Research Policy 42 (2013) 295–301

Contents lists available at SciVerse ScienceDirect

Research Policy

journal homepage: www.elsevier.com/locate/respol

Peer review versus citations – An analysis of best paper prizes

Tom Coupé*

Kyiv Economics Institute at the Kyiv School of Economics, Yakira 13, 04119 Kyiv, Ukraine

ARTICLE INFO

Article history:
Received 5 July 2011
Received in revised form 27 April 2012
Accepted 16 May 2012
Available online 4 June 2012

JEL classification:
A 10

Keywords:
Peer review
Citations
Academic quality
Performance evaluation

ABSTRACT

In this paper, I analyze the 'best paper' prizes given by economics and finance journals to the best article published in their journal in a given year. More specifically, I compare the citations received by best paper prize-winning papers to citations received by papers that are awarded runner up prizes and to citations received by non-winning papers. In this way, I evaluate to what extent evaluation outcomes based on peer review correspond to evaluation outcomes based on citation counts. The data show that the paper that gets the 'best paper' prize, is rarely the most cited paper; is, in a small majority of cases, cited more than the runner up papers and is, in most cases, cited more than the median paper. I also explore whether characteristics of the prizes or the papers correlate with this difference in outcomes between peer review and citation counts and find there is no easy way to reduce the difference in outcomes between these two evaluation methods

© 2012 Elsevier B.V. All rights reserved.

1. Introduction

In an attempt to make the distribution of research money more performance based, several governments have developed research assessment systems that evaluate the research output of university departments. In the UK, for example, there is the Research Excellence Framework (the former Research Assessment Exercise), in Australia there is the Excellence in Research for Australia (the former Research Quality Framework) and in Italy, there is the Valutazione Quinquennale della Ricerca (VQR) (which replaced the Valutazione Triennale della Ricerca (VTR)). As illustrated by the name changes, many of these governments are still in the process of fine-tuning their assessment systems.

One of the important questions in this search for the ideal research assessment system is whether such system should be based on bibliometrics (i.e. citation counts) or peer review. Some systems include bibliometrics, like the Flemish system (Debackere and Glänzel, 2004), while other focus on peer review like the UK Research Assessment Exercise. That this question is controversial is best illustrated by the heated debates in the UK that were caused by the proposal to replace the peer review based Research Assessment Exercise by the Research Excellence Framework in which bibliometrics would play a much more important role (see for example, Corbyn, 2009). In the end, the Higher Education Funding Council (HEFCE) made the use of citation counts optional rather than imposed (see Richardson, 2011).

This paper contributes to this discussion by evaluating to what extent decisions based on peer review correspond to decisions based on citation counts. Both peer review assessments and citation counts can be considered as imperfect measures of the true academic quality that the assessment systems are trying to capture. Many authors indeed have illustrated the problems with peer review (see Bornmann, 2011 for a review). It has been shown that important papers are sometimes not recognized by reviewers (Gans and Shepherd, 1994), that the more productive economists do not necessarily get selected by the best universities (Smeets et al., 2006) or that not only quality matters in whether a grant proposal (Broder, 1993), a paper (Blank, 1991) or a candidate for honorary fellowships gets accepted (Hamermesh and Schmidt, 2003). Similarly, the use of citations to measure academic quality is not undisputed (see for example, Bornmann and Daniel, 2008 for a review). Citation counts are imperfect measures as, amongst others, they can be manipulated through self-citations, they include 'negative' citations (when a paper is cited as an example of how not to do something) and authors can cite selectively, only referring to works of their friends. Given that both peer review and citation counts measure academic quality with error, it is unlikely that an evaluation based on one of them will correspond perfectly to the evaluation based on the other. The questions I try to answer in this paper are, first, how different are the outcomes of an evaluation based on peer review from the outcomes of an evaluation based on citation counts and second, what factors can explain these differences.[1]

* Tel.: +38 044 492 8012; fax: +38 044 492 8011.
E-mail address: tcoupe@kse.org.ua

[1] Another interesting question is whether peer review or citation counting gives the best proxy for academic quality. To be able to answer this question, however,

0048-7333/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.respol.2012.05.004

This paper is not the first paper that studies the link between citations and peer review. Mahdi et al. (2008), for example, find that for most of the 2001 RAE assessment units (which broadly correspond to specific disciplines and university departments) there is a significant rank correlation between the 2001 RAE (peer review) ranking and a citation ranking based on publications submitted for the 2001 RAE. At the same time, they stress that a significant correlation does not necessarily mean a very high correlation as for many assessment units the rank correlation is only between 0.4 and 0.7. In this paper, rather than studying university-wide aggregates, I focus on the link between citations and peer review at the level of individual papers. This focus on individual papers not only generalizes the analysis that exists at the aggregate level but is also closer in spirit to the UK assessment system that in 2008 represented assessment scores in the form of 'quality profiles,' which gives the distribution of the 'research activity' of a given assessment unit over 5 different quality levels and hence implicitly requires the separate assessment of each specific 'research activity'.[2]

Studies that focus on the so called 'predictive power' of peer review (that is, the extent to which peer review 'predicts' citations) rarely study individual articles and if they do, mostly focus on medical journals (see Bornmann, 2010, 2011). In addition, these articles typically compare citations of papers published in a given journal to the citations of papers that were rejected by that journal but published in another journal. The problems with this approach are twofold. First, when papers are resubmitted to a different journal they typically are different from when they were submitted initially as authors will try to incorporate the comments of the initial set of referees. Second, the citations of an article might be influenced by the journal in which it has been published, making it hard to compare citation counts of articles from different journals. In my case, I avoid these problems by using data from best paper competitions, in which juries evaluate articles within a given journal.

Every year, several journals give a prize to the 'best' article published in their journal over a specific period of time. These best paper prize competitions are a good example of peer review as the jury consists of a number of scientists who are asked to evaluate all the articles published in the journal in a given time period. Most often peer review is used in a similar 'short-run' context–referees, when evaluating papers for publication, and senior faculty, when deciding about the hiring of assistant professors, do not have much information about how important a paper will become or how productive a job market candidate will be. Also in the context of peer review based research assessment systems, many papers that need to be evaluated will only be published shortly before the review process start and hence only very short citation windows will be available for these.[3] For example, in 2014 the peer review panels in the next UK Research Excellence Framework will evaluate publications in the period 2009–2013.

In this paper, I check whether papers that were selected 'best paper' by a given journal in a given year also turn out to be the highest cited paper among all papers published in that journal and competing for that year's best paper prize.[4] I use best paper competitions from the four top journals in finance, the Journal of Finance (JoF),[5] the Journal of Financial Economics (JFE), the Review of Financial Studies (RFS) and the Journal of Financial and Quantitative Analysis (JFQA). I chose the first three finance journals because in addition to the best paper prize, they also award runner up prizes to the second, and sometimes even, third best paper, which allows me to check not only whether best paper prizes are best in terms of citations but also to compare the citations of best paper prize winners and runner up prize winners. I add the fourth finance journal, and three economics journals with a long standing tradition of giving a best paper prize (the Journal of Economic History (JEH),[6] the Southern Economic Journal (SEJ) and the Canadian Economic Journal (CJE) to have a wider variety of characteristics of prizes and juries. This allows getting some tentative evidence about whether these characteristics influence the extent to which peer review and citation counts lead to the same conclusions.

I find that only in a small number of cases, the best paper is the most cited paper. I also find that in a large majority of cases, the best paper is cited more than the median paper in competition for the best paper prize, and that in a small majority of cases the best paper prize has a higher citation count that the runner up paper(s). This suggest that 'subjective' peer review will often coincide with 'objective' citation counts when distinguishing between highly cited and infrequently cited papers, but that differences between the two methods will be larger when a distinction has to be made among highly cited papers.

I do not find strong evidence that the difference between peer review and citation counts is related to characteristics of the prize, such as the amount of prize money or the number of jury members or how many years the prize has been awarded. There is some evidence, however, that using longer citation windows to count citations improves the match between the two evaluation methods if this match is measured in terms of the percentage of prize winning papers that are cited more than the median. As far as characteristics of the papers are concerned, I do find some evidence that the difference between peer review and citations counts is related to page length (with longer papers being more likely to have received a best paper prize after controlling for citations).

The remainder of the paper is organized as follows. Section 2 presents descriptive information about the best paper prizes in economics and finance. Section 3 focuses on the sample I use in this paper and present the basic analysis. In Section 4, I provide econometric estimates that model the chance a paper has won a prize as a function of the characteristics of that paper. Section 5 concludes.

2. The best paper prizes in economics and finance

Out of the 100 most cited economics and finance journals in 2009, 26 journals currently have a regular best paper prize.[7] An

one would need to know the true academic quality which is unknown. Note also that even if the citation counts and peer review give the same outcome, this does not necessarily mean that both reflect the same underlying true quality as there is the possibility that both could be biased in the same way.

[2] In previous RAEs, the assessment categories also reflected the need for each research activity to be evaluated separately, for example a 5* meant "Research quality that equates to attainable levels of international excellence in more than half of the research activity submitted and attainable levels of national excellence in the remainder."

[3] A recent study by Waltman et al. (2011) gives for several exact sciences the correlation between citations counts of specific articles over different time horizons. For mathematics, the citation count after the first year is relatively weakly correlated with the citations count after five years (around 0.33). Given that publication lags in economics and finance are longer than in mathematics (see Table 2 in Ellison, 2002) this correlation is likely to be even weaker for economics and finance journals.

[4] One could argue that the extra attention that a winning paper gets will increase its citation count. If this is the case, my results would be biased towards finding that citations and peer review give similar results. The extent of such bias is likely to be small however as the winners are announced only once, at an association meeting, in a newsletter or in an announcement in the journal itself. After that, one has to search really hard to find which paper won the prize. Indeed, for several journals I was not able to find all prize winners, even after an extensive Internet search. Hence, it is unlikely that such short run extra attention would significantly affect the long term citation count.

[5] The conference issues of the JoF are excluded from the competition and hence from the sample.

[6] I excluded the notes and discussions from the sample as they are not considered for the best paper prize.

[7] Using ISI's 2009 Journal Citation Reports. I combined the journals classified by ISI as 'economics' and 'business and finance' and deleted the pure

Internet search further revealed 19 more journals that have a best paper prize but are outside the top 100 economics and finance journals.

In economics and finance, awarding a prize is a relatively new phenomenon. The first prize in the sample of prizes appeared in 1960 (the Graham and Cobb award of the Financial Analyst Journal), the median year a journal awarded its prize for the first time is 1993. Only 17 prizes (out of 45) were first awarded before 1990. A typical journal (out of those awarding a prize) existed for 23 years before it started awarding a best paper prize.

There are several reasons journals have established these prizes. Some journals establish a prize to recognize the contributions of a specific economist. The Journal of International Economics for example writes: 'The award is intended to honor Jagdish Bhagwati for his many contributions to the field of international economics (2000, p13)'. About half of all prizes are named after an influential economist. Some of these are still alive (Fama-DFA prize, Jensen Prize) but most of them are named after dead economists and in some cases the death of that person is the direct reason to establish the prize (for example, the Richard Stone Prize of the Journal of Applied Econometrics or the H. Gregg Lewis Prize of the Journal of Labor Economics). Other prizes are named after the journal and sometimes a prize is named after the firm that sponsors the prize.

Some journals hope that having a prize will attract good papers. In 1990, the editor of the Economic Journal, for example, wrote: 'I hope that the existence of the prize will encourage high-quality submissions (1990, p. I)'. Or more general, they hope to stimulate research in their field like the Smith-Breeden Prize ('to promote excellence in research in financial economics', Journal of Finance, 1990, p. 1). Note that based on the above one could argue that, for a journal and its jury, a best paper is thus the paper that helps most to 'attract good papers', that does most 'to promote excellence' or that does most to honor the economist or firm after which the prize has been named.[8] If this is indeed the case, then some juries might sacrifice some pure academic quality to reach these non-academic goals, which provides another illustration that peer review, like citation counts, is likely to provide an imperfect measure of the true academic quality.

As mentioned in the introduction in this paper, I focus on the best paper competitions from four finance journals, the Journal of Finance (JoF), the Journal of Financial Economics (JFE), the Review of Financial Studies (RFS) and the Journal of Financial and Quantitative analysis (JFQA), and three economics journals, the Journal of Economic History (JEH), the Southern Economic Journal (SEJ) and the Canadian Economic Journal (CJE). As one can see from Table 1, there is quite some variation in the characteristics of these best paper prize competitions.

In most cases, the best paper is selected by a relatively small committee, most often consisting of editors. Exceptions are for example the Journal of Financial Economics where subscribers and the Review of Financial Studies where society members vote. Some juries are careful and select 'outstanding' papers, but most often one article is said to be the 'best' article among those that have been published during the previous year(s).

It is also worth noting that finance journals have prizes that carry substantially higher prize money. The largest amount of prize money is offered by the Journal of Finance. Since 1997, the

Table 1
The best paper prize competitions of JoF, JFE, RFS, JFQA, JEH, SEJ and CJE.

	JoF	JFE	RFS	JFQA	JEH	SEJ	CJE
Competition	Smith Breeden prize/Brattle prize	Jensen prize/Fama-DFA prize	Barclays Global Investors Michael Brennan award	William F. Sharpe award	Arthur Cole prize	Georgescu Roegen prize	Harry Johnson prize
Aim	"For the top three papers in JoF in any area other than corporate finance"/"for outstanding papers on corporate finance"	"Best paper in corporate finance and associations"/"Best paper in capital markets and asset pricing"	"Rewards important research in the field of finance", "To the best paper published in RFS"	"To foster excellence in financial research", "For the best article published each year"	For the best article	For the best academic article	For the best paper published
First awarded in	1989/1999	1997/1997	1989	1999	1966	1987	1978
Procedure	Associate Editors of JoF give top 3 – then votes are summed	Vote by subscribers	Selection by Executive Editor, Co-Editors, and Associate Editors of RFS	Nomination by JFQA readers, subscribers, and Associate Editors, followed by selection by Associate Editors.	Selection by the Board of Editors	The Editor of the SEJ, and two Vice Presidents of the Southern Economic Association.	Selected by a committee of three
Prize amount ($)	10,000/5000 (*2)	5000/2500	20,000/7000	5000	0	An engraved plaque and a set portion of the earnings on the Georgescu-Roegen endowment	5000
Median Citations 1997–2007[a]	22	18	12	6	3	2	3

[a] The median number of citations for articles in the journal that were published in the period 1997–2007.

accounting and business journals. I allocated the prize of the European Economic Association to its new journal, the Journal of the European Economic Association rather than to its earlier journal, the European Economic Review. The special issue prize of the Journal of Corporate Finance and the former Herman Daly Prize of 'Ecological Economics' are not included in the number 26. The list of prizes can be found on https://docs.google.com/leaf?id=0B_QokXC4wUSYNjEwZWIxYjgtNTZjYS00NTBkLWJiZjMtZjQ0ZTg1NTliZGZh&hl=en.

[8] Very few journals explain why a paper is selected to be the best.

Smith-Breeden prize has been worth $10,000 for the winner and $5000 for each of the two 'distinguished' papers. More recently, an additional prize has been established to award the best paper on corporate finance. This 'Brattle Prize' is worth $10,000 for the winning paper and $5000 for the second paper. The Journal of Financial Economics is also very generous: both the Jensen Prize and the Fama-DFA prize give $5000 to the first and $2500 to the second paper. The 'Barclays Global Investor best paper award' of the Review of Financial Studies at $20,000 represents the richest best paper prize in economics and finance. In comparison, the prize money of the economics journals is negligible; the Journal of Economic History's Arthur Cole prize is even purely symbolic.[9]

Finally, in addition to the difference in prize money, the finance journals also have higher median citations than the economics journals. At the same time, the economics journals in the sample started awarding prizes earlier.

3. Data and descriptive results

In my sample, I have 138 'competitions' which were organized in the period 1987–2007. I do not take into account the more recent competitions as I want to give sufficient time for the citation counts to become meaningful. Citation counts were obtained in the period July–August 2010, hence the citation counts cover periods from 2.5 years to 21.5 years. These 138 competitions come from 118 different 'volumes', one year periods for specific journals that include all articles that participate in a given competition. Twenty of these volumes had 2 competitions simultaneously (each for a different subarea, like the Jensen Prize and the Fama Prize in the JFE), so I have 98 single competition volumes. The advantage of single competition volumes is that there is one prize winner that can be compared to the highest cited article. In case there are two prizes, each for a subarea, it is hard to determine to which subarea the highest cited article belongs.[10] Given that sometimes there are joint first prize winners, and that in many cases, several runners up are identified, I have more prize winners than competitions, respectively 143 first prize winners, and 88 runner up winners. Table 2 gives data disaggregated by journal.

For each article, I collect citation count data from the ISI's Web of Science.[11] Table 3 uses these citation data from the finance journals in the sample to give descriptive statistics on the main questions of interest.

From Table 3, one can see that the best paper prize winner of the finance journals in the sample is rarely the highest cited article: 10% of the best paper prize winners of the single competition volumes and 25% of the winners of double competition issues turned out to be the highest cited article.[12] At the same time, these probabilities are clearly higher than what one would get if the jury randomly chose the best paper article: as the median single competition has 39 competitors and the median double competition (with two 'best' papers) 77 competitors, in both cases the chance to randomly pick the highest cited paper is about 2.5% (1/39 and 2/77 respectively). Note further that also several runner up papers turned out to be the highest cited paper.

For the single competition volumes, the best paper prize juries selected in 65% of the competitions a best paper prize that has a higher citation count than the runner up paper(s). For double competition volumes, this is lower at 51%. Again this is better than random: if one would pick randomly a paper out of two papers, in 50% of the cases one would pick the highest cited paper. If one has three papers, like in the case where one has one winner and two runners up, the probability that a randomly picked paper is the highest cited paper decreases to 33%. Based on the competitions in the sample that have a varying number of runners up prizes, the average probability of randomly picking the highest cited paper out of the set that consists of the winning and runner up papers is 42%.

The Editor's report from the JoF provides us with additional information. At the JoF, the voting is held in two rounds. In the first, all associate editors can nominate papers, and then the associate editors have to select among the nominated papers. From the Editor's report I also have information about which papers were nominated, that is, thought to be among the top 3 papers[13] by at least one associate editor. As an example, in 1989, 26 associate editors nominated 23 different papers. Overall, a striking 43% of all papers get nominated. Of these nominated papers, 70% score better than the median cited paper of their respective competition. In 76% of the competitions, the highest cited paper is included among the nominated papers.

Table 4 gives the statistics for the three economics journal, all of which only give one prize per year. I get similar results as for finance journals: the best prize winner is rarely the most cited paper but is often more cited than the median paper.[14]

Based on the descriptive statistics so far, one can conclude that if one focuses on the 'best' paper, there is quite some difference between the peer review outcome and the citation-based outcome. Also, one can conclude that these two methods coincide more often when using a lower threshold. In about 75–80% of the competitions, awards are given to papers that have more citations that than the median number of citations of their competitors.

When looking at possible explanations for this difference in outcomes based on citations and based on peer review, I find little evidence that higher stakes in terms of prize money improves or deteriorates the match between both evaluation methods. From one side, higher stakes could stimulate the jury members to take their task more seriously. From the other side, higher stakes also means that the jury members have more reasons to bias their decisions. I do find that the RFS and the JoF, which offer the highest prize money, have slightly more prize winning papers that are cited more than the median, compared to the other finance journals. At the same time, the JFQA and JFE offer less prize money but they have more prize winners that also were the highest cited article. In addition, the SEJ has substantially less money to offer than the finance journals but still does as good as the best finance journals in terms of prize winning papers that are cited more than the median.

Similarly there is little evidence that the size of the selection committee is crucial. The JFE lets its numerous subscribers decide, while the SEJ has only a 3 person jury consisting of the Editor and the two Vice Presidents of the Southern Economic Association. Both journals do as well in terms of the share of prize winning papers that are cited more than the median. The score on this criterion also does not seem to be related to how long the prize exists; the young JFQA prize does as good (or bad) as the old JEH and CEJ prizes.

[9] About three-fourths of the prize-giving journals in the full sample of 45 journals have a monetary prize for the best paper, with 10% of all prize giving journals, mainly the finance journals, also giving monetary prizes for the second or third place. Conditional on giving a prize, the mean amount of money for the first prize winner is about $3000 (median $2000).

[10] Papers can be and have been nominated for both categories.

[11] The version of WoS the author has access to have citation data from 1987 onwards.

[12] Deleting self citations is unlikely to change this as the gap between the most cited paper and the citations of the best paper prize is typically more than 10%. In fact, the median prize winner has 23% of the citations of the most cited paper.

[13] Each editor can nominate three papers. The average ratio nominated papers on associated editors is about 0.7. From 2006 onwards the report gives the list of 'finalists', which seem to be the most frequently nominated papers, rather than of 'nominations', so I exclude 2006 and 2007 from these calculations.

[14] Given that the median number of articles in these economics competitions is 55, the chance to pick the most cited paper at random is less than 2 percent.

Table 2
The composition of the sample of best paper prizes.

	Overall	JoF	JFE	RFS	JFQA	JEH	SEJ	CJE
Number of competitions	138	28	22	18	9	20	20	21
Number of different volumes	118	19	11	18	9	20	20	21
Number of volumes with 2 competitions	20	9	11	0	0	0	0	0
Number of volumes with 1 competition	98	10	0	18	9	20	20	21
Number of competitions with first prize and runner up prize	59	28	22	9	1	0	0	0
Number of first prize articles	143	30	23	19	10	20	20	21
Number of runner up articles	88	49	28	9	2	0	0	0

Table 3
Peer review versus citation counts for finance journals (numbers are percentages).

	Overall	JoF	JFE	RFS	JFQA
For single competition volumes					
Is any of the first prize winners the most cited paper?	0.10	0.1	–	0.05	0.2
Is the first prize paper more cited than the runner up prize paper(s)?	0.65	0.5	–	0.78	1
Is any of the runners up the most cited paper?	0.15	0.2	–	0.11	0
For double competition volumes					
Is any of the first prize winners the most cited paper?	0.25	0.11	0.36	–	–
Is the first prize paper more cited than the runner up prize paper(s)?	0.51	0.39	0.62	–	–
Is any of the runners up the most cited paper?	0.10	0.11	0.09	–	–
Overall					
Is an award given to an article that is cited more than the median article	0.81	0.86	0.78	0.79	0.67

Table 4
Peer review versus citation counts for economics journals (numbers are percentages).

	Overall	JEH	SEJ	CJE
Is any of the first prize winners the most cited paper?	0.066	0.15	0.05	0
Is an award given to an article that is cited more than the median article	0.75	0.7	0.85	0.71

Table 5
Peer review versus citation counts for different time periods (numbers are percentages).

	Finance	Finance < 2001	Finance > 2000	Econ.	Econ. < 1997	Econ. > 1996
Is an award given to an article that is cited more than the median article	0.81	0.90	0.73	0.75	0.83	0.68
Is any of the first prize winners the most cited paper? (single volumes)	0.1	0.08	0.14	0.066	0.067	0.065
Is any of the first prize winners the most cited paper? (double volumes)	0.25	0.17	0.29			

The median year of publication for the finance journals in the sample is 2001, for economics journals 1997

I do however find that the match between citation counts and peer review is somewhat better for the prizes that have been awarded for a longer period, at least in terms of the percentage of prize winning papers that are cited more than the median. Table 5 shows that if I split the sample of finance journals into two time periods based on the median publication year in the sample, I find that the percentage of prize winning papers that are cited more than the median is higher in the earlier period (0.9) than in the later period (0.73). The same is true for economics journals (0.83 versus 0.68). In fact, this is true even if I disaggregate to the level of the journals, for all journals except the JFAQ (which started awarding prizes only in 1999 and hence has only few papers receiving a prize before 2001). This is consistent with the idea that the correspondence between peer review and citation counts is bigger in the long run when citation counts have converged to their 'steady state'.[15] At the same time however, I do not find, in the earlier periods a higher percentage of best paper prize winners that also are the highest cited paper. In the earlier period, finance journals even score worse on this criterion.

So far I have checked whether the characteristics of the prize (discipline, amount of prize money, size of jury, years in existence or time when awarded) are related to how well outcomes based on citation counts correspond to outcomes based on peer review. In the next section, I abstract from these differences between prizes and instead will look at whether characteristics of papers can explain the difference between peer review and citation counts. More specifically, I will use regression analysis to study, within given competitions, what variables, besides the number of citations, correlate with the chance to have been awarded a best paper prize.[16]

4. Econometric results

Hamermesh and Schmidt (2003) show that in a regression that models the chance of being selected as an Econometrica fellow, characteristics like location and area of specialization matter in addition to citations. In a similar spirit, I first ran a regression that uses the citation count to explain which article wins the best paper prize, and then check whether other variables can add explanatory power. This will tell us what characteristics of papers correlate with

[15] It is also consistent with juries in the past being 'better' or papers in the past being easier to judge.

[16] Alternatively, one could also look at what are the determinants of citation counts, other than being awarded a prize. Given the focus on best paper prizes, I chose to use winning a prize as a dependent variable and citation count as an explanatory variable.

Table 6
Explaining the chance to win a prize.

	Winning Paper	Winning Paper	Best/Runner Up
	(1)	(2)	(3)
Highest cited	1.114	1.02	0.038
	(0.23)	(0.05)	(0.1)
Total cites/100	2.048***	1.863***	0.489***
	(5.09)	(4.42)	(5.82)
# Pages		1.064***	0.053***
		(5.89)	(6.64)
Two authors		1.02	−0.060
		(0.11)	(−0.38)
Three authors		0.699	−0.382*
		(1.26)	(−1.8)
Four authors or more		1.93	0.051
		(1.51)	(0.13)
Pseudo R^2	0.045	0.078	
# Groups	118	118	118
# Observations	6252	6252	6252

I ran conditional regressions in (1) and (2); while for (3) I used a rank order logit specification. Odds ratios are given for (1) and (2), coefficients for (3), T-stats are in parenthesis. The omitted category is papers with one author.

* Means significant at the 10% level.

*** Means significant at the 1% level.

the difference between outcomes based on citation counts and peer review.

More specifically, I start by regressing a dummy that reflects whether a paper wins a prize (first or runner up) on the citation count (divided by 100) and on a dummy that is one for the highest cited papers. I use a conditional logit regression specification, thus controlling for competition (volume-journal) specific differences and focusing on the within competition dimension.[17]

I then add, as explanatory variables, 2 easily observable characteristics of the papers, the number of pages and the number of authors. Both of these variables could potentially affect the probability a paper wins a prize once one controls for citation counts. The number of authors can be correlated with the difference between peer review and citations for two reasons–first, given there are more coauthors it is more likely that one of the jury members is linked to one of the coauthors and hence can be biased in favor of giving the best paper prize to that paper. If this is the case, one would expect, after controlling for the citation count, a positive coefficient of the number of coauthors. For a given number of citations, a paper written by more coauthors would then have a higher chance to win the prize. From the other side, more coauthors means the citation count can be more inflated by citations of friends–hence, after controlling for the total number of citations, the number of coauthors could have a negative effect of getting the best paper prize. In other words, by including the number of coauthors, I check whether jury members were influenced relatively more, as compared to citing authors, by the number of coauthors. In a similar way, the second variable I add, the number of pages will reflect who attached a larger weight to the length of the paper, the jury or the citing authors.

A third specification uses the rank ordered logit model (Beggs et al., 1981), which allows the dependent variable to be different for best papers, runner up papers and non-winning papers (Table 6).

I find that total citations increase the chances of the paper having been awarded a prize in a best paper competition, with the odds ratios approximately doubling for each extra 100 citations. The significant coefficient on the citation count also implies that the highest cited paper has the highest chance to have won a prize, even though the insignificance of the coefficient on the dummy for the highest cited paper indicates there is nothing extra coming from the status of being the highest cited paper per se. Noteworthy is the low pseudo R^2, which indicates there are many other reasons other than the citation count that, explain why some papers got a prize and others do not. I further find evidence that even after controlling for the number of citations, longer papers have a substantially higher chance to have won a prize, indicating that paper length influences jury members relatively more than citing authors. There is little evidence, however, that the number of authors matters.

As a further check, I next limit the sample to prize winning papers and compare best paper winners and the runners up. The advantage of this is twofold. First, this allows me to use only direct competitors in a conditional logit regression. Remember that for volumes with two competitions I do not know which papers of a volume should be allocated to which competition. Second, for this smaller sample of 149 papers, I can more easily collect additional information. I collect several additional explanatory variables following Hamermesh and Schmidt (2003). The percentage of authors who are located in the US, respectively are female, allows me to check whether after controlling for citations, the jury members were influenced by the location of the authors or by their gender.[18] A dummy variable, based on the abstract, of whether the paper is theoretical, empirical or combines theory and empirics allows me to check whether juries are influenced by the focus of the paper, again after controlling for citation counts.

I also collect information on the citations of each prize winning paper in Google Scholar, providing me with an alternative indicator of the impact of a paper.[19] I find that the correlation between Google Scholar citations and ISI citations is high at 0.98. However, only for 52 out of 59 competitions which have both best papers and runners up, the ordering based on Google citations and ISI citations is the same (Table 7).

In all 4 specifications, I find that being the highest cited paper among the winning papers increases the odds to having received the best paper prize rather than the runner up prize, with the odds ratio being 1.7–1.85. Total citations are found to be insignificant, meaning that what counts is being higher ranked, rather than how much more one is cited. Note that this is the opposite of what I found in the full sample (Table 5) where I compared winning to non-winning papers and found that what matters is the citation count rather than being the highest cited paper.

Point estimates further suggest that having female authors reduces the chances to have received the best paper prize (relative to receiving the runner up prize) while page length, having US affiliated scholars or having an empirical focus improves those chances, though none of these factors are found to be significant.

5. Discussion

In this paper, I documented the difference in evaluation outcomes based on peer review and citation counts. I find that award winning papers have a significantly higher number of citations, that they have significantly more chance to be cited more than the median paper in a volume but also that in less than a quarter of the cases, the best paper is the highest cited paper. I also find in a small majority of cases the best paper prize has a higher citation count that the runner up paper(s). These findings suggest that 'subjective' peer review will often coincide with objective citation

[17] Thus I control for differences between journals and between years.

[18] Like in the case of the number of pages or coauthors, given that citations themselves could also be influenced by the location and gender, one could state that I check whether jury members are influenced relatively more, as compared to citing authors, by the location of the authors of the paper or by their gender.

[19] The use of Google Scholar data to measure the impact of papers is somewhat controversial as Google Scholar mainly relies on citations in documents that are freely accessible and cannot count citations in papers available in fee-based journals published by major publishers (see Bornmann et al., 2009).

Table 7
Explaining which paper wins and which paper becomes a runner up.

	Best paper (1)	Best paper (2)	Best paper (3)	Best paper (4)
Highest ISI cited	1.815*	1.853*		
	(1.85)	(1.80)		
Highest Google cited			1.745*	1.69
			(1.69)	(1.49)
Total ISI cites/100	1.073	1.025		
	(0.45)	(0.15)		
Total Google cites/100			1.028	1.021
			(0.73)	(0.54)
Theory		0.753		0.727
		(−0.69)		(−0.76)
Empirics and theory		0.422		0.465
		(−1.25)		(−1.13)
Share US affiliated		4.835		4.075
		(1.4)		(1.26)
Number of authors		1.093		1.109
		(0.34)		(0.4)
Share female authors		0.798		0.740
		(−0.27)		(−0.34)
# Pages		1.016		1.015
		(0.59)		(0.53)
Pseudo R^2	0.06	0.11	0.07	0.11
# Groups	59	59	59	59
# Observations	149	149	149	149

We run a conditional logit regression. Numbers in the table are odds ratios. T-stats are in parenthesis.

* Means significant at the 10% level.

counts when distinguishing between highly cited and little cited papers, but that differences between the two methods will be larger when a distinction has to be made among highly cited papers.

I also investigated whether the extent of the difference between the outcomes of these two evaluation methods depend on the characteristics of the prize like the discipline (finance versus economics) in which it is awarded, how many years the prize has been awarded by the journal, the amount of prize money and the age of the journal that awards the prize. I did not find evidence in the data that these matter but I did find that prizes that were awarded earlier and hence for which a more long run citation count is available, were more likely to be cited more than the median, than prizes that were awarded more recently. This is consistent with the idea that in the long run the difference between peer review and citations will be less important than it would look like in the short run. At the same time I did not find that in the longer run the most cited paper was also more likely to be the best paper prize winner.

I further found that some characteristics of the papers themselves, other than citations, influence the chance to have won a prize. For example, I find that not only more cited papers but also longer papers have a significantly higher chance to have won a prize in a paper competition, while I find no evidence that the number of authors matters. For the choice between a runner up paper and a best paper, the highest cited paper is more likely to have won the best paper prize, though how big the difference in citations is, is not important, nor are the characteristics of the authors of the paper or the page length.

The interpretation of what I have documented depends crucially on what one believes to be true academic 'quality'. If one believes that expert opinion is the correct measure, then one can interpret the above results as showing that citation counts do not reflect academic quality that well, especially among highly cited papers. This would have important policy implications, for example in the case of UK's Research Assessment Framework, which foresees that citation counts will be made available to some expert panels. My findings could be used to argue that providing such information could endanger the correctness of the classification of departments at the top of the quality distribution as it would allow jury members to attach an unjustified weight to the highest cited articles.

If one believes in citation counts reflecting the true impact of a paper, one can interpret the above results as indicating that experts either cannot distinguish quality ex ante that well or even that the above results show the biased judgments of experts. In the framework of UK's research Assessment Framework, my results thus could be used to argue for a bigger role of citation counts and a reduced role of the expert panels.

My preferred interpretation is situated somewhere between the two above more extreme interpretations: the difference in outcomes should be seen as an illustration that citations and peer review both measure true academic qualities with error. As this paper has shown, neither changing the characteristics of the peer review mechanism used nor correcting for specific characteristics of papers (at least for those characteristics of papers and prizes tested in this paper) offer a clear and easy way to reduce the gap between these evaluation methods.

Acknowledgments

I thank Victor Ginsburgh and Daniel Hamermesh for helpful comments on an earlier version of this paper, Jim Davis for proofreading the paper, the Belgian federal government for financial support (PAI 4/28) and seminar participants at the Kyiv School of Economics. Any remaining errors are mine.

References

Beggs, S., Cardell, S., Hausman, J., 1981. Assessing the potential demand for electric cars. Journal of Econometrics 17, 1–19.
Blank, R., 1991. The effects of double-blind versus single-blind reviewing: experimental evidence from the American economic review. American Economic Review 81, 1041–1067.
Bornmann, L., 2010. Does the journal peer review select the best from the work submitted? The state of empirical research. IETE Technical Review 27, 93–95.
Bornmann, L., 2011. Scientific peer review. Annual Review of Information Science and Technology 45, 199–245.
Bornmann, L., Daniel, H.-D., 2008. What do citation counts measure? A review of studies on citing behavior. Journal of Documentation, 45–80.
Bornmann, L., Marx, W., Schier, H., Rahm, E., Thor, A., Daniel, H.-D., 2009. Convergent validity of bibliometric Google Scholar data in the field of chemistry – citation counts for papers that were accepted by Angewandte Chemie International Edition or rejected but published elsewhere, using Google Scholar, Science Citation Index, Scopus, and Chemical Abstract. Journal of Informetrics, 27–35.
Broder, I., 1993. Professional achievements and gender differences among academic economists. Economic Inquiry, 116–127.
Corbyn, Z., 2009. Hefce Backs Off Citations in Favour of Peer Review in REF. Times Higher Education.
Debackere, K., Glänzel, W., 2004. Using a bibliometric approach to support research policy making: the case of the Flemish BOF-key. Scientometrics, 253–276.
Ellison, G., 2002. Evolving standards for academic publishing: a Q-R theory. Journal of Political Economy, 994–1034.
Gans, J., Shepherd, G., 1994. How the mighty have fallen: rejected classic articles by leading economists. Journal of Economic Perspectives, 165–179.
Hamermesh, D., Schmidt, P., 2003. The determinants of econometric society fellows elections. Econometrica, 399–407.
Mahdi, S., D'Este, P., Neely A., 2008. Citation counts: are they good predictors of RAE scores? A bibliometric analysis of RAE 2001. AIM Cranfield School of Management Working paper.
Richardson, M., 2011. The research excellence framework: revisiting the RAE. Research trends, March. http://www.researchtrends.com/issue22-march-2011/the-research-excellence-framework-revisiting-the-rae-2/read on 26/04/2012.
Smeets, V., Warzynski, F., Coupé, T., 2006. Does the academic labor market initially allocate new graduates efficiently? Journal of Economic Perspectives, 161–172.
Waltman, L., van Eck, N., van Leeuwen, T., Visser, M., van Raan, A., 2011. Towards a new crown indicator: an empirical analysis. Scientometrics, 467–481.