitschriften, *Perspektiven der Wirtschaftspolitik*, dieses Heft.
- Schulze, G., Warning, S. und C. Wiermann (2007): What (and how long) does take to get tenure? – The Case of German Economics and Business Professors, University of Freiburg, iep working papers, no. 4 (http://www.vwl.uni-freiburg.de/fakultaet/sopo/discussion_papers.htm).
- Theoharakis, V. und K. Axaroglou (2003): Diversity in economics: an analysis of journal quality perceptions, *Journal of the European Economic Association*, 1: 1402–1423.
- Ursprung, H. und M. Zimmer (2007): Who is the "Platz-Hirsch" of the German Economics Profession? A Citation Analysis, *Jahrbücher für Nationalökonomie und Statistik*, Band 227, Heft 4, im Erscheinen.

Anhang

A.1 Deskriptive Statistiken der Originallisten

Tabelle A1: Deskriptive Statistiken der Originallisten

<i>Variable</i>	<i>Mittelwert</i>	<i>Std. Abw.</i>	<i>Min</i>	<i>Max</i>
Rb	5,6361	10,9720	0,00	100,00
RbR	2,3257	1,5004	1	6,00
Wien	3,1353	0,9407	1,00	5,00
VHB	6,8688	1,5460	1,89	9,93
VHBR	3,4477	1,3800	1,00	6,00
CL	0,1324	0,1288	0,08	1,00
KMS	7,3169	16,0521	0,00	100,00
TA	10,8920	16,1031	1,20	91,64
BH	2,3233	0,7859	1,02	4,83

A.2 Die iterative Methode zur Imputation der Listen

Bei der iterativen Imputationsmethode wird die Basisliste jeweils auf alle möglichen Kombinationen von unabhängigen Variablen (andere Listen) regressiert. Weisen alle Variablen einen auf dem 10 Prozent Niveau signifikanten Einfluss auf, dann dient dieses Modell zum Imputieren fehlender Werte in der Basisliste, wobei größere Modelle, d.h. die mit mehr unabhängigen Variablen, immer bevorzugt zum Imputieren herangezogen werden. Infolge der relativ hohen Korrelation der unabhängigen Variablen untereinander handelt es sich hierbei um eine eher konservative Methode der Identifikation der Modelle.

Konkret bilden wir alle Permutationen aus allen oben beschriebenen Listen, die nicht die Basisliste sind: Das ergibt eine Regression mit sechs abhängigen, 6 über 5 Regressionen mit fünf abhängigen, 6 über 4 mit vier unabhängigen, 6 über 3 Regressionen mit drei abhängigen, 6 über 2 Regressionen mit zwei abhängigen und 6 über 1 Regressionen mit einer abhängigen Variablen, insgesamt 63 Regressionsgleichungen. Beginnend mit dem umfangreichsten Modell werden solange geordnete Probitschätzungen durchgeführt und schrittweise die Zahl der Regressoren verringert, bis in einem Modell alle Variablen signifikant sind. Bei mehreren Modellen mit der gleichen Anzahl von signifikanten unabhängigen Variablen wird zunächst das Modell mit dem höchsten Pseudo-R² ausgewählt, um den ersten Imputationsschritt auszuführen. Wir imputieren die Werte für die Zeitschriften, die in der Basisliste fehlen, aber in allen unabhängigen Listen enthalten sind anhand der geschätzten Parameterwerte, indem wir diesen Zeitschriften die Werte der Basisliste zuordnen, die die höchste Prognosewahrscheinlichkeit haben. Nach dieser ersten Imputation werden dann die durch Imputation verlängerte Basisliste auf alle übrigen Modelle regressiert, um mit denselben Kriterien zu prüfen, welches Modell im nächsten Schritt für die Imputation geeignet ist. Diese Iterationen führen wir solange durch, bis die Basisliste auf die Vereinigungsmenge aller in den Listen enthaltenen Zeitschriften aufgefüllt ist. Dabei bleiben die Werte der ursprünglichen Basisliste während des gesamten Verfahrens unverändert.

Die Ergebnisse des iterativen Imputationsverfahrens zeigen bei der paarweisen Korrelation mit den imputierten Listen nach dem randomisierten Verfahren eine sehr hohe Korrelation.