View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Research Papers in Economics

University of Konstanz
Department of Economics

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

Robert Hofmeister and Matthias Krapf

Working Paper Series
2011-37

http://www.wiwi.uni-konstanz.de/workingpaperseries

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

Robert Hofmeister and Matthias Krapf*

September 29, 2011

Abstract

Using data on the B.E. Journals that rank articles into four quality tiers, this paper examines 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. 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.

Keywords: Peer Review, Research Evaluation, Citations, Journal Quality.

JEL Classification Numbers: A10, A14.

---

*Hofmeister: University of Konstanz (robert.hofmeister@uni-konstanz.de); Krapf: University of Vienna (matthias.krapf@univie.ac.at). The authors are grateful to Florian Englmaier, Jan Häußler, Leo Kaas, Bertrand Koebel, Winfried Pohlmeier, Jörg Stoye, Heinrich Ursprung, Christian Zimmermann, two anonymous referees and the managing editor, Don Fullerton, for helpful comments and suggestions.

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

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

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 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 2 describes the particular set-up used in our analysis. Section 3 presents the data and descriptive statistics, and Section 4 the econometric results. Section 5 concludes.

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

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

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.

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

Table 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.

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

[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`.

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. For the remainder, we introduce an 'other JEL' variable, which captures JEL code categories A, B, M, N, P, R, and Z that all have less than 20 observations, respectively, as well as unclassified articles. Finally, table 1 shows that the number of articles published in all three journals has increased over time.

Table 2 presents descriptive statistics of the citation variables as found in the *IDEAS* 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

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

Table 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 2: DESCRIPTIVE STATISTICS: CITATION VARIABLES

criterion	(sub)set	observations	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
	Frontiers/Advances	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
	Frontiers/Advances	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

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.

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

[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.

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 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 Econometric Analysis

4.1 Determinants of Editorial Sorting

In this section, we analyze the determinants of editorial sorting. Table 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, 2007). 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 female authors is not significant in any spec-

[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.

ification. 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.

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 shows the results. In the Appendix we present robustness checks with raw citations and recursively weighted citations as dependent variables.

Specification (1) includes dummies for journals, tiers, the respective interaction terms, and publication year dummies. Articles published in the *BEJEAP* which

Table 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 D			0.121	0.0955
			(0.114)	(0.115)
JEL E			-0.207	-0.228*
			(0.137)	(0.138)
JEL F			-0.0654	-0.0702
			(0.169)	(0.170)
JEL G			-0.338	-0.326
			(0.209)	(0.210)
JEL H			-0.0296	-0.0662
			(0.154)	(0.155)
JEL I			-0.0231	0.0126
			(0.197)	(0.198)
JEL J			-0.128	-0.159
			(0.153)	(0.155)
JEL K			-0.0132	-0.0585
			(0.232)	(0.235)
JEL L			-0.00788	0.00605
			(0.142)	(0.143)
JEL O			-0.199	-0.228
			(0.138)	(0.139)
JEL Q			0.0921	0.133
			(0.203)	(0.205)
other JEL			0.117	0.115
			(0.153)	(0.155)
year dummies	no	yes	no	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.

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[9] 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 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 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

[9] See the Appendix for the results.

Table 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)
Africa			-1.079	-0.846
			(0.687)	(0.688)
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)
North America			-0.0341	-0.0381
			(0.0542)	(0.0553)
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)
JEL C			0.144***	0.0807**
			(0.0378)	(0.0388)
JEL D			-0.210***	-0.207***
			(0.0357)	(0.0357)
JEL E			-0.0414	-0.0267
			(0.0455)	(0.0464)
JEL F			0.120***	0.149***
			(0.0455)	(0.0456)
JEL G			-0.938***	-0.950***
			(0.0795)	(0.0801)
JEL H			0.0192	0.00875
			(0.0456)	(0.0456)
JEL I			-0.166**	-0.124*
			(0.0644)	(0.0651)
JEL J			0.172***	0.197***
			(0.0419)	(0.0417)
JEL K			0.384***	0.466***
			(0.0627)	(0.0638)
JEL L			0.349***	0.340***
			(0.0401)	(0.0404)
JEL O			0.0848**	0.120***
			(0.0368)	(0.0375)
JEL Q			-0.469***	-0.498***
			(0.0734)	(0.0743)
JEL other			0.0636	0.0492
			(0.0436)	(0.0436)
Constant	1.443***	1.324***	1.470***	1.103***
	(0.0573)	(0.0645)	(0.0767)	(0.0863)
year dummies	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

institutions attract significantly fewer citations. The coefficients of the geographic variables are somewhat different from those detailed in Table 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.[10]

Just as Ursprung and Zimmer (2007), we find that citation intensity varies significantly across fields. 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 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,

[10] Recall that no reference category for geographic location is required because the sum of their means is larger than one (see Table 1).

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.

4.3 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 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 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.

Table 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

For the *Frontiers* and *Advances* "journals" we find that 68.3% of all published articles 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.[11] Assuming that the *Frontiers*, *Advances*, and *Contributions*

[11] Suppose acceptance for publication was purely random. Then type I error would approach one and

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.

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),

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.

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.

References

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 (Jahrbcher fr Nationalkonomie 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.

A Appendix

Table 6 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 6: Poisson Estimates.

	cites	w cites simple	w cites recursive
	(1)	(2)	(3)
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)
Africa	-0.407	-0.846	-1.627
	(1.017)	(0.688)	(3.743)
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)
North America	-0.210*	-0.0381	0.0272
	(0.109)	(0.0553)	(0.189)
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)

	cites	w cites simple	w cites recursive
JEL C	-0.0505	0.0807**	0.0177
	(0.0776)	(0.0388)	(0.136)
JEL D	-0.144**	-0.207***	-0.145
	(0.0698)	(0.0357)	(0.123)
JEL E	0.0223	-0.0267	-0.0539
	(0.0951)	(0.0464)	(0.157)
JEL F	0.126	0.149***	0.203
	(0.0898)	(0.0456)	(0.155)
JEL G	-0.712***	-0.950***	-0.882***
	(0.150)	(0.0801)	(0.272)
JEL H	-0.108	0.00875	0.0156
	(0.0912)	(0.0456)	(0.159)
JEL I	-0.245*	-0.124*	-0.0734
	(0.129)	(0.0651)	(0.229)
JEL J	0.107	0.197***	0.119
	(0.0841)	(0.0417)	(0.147)
JEL K	0.225*	0.466***	0.378
	(0.130)	(0.0638)	(0.230)
JEL L	0.157*	0.340***	0.278*
	(0.0799)	(0.0404)	(0.142)
JEL O	0.0923	0.120***	0.0755
	(0.0754)	(0.0375)	(0.130)
JEL Q	-0.216*	-0.498***	-0.445*
	(0.121)	(0.0743)	(0.254)
other JEL	0.184**	0.0492	0.0501
	(0.0828)	(0.0436)	(0.153)
2001*BEJEAP	0.209	0.220***	0.241
	(0.164)	(0.0824)	(0.295)
2002*BEJEAP	-0.287	-0.613***	-0.532
	(0.246)	(0.143)	(0.520)
2003*BEJEAP	0.00755	0.0130	0.0955
	(0.132)	(0.0659)	(0.236)
2004*BEJEAP	0.148	0.0202	0.195
	(0.109)	(0.0573)	(0.202)
2005*BEJEAP	0.383***	0.337***	0.367
	(0.121)	(0.0626)	(0.228)
2001*BEJM	0.732***	0.644***	0.709**
	(0.181)	(0.0832)	(0.279)
2002*BEJM	-0.180	-0.497***	-0.307
	(0.227)	(0.120)	(0.383)
2003*BEJM	-0.213	-0.235***	-0.0382
	(0.185)	(0.0880)	(0.283)
2004*BEJM	0.663***	0.637***	0.692***
	(0.134)	(0.0631)	(0.217)
2005*BEJM	0.167	-0.142**	-0.0861
	(0.143)	(0.0717)	(0.244)
2001*BEJTE	0.709***	0.593***	0.822*
	(0.236)	(0.123)	(0.422)
2002*BEJTE	0.687***	0.466***	0.698
	(0.235)	(0.122)	(0.424)

	cites	**w cites simple**	**w cites recursive**
2003*BEJTE	0.415*	0.206*	0.304
	(0.226)	(0.120)	(0.422)
2004*BEJTE	0.540**	0.155	0.290
	(0.227)	(0.128)	(0.446)
2005*BEJTE	-0.179	-0.381**	-0.137
	(0.290)	(0.155)	(0.512)
Constant	0.132	1.103***	-1.578***
	(0.165)	(0.0863)	(0.306)
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