Inefficient allocation of resources in science

Adam Narkiewicz

April 2014

Abstract

Researchers' incentives to publish in top-tier journals are examined by a theoretical model. Using incentives to deliver optimal quality of research as a benchmark, the model concludes that the former incentives result in misallocation of resources and welfare loss. The main reasons behind it are lack of unpublishable but otherwise valuable research, misalignment between effort and value of research, and suboptimal choices of research topics. A bibliographic data set is used to empirically test predictions of the model. Indeed, while choosing a field of study, researchers tend to be more focused on their ability to publish rather than on their ability to deliver valuable research. Alternative methods of evaluating research output are suggested to mitigate these problems.

JEL codes: A11, J24, L15

1 Introduction

Measuring impact of scientific output has been increasingly becoming a subject of a heated debate. On one hand, nominal academic research spending increased from \$18.9 billion in 1992 to \$65.8 billion in 2012 in the United States alone (see NSF, 1993, and NSF, 2013). This is an average increase per year of 6.4% in nominal and 4% in real terms. On the other hand, advances in information technology and availability of data allow for better understanding of impacts research projects have. This in turn results in growing popularity of measures like impact factors or h-indices.

Measuring value of research output is a challenging task with a numerous obstacles. Nevertheless, despite their shortcomings, various measures of research impact and quality are used commonly, and they support decisions that shape allocation of resources in science (see for example Hirsch, 2005). The quality of these measures is thus of great importance as they guide choices governing a multi-billion dollar markets.

These choices include, for example, answers to following questions: What research topics should receive more grants? Who should receive grants? Whom to employ at our research department? Whom to promote? What salary to offer? Answer to each of these questions usually depends on a multitude of factors, but almost always there are some numerical measures involved that allow for a straightforward comparison between two options (see Stephan, 2010, for an overview of measures used to allocate grants).

In this article I focus on a measure which is popularly used to justify hiring and salary decisions at research departments all over the world. The measure in question is a number of papers published in top-ranked journals. The goal of this paper is to propose and give credibility to the notion that this measure results in a misallocation of resources. In particular, that it results in a suboptimal amount of effort a researcher invests into a research project as well as suboptimal choices of research topics. To achieve this objective I build a theoretical model and the empirically test some of its predictions.

The question of essence is whether people responsible for these decisions really take the aforementioned measure into account. A sizeable research literature explicitly or implicitly assumes that productivity of a researcher is described by his or her ability to publish in top-ranked journals (see for example Laband, 1985; Levin and Stephan 1991; Oster and Hamermesh, 1998; Mittal, Feick and Mursched, 2008; O'Keefe and Wang, 2013). Furthermore, according to Gomez-Mejia and Balkin (1992) indeed such a measure can be in some cases the most important determinant of faculty's pay as opposed to other potential measures. Singh, Haddad, and Chow (2007) as well as Franzoni, Scellato, and Stephan (2011) present a lot of anecdotal evidence in favor of this claim and Oswald (2007) provides details on why using numbers of publications in top-tier journals can be wrong.

Is there any measure that could be employed in place of the number of publications in top-tier journals? Number of citations per paper is often suggested as a good proxy for paper's quality (see Medoff, 2003, for comprehensive discussion). Various derivatives of this measure are increasingly often used to evaluate researcher's output; especially Hirsch (2005) inspires more and more followers. Hamermesh, Johnson, and Weisbrod (1982) as well as Sauer (1988) present evidence that numbers of citations can be used as a determinant of faculty pay in economic institutions. I argue that use of these measures is superior to use of journal-based measures.

Therefore, throughout this paper the following simplifying assumptions are made:

1. Article's quality, impact, and value are synonymous and can be measured by number of citations to this article. A research project is and endeavor aimed at delivering an article.
2. A top-tier journal is a journal with a high-ranking impact factor and a publication in such a journal is a certification of the quality of an article.
3. Number of publications in a top tier journal is a measure that is employed to reward faculty. This measure is at least dominant among other measures at least within some groups of researchers. Rewards can be pecuniary (e.g. salary, employment opportunities) or non-pecuniary (e.g. prestige).

The paper consists of six sections. The second section describes a simple game-theoretic model built on the above assumptions which is aimed at identification of potential sources of welfare loss. The following sections of the paper contain data analysis aimed at testing whether predictions of the theoretical model can be confirmed empirically. Section three explains how the theoretical model from the section two can be tested empirically. Section four describes source of data and how data were obtained and justifies use of particular variables. Section five presents and discusses results of estimation. Finally, the sixth section discusses general results and suggests policy implications.

2 Theoretical framework

To illustrate potential sources of welfare loss, let's introduce a simple game. Assume that researchers are evenly distributed over a unit square. There are two fields of study (let's name them X and Y) and in each field of study there is one journal publishing only papers that are related to its field of study.

A researcher is described by a pair $(x, y) \in [0,1] \times [0,1]$. Researchers are heterogeneous with respect to ability to deliver socially valuable research. For a researcher at (x, y) cost of delivering research of value v is $c_x = xv^2$ for the field X and $c_y = \alpha yv^2$ for the field Y. Coefficient α denotes relative easiness of doing research in the field Y against the field X. If $\alpha = 1$, it is equally easy to do research in both fields. α is the same for all researchers.

For simplicity, let's assume that every researcher is involved in up to one research project and every research project relates to only one field of study. Each journal serves up to 1/3 of the total population of researchers. Therefore, there will always be at least 1/3 of researchers whose work has not been published. Journals seek to maximize total social value of the research they publish, given their limited space.

Assume that the researchers' payoffs depend on journals' reputations (R_x and R_y respectively for fields X and Y). Therefore, a researcher's utility is $U = R_x - xv^2$ or $U = R_y - \alpha yv^2$ depending on value of v she chooses and whether she chooses to publish in the field X or Y.

In the first stage of the game, researchers choose the field and the value of research they are going to deliver. In the second stage journals accept research for publication and researchers receive payoffs.

The equilibrium of this game is depicted on Figure 1. The figure was created using parameters $R_x = 1$, $R_y = 1.3$, and $\alpha = 0.9$. Researchers who got admitted to publish in the first journal (field X) are in the area ZDXW. Researchers who got admitted to publish in the second journal (field Y) are depicted by area ABYWX. A third of the researchers (area WYCZ) do not involve in any research project. All researchers who manage to publish in a particular journal deliver exactly the threshold value that is required to get a publication, because further increase of value of their research yields higher costs without higher benefits.

Now, let's tweak the game so that the researchers' payoffs depend on value of the research and equal either $U = v - xv^2$ or $U = v - \alpha y v^2$ depending on the field chosen. Then, each researcher produces research even though not all of it gets published. Assuming that a social planner would take total value of research as output and cost of conducting research as input, this game yields an outcome that is socially optimal. Figure 2 depicts welfare loss induced by the payoff being based on journals' reputations instead of research value; it uses the same parameters as Figure 1.

Under the socially optimal setting, the researchers split along AQ line and researchers in ABCQ area produce research in the field Y, meanwhile researchers in area AQD produce research in the field X. Area APX depicts researchers who in the original setting choose field Y even though it is socially optimal for them to choose field X. Area PRW depicts researchers who choose field X even though it is socially optimal for them to choose field Y. Areas XLKD and ABNMX depict researchers who deliver less value than is socially optimal for them given the topic they choose. Areas MNYW and KLWZ depict researchers who deliver more value than is socially optimal for them given the topic they choose. Area WYCZ depicts researchers who deliver no value at all even though it is socially optimal for them to conduct research.

The central feature of this model is the difference between reputations of journals which in reality can be translated into difference in prestige across different fields. Since in reality the relationship between journals and fields is many-to-many, researchers' choice of the field is guided by field's prestige derived from the reputation of journals that are likely to publish research from this field.

Researchers choose fields according to their perceived reputation and field reputation depends on the value of research delivered by researchers. This establishes a two-way relationship for which a long-run equilibrium can be identified. Indeed, assuming that reputation in the next period in our game is equal to value of research published by a paper in the previous period, equilibrium exists and it is stable. Moreover, as long as easiness to deliver research in both fields is equal, both fields will have the same reputation (prestige) in the long run. Since we don't observe convergence in fields' prestige, a preliminary conclusion from the model is that different fields have different prestige because it is easier to produce research in some of them and harder in others. Difference in easiness to deliver value in various fields will undergo empirical scrutiny in the empirical part of the paper.

Based on this simple model, the potential sources of inefficient allocation of resources in science can be identified as the following:

1. Some researchers may give up on socially valuable research because they can't get published.
2. Some researchers may deliver less value than is socially optimal because they know they have a "publishable chunk" already.
3. Some researchers may spend too many resources on their research just to get a publication.
4. Some researchers may choose a field of study that is inconsistent with their skills just because some fields are more prestigious than others.

Is it possible that any of these can be real phenomena? Let's consider the following situations:

1. A researcher has a possibility to put together a data set which may be used by somebody else in their research project.
2. A researcher has spotted a flaw in somebody else's article as she is not able to replicate it.
3. A researcher was able to confirm another research project by replicating it with another data set.
4. A researcher conducts a research project in which she fails to obtain statistically significant results even though confidence intervals rule out any economic significance.
5. A young researcher without established reputation writes a controversial paper.

All these are situations where valuable work could be done. In all these situations however it is hard to get published. With few exceptions, most reputable journals will not publish information that somebody has an “interesting” database unless solid conclusions are drawn from it. As Hamermesh (2007) points out, replication studies are heavily discriminated against in spite of their unambiguous importance (for discussion, see for example Dewald, Thursby, and Anderson, 1986). De Long and Lang (1992) as well as Stanley (2005) find that publication bias causes among others exclusion of studies with non-significant results. Finally, remember how hard it was for Akerlof (1970) to publish his original paper on information asymmetry. As a result, researchers are discouraged from pursuing these productive activities because they are valued not according to the actual impact but according to the prestige of the journal while no journal wants to publish them.

Given the paper is destined for publication, there is still a lot of potential sources of waste arising from manipulations that have nothing to do with increasing quality of the paper but require resources or actually decrease value of the research project. These include:

1. Splitting one research project into a number of “publishable chunks”.
2. Leaving mistakes (or willful deceits) that are likely to be overlooked by referees but improve paper’s chances.
3. Working on formalities to make a paper compatible with publisher’s format.
4. Working too much because of unreasonable demands from referees.

There aren’t many studies that would make an attempt at empirically quantifying magnitude of these effects, but it is not hard to find anecdotal evidence for at least some of them. For example Frey (2003) provides compelling argument against unreasonable demands put on an author by the peer-review process and Ellison (2000) provides a lot of insight into mechanisms forming these unreasonable demands.

The last source of waste arises from the fact that some researchers can choose research topic that delivers less value but can be published in a better journal. The empirical strategy to test this hypothesis is presented in the remaining part of the paper.

3 Empirical strategy

The first hypothesis of interest is whether in some fields it is harder to carry out research than in others. Testing this hypothesis allows us to reject a preliminary conclusion of a theoretical model derived in the previous section. Failing to reject it is a necessary condition for the model to be accurate as equality in easiness for the fields would mean that there is a need to identify alternative reasons for different reputations of fields. Unfortunately due to methodological difficulties I refrain from trying to establish causality between field difficulties and field reputations. Identifying one potential reason for differences in reputations must suffice for now.

For that purpose I will assume that value of research is a function of researcher's overall ability, field's easiness to produce research and a random disturbance. This relationship can be modeled as

$$v_{ij} = a_i + \sum_{k=1}^{K} I_{ijk} b_k + \varepsilon_{ij}$$

where:

v_{ij} is value of research project number j coauthored by researcher i,

a_i is researcher i's ability to create valuable research (average value of research delivered),

I_{ijk} indicates whether research project number j coauthored by researcher i belongs to field k,

b_k denotes average effect of field k on the quality of the research project, and

ε_{ij} is a random disturbance.

The parameter of interest is b_k. Under the null hypothesis the fields have the same difficulty and $b_k = 0 \; \forall k$.

Any potential endogeneity in this setting would mean that researchers tend to deliver more valuable research in some fields, but neither because of field's characteristics (which we intend to capture) nor because of researcher's characteristic which is already included in the equation as a control for self-selection of researchers.

A plausible source of such an endogeneity would be a time trend in average perceived value of research and a time trend in a number of papers published in that particular field. A remedy for this problem is to control for time. Then any trend in variables of interest will not cause endogeneity. Thus, the final form of the model is:

$$(1)\quad v_{ij} = a_i + \sum_{k=1}^{K} I_{ijk} b_k + T_{N_{ij}} + \varepsilon_{ij}$$

where:

N_{ij} is a year in which research project number j coauthored by researcher i has been published and

T_n is an effect of year n on value of research.

Since de-meaning will pick up any researcher-specific fixed effects any occurrence of serial correlation of the error term is unlikely. The only potential source of inconsistency of covariance matrix is heteroscedasticity. It is then better to use heteroscedasticity robust standard errors. Including proper controls and taking care of heteroscedasticity should be enough to get consistent estimators for both the parameters of interest and their covariance matrix.

After showing that fields indeed differ in their difficulty levels, I will focus on showing that researchers may choose their fields of study not according to their skills but according to perceived prestige of a field. This is the main empirical conclusion that is drawn in support of the theoretical model presented in section two.

I assume that each researcher has some information about what and how influences her utility and makes decisions based on this information. Sources of such utility can be for example (a) getting tenure, (b) salary, (c) reputation as a productive scientist. A researcher makes choices about what field of study she should choose in order to maximize this utility based on her information. For the purpose of this study, I assume that pieces of information relevant to making such choices are (a) her ability to deliver high quality research in each field and (b) her ability to get a publication in a top-tier journal.

At the onset of her career, a researcher has some prior information that guides her decisions, and over time this information improves. Most importantly, over time a researcher is able to observe results of her work, its perceived value, and how this value affects her utility. Then she adjusts her choices. This is an ongoing process for every active researcher. However, for every researcher I analyze, I need a cut-off date that separates period when she observes her ability from the period when she makes choices based on her observations. This simplification allows me to establish causality.

For estimation purposes, I assume that when researcher decides what field to choose for her next research project, she takes into account her relative ability to create value in the field and the field's prestige. Formally, probability that she chooses field k can be modeled as:

$$P(f = k) = f(\breve{a}_{ik}, r_k, \varepsilon)$$

where:

f denotes the outcome field,

k is a field number,

$\breve{a}_{ik}$ is ability of the researcher i in field k relative to her ability in other fields,

r_k is reputation or prestige of field k, and

ε is a random disturbance.

Relative ability of the researcher i in field k is derived by dividing average value of research projects that belong to field k by overall average value of research projects for that particular individual. The choice set for each researcher (that is set of ks that can appear on the left hand side of the equation) is the set of fields that she had previously published in. This effectively means that I restrict the sample to informed decisions and disregard uninformed decisions.

The hypothesis is based on the comparison between the two games described in the previous section. According to the former game, researchers will consider both their relative ability and reputation of the field while making the choice. This is undesirable since in the socially optimal case, the researchers will take into account their relative ability only, as shown in the latter game. Hence the null hypothesis is that r_k has no effect on $P(f = k)$.

The simplest way to test it is to build a linear share regression model:

$$p_{ik} = \alpha\breve{a}_{ik} + \beta r_k + \varepsilon_{ik}$$

where observations are derived from researchers' choices made after their relative ability in the fields has been observed. Moreover:

p_{ik} is a fraction of research projects undertaken by researcher i in field k after ability has been observed,

$\breve{a}_{ik}$ is relative ability of the researcher i in field k,

r_k is reputation of field k, and

ε_{ik} is a random disturbance.

The null hypothesis is $\beta = 0$.

Reputation of a field depends in long run on researchers' decisions about whether to publish in this field and how much value to deliver. This effectively means that our left-hand side variable contributes to the value of our right-hand side variable. However, the magnitude of this contribution is extremely low compared to the reverse relation. Therefore I am going to neglect this as a potential source of endogeneity.

Endogeneity could also arise if fields of study had some characteristics that correlate with the reputation and which are taken into account by a researcher making a decision. A plausible example of such a characteristic is popularity of the field. I am going to include it as a control variable.

Finally, if a researcher had chosen a field before, she might have accumulated knowledge and experience in this field as well as connections to potential co-authors that make her more likely to choose this field in future despite her low productivity. In order to separate effect of observed ability it is hence important to control for volume of past experience with the field. Therefore, the final formula for this model is:

$$(2)\ \ p_{ik} = \alpha\breve{a}_{ik} + \beta r_k + \gamma c_k + \delta q_{ik} + \varepsilon_{ik}$$

where:

c_k is popularity of the k-th field and

q_{ik} is a fraction of research projects undertaken by researcher i in field k before ability has been observed ("field exposure" from now on).

Note that it is unlikely for a strong heteroscedasticity to exist in this model. What can cause inconsistency of the conventional covariance matrix is potential autocorrelation of the errors terms within clusters encompassing observations coming from the same author. This can be alleviated by using clustering in the regression procedure.

A significant fraction of researchers may have choice sets that are superior for the sets of previously chosen fields. For such researchers the left hand-side variables within their clusters add up to one. This however should not be a problem because right-hand side of the equation does not add up to one within the cluster. Hence, there is no threat that covariance matrix could be singular.

4 Data

Table 1 summarizes proxies for variables and definitions of classes necessary for estimation. Economics has been chosen because of my familiarity with it and its widespread usage of JEL codes. I follow Grijalva and Nowell (2008) using JEL codes, as they are clear and easy indicators of a "field of study" necessary to carry out analysis in accordance with the previous section.

The data were collected from a publicly available database of articles and papers in economics, maintained by RePEc on Nov 8th, 2013. The data about articles, authors and citations were procured in accordance with instructions on their website[1]. The data about impact factors of journals were also taken from RePEc[2]. The raw data were parsed using a C++ program[3] and manipulated with MS Access and MS Excel to obtain tables suitable for a statistical package.

There are a number of other databases that include metadata of scientific articles. Unfortunately, most of them do not contain information necessary to carry out this analysis. For example, Web of Science database seems to be one that has been often used when it comes to citation analysis. It has been

[1] http://ideas.repec.org/getdata.html
[2] http://ideas.repec.org/top/top.journals.simple.html
[3] Source code available from author upon request.

criticized for its exclusion of books and arbitrary choice of journals it includes in its indices (see Archambault et al., 2006, for literature on problems related to Web of Science coverage). But most importantly, it does not include information on JEL codes for listed articles. The same issue occurs with Google Scholar or Microsoft Academic Search. Science Direct has JEL codes and citations but its scope is limited to Elsevier publications. Some databases dedicated specifically to Economics provide JEL codes in their metadata repositories. Among them, EconLit does not have information on citations. SSRN seems to have full set of information needed but information there is often outdated as it relates to working papers before they have been published. An important part of this study is analysis of journals in which the articles were published which rules out SSRN.

RePEc has a database that contains information on over 1.4 million published articles and working papers out of which 0.4 million were JEL coded. Information is entered by volunteers as well as by publishing companies. There are 1917 economics journals covered in RePEc as opposed to 335 covered by Web of Science. It also has working papers in its database. Moreover RePEc has an easily accessible database which is available freely to everybody interested.

RePEc's database is by no means perfect. It has a number of issues coming mostly from the fact that there is no central authority to enforce quality of entries. According to RePEc itself, 1% of the entries are incorrect[4]. Moreover, some of the journals are not covered fully, although this issue pertains mostly to less popular journals. Finally, content of the database relies heavily on the use of text processing programs, which are not always able to parse content of the documents properly.

Despite its shortcomings, RePEc is a valuable source of data about patterns in authors' behavior. It has been used before for research purposes for example by Krapf, Ursprung and Zimmermann (2014). A thorough discussion of content, capabilities and limits of RePEc database can be found in Zimmermann (2013).

The data set used to estimate equation (1) was constructed by using every coauthor-article combination for which: (a) year of publication of the article was identified and it was 1994 or later, (b)

[4] http://econpapers.repec.org/check/

JEL codes of the article were identified, (c) the author of an article was identified. As a result I got 85,083 observations. Summary statistics of this data set has been presented in Table 2. Note that variables indicating field do not add up to 1 because one paper can belong to a number of fields.

Before I started construction of a data set for estimation of equation (2), it was necessary to get information on each field that was going to be used as explanatory variables in the regression. The fields are summarized in Table 3. Every field has the average number of citations to an article from that field, its size measured by number of articles, and three measures of reputation: number of papers published in top 10 journals[5] ranked by impact factors (TOP10), number of papers published in top 50 journals ranked by impact factors (TOP50) and average impact factor of journals that publish papers from that field (IF). [Add a list from ISI and comment according to Ranking Leading Econometrics Journals Using Citations Data from ISI and RePEc].

Each row of the dataset used to estimate equation (2) is an element of a researcher's choice set. Only researchers whose publication activity spanned two or more years were taken into account. This period for each researcher was split into a "learning period" and a "decision period". Activity from the "learning period" was used to compute explanatory variables. Activity from the "decision period" was used to compute the dependent variable. Because the volume of papers registered in RePEc database increases over time, activity period was split so that for each researcher .79 of it covers "learning period" and subsequent .21 covers "decision period", rounded to whole years. These proportions have been chosen so as to maximize overlap between choice sets and fields chosen in the decision period, and therefore to maximize sample size. The outcome data set contains 45,866 field-researcher observations.

Along with the data based on published articles, an indicator whether researcher is working in the United States of America has been constructed. This was done based on email addresses of the researchers. Email addresses ending with ".edu", ".gov", and ".us" were treated as indication that

[5] These are: The Quarterly Journal of Economics, Journal of Economic Literature, Econometrica, Journal of Economic Growth, Journal of Financial Economics, Review of Economic Studies, Journal of Political Economy, Journal of Economic Perspectives, Economic Policy, Brookings Papers on Economic Activity

researcher works in the United States of America. There are 5,921 observations falling into this category. Summary statistics are presented in Table 4.

5 Results

The results of estimation of equation (1) are presented in Table 5. Year 1994 dummy has been omitted due to co-linearity with other year dummies. It is clear from these regressions that some fields (for example M: Business Administration and Business Economics, Marketing, Accounting; and O: Economic Development, Technological Change, and Growth) are easier to get a citation than others (for example K: Law and Economics; and L: Industrial Organization). The magnitudes of these differences are big, given the average number of citations for an article. For example, on average, Law and Economics articles receive 1.5 citations less than Business Administration and Business Economics, Marketing, Accounting articles. The standard error of this estimate is around 0.4. Given the average numbers of citations this is a difference of approximately 20%. Note that subtracting column “coefficient” in Table 5 from column “Average # of citations” in Table 3 gives us average ability of a researcher publishing in a field.

Equation (2) has been estimated in 6 variants and results of these estimations are presented in Table 6. First, each of the three field reputation measures is used to ensure robustness of the results with respect to this choice. Secondly, every estimation is carried out for the full sample and for a US researchers subsample. Finally, for every estimate there are provided standard errors with and without clustering. There are 14,587 clusters (i.e. authors) in the full sample and 1,950 clusters in the US subsample.

In all regressions all coefficients have signs that are consistent with predictions. Field exposure is consistent across specifications both in magnitude and statistical significance. A 10 percentage point (pp) increase in fraction of research that has been done by a researcher in a particular field of study in the past, increases the chances of choosing this field in future by around 3 pp.

Author's ability is always significant for the entire sample at 0.001 level but never significant for the US subpopulation. This may indicate that especially researchers in the United States are prone to ignore their ability and pragmatically choose what is in their best interest according to popularly used measures. Note relatively weak economic significance of this variable. An increase in ability by a standard deviation causes an increase in probability of choosing a field between 0.0004 and 0.0038 across specifications.

Field's reputation, the variable of interest, is significant at 0.05 level across all specifications except when IF measure of reputation is used for the entire sample. This pattern is also consistent with the notion that especially researchers in the US are more likely to pay attention to field's reputation than their own ability when choosing the field. Similarly to author's ability, this variable also has relatively weak economic significance. An increase in field's reputation by a standard deviation causes an increase in probability of choosing a field between 0.0019 and 0.0107. Impact of field's reputation is bigger than impact of author's ability in all specifications except when IF measure of reputation is used for the entire sample. It is not likely that small economic significance invalidates the point of the study, since a lot of information has been lost while preparing the data for estimation. Simplifying assumptions like arbitrary splitting researchers careers in two is one of the examples why magnitudes of coefficients can be underestimating real effect of the underlying mechanisms. A lot more accurate study will be needed to make an estimate of actual welfare loss expressed in dollars.

Field's popularity also is statistically significant at 0.05 level across all specifications. A field that is twice as big adds between 1.1 and 2.4 pp to the probability that researcher will choose that field in future, depending on specification. This indicates that holding everything else constant, authors have a tendency to switch to more popular topics. The general results are robust to whether field popularity is measured by raw number of papers in the field or by its natural logarithm.

6 Conclusions

Taking the results literarily, my empirical analysis shows then when choosing a field of study, researchers care slightly more about the frequency with which papers from this particular field of study get published in top-tier journals than about the numbers of citations they can get. This supports the claim that incentive structure for researchers is at least partially focused on the numbers of papers they published in top-tier journals. And my theoretical model shows that acting in order to maximize number of publications in top-tier journals leads to a welfare loss.

There is no question that a measure to evaluate research output both with respect to quantity and quality is needed. People making decisions do not have time and resources to carefully check every piece of work for its quality. The question then, is how to construct such a measure so that it can be used in practice and so that it minimizes distortion of human behavior.

In practice, two sources on information about quality of research projects proved to be useful and widely accepted despite their imperfections: journal certification and citation-based measures (see Fray and Rost, 2010, for their culprits). Number of papers published in top-tier journals is one of them. The fact that an article was published in a top-tier journal is a signal that the article is valuable research only as long as it does not have many citations and we are comparing it to another article of the similar age that does not have many citations either. Over time, strength of this signal fades away as compared with number of citations which indicates how many researchers use the article's contribution in their own research and as an alternative signal, it gains accuracy with time. To the further detriment of publication in a top-tier journal as a signal, as Ellison (2011) points out, an increasing number of high-value papers never undergo peer review and are never published in a journal. Instead, they float on the Internet in form of working papers and receive numerous citations from both formally published and unpublished sources. Finally, publication in a good journal not only provides less information than number of citations but it also alters human behavior in a way that leads to suboptimal quality of their work, as my theoretical model indicates.

Use of bibliographic measures for evaluating documents has been booming during past two decades. Page et al. (1999) developed an algorithm based on bibliographic measures that evaluate websites. This idea proved to be practical and extremely efficient in weeding out low-quality documents as indicates the commercial success of Google Inc. However, the analogy between websites and scientific articles has its limitations. In scientific community there is often pressure to evaluate research projects even before they have been published, without giving them any time to accumulate citations. Therefore, it may be worthwhile to consider to proxy value of such research paper with an impact factor of the journal it is being published in. On the other hand, research that is a few years old can be easily judged by its citation count adjusted for the age of the article. Exact construction of such a measure would depend on its purpose – whether it is used to compare candidates for a job opening, ranking of researchers, or as a performance indicator for an entire research institution.

Use of bibliographic measures is increasingly easier thanks to popular and free search engines. For instance, Google Scholar counts all documents that cite a particular research project and eliminates duplicates. Card and DellaVigna (2013) describe advantages and provides an example how data from this tool can be used in research. As long as the information on citation numbers is neglected and coarser measures are in place, resources in science are mismanaged in spite of an available easy fix.

References

Akerlof, George A. "The market for ‘lemons’: quality uncertainty and the market mechanism." Aug 84.3 (1970): 488-500.

Archambault, Éric, et al. "Benchmarking scientific output in the social sciences and humanities: The limits of existing databases." Scientometrics 68.3 (2006): 329-342.

Card, David, and Stefano DellaVigna. Nine facts about top journals in economics. No. w18665. National Bureau of Economic Research, 2013.

De Long, J. Bradford, and Kevin Lang. "Are all economic hypotheses false?." JOURNAL OF POLITICAL ECONOMY-CHICAGO- 100 (1992): 1257-1257.

Dewald, William G., Jerry G. Thursby, and Richard G. Anderson. "Replication in empirical economics: The journal of money, credit and banking project." American Economic Review 76.4 (1986): 587-603.

Ellison, Glenn. “Evolving standards for academic publishing: A qr theory.” No. w7805. National bureau of economic research, 2000.

Ellison, Glenn. "Is peer review in decline?." Economic Inquiry 49.3 (2011): 635-657.

Franzoni, Chiara, Giuseppe Scellato, and Paula Stephan. "Changing incentives to publish." science and engineering 5 (2011): 6.

Frey, Bruno S. "Publishing as prostitution?–Choosing between one's own ideas and academic success." Public Choice 116.1-2 (2003): 205-223.

Frey, Bruno S., and Katja Rost. "Do rankings reflect research quality?." Journal of Applied Economics 13.1 (2010): 1-38.

Gomez-Mejia, Luis R., and David B. Balkin. "Determinants of faculty pay: An agency theory perspective." Academy of Management journal 35.5 (1992): 921-955.

Grijalva, Therese C., and Clifford Nowell. "A guide to graduate study in economics: Ranking economics departments by fields of expertise." Southern Economic Journal (2008): 971-996.

Hamermesh, Daniel S., George E. Johnson, and Burton A. Weisbrod. "Scholarship, Citations and Salaries: Economic Rewards in Economics." Southern Economic Journal 49.2 (1982).

Hamermesh, Daniel S. "Viewpoint: Replication in economics." Canadian Journal of Economics/Revue canadienne d'économique 40.3 (2007): 715-733.

Hirsch, J. E. "An index to quantify an individual's scientific research output." PNAS 102.46 (2005): 16569-16572.

Krapf, Matthias, Heinrich W. Ursprung, and Christian Zimmermann. "Parenthood and productivity of highly skilled labor: evidence from the groves of Academe." Federal Reserve Bank of St. Louis Working Paper Series 2014-001 (2014).

Laband, David N. "An evaluation of 50" Ranked" economics departments: By quantity and quality of faculty publications and graduate student placement and research success." Southern Economic Journal (1985): 216-240.

Levin, Sharon G., and Paula E. Stephan. "Research productivity over the life cycle: Evidence for academic scientists." The American Economic Review (1991): 114-132.

Medoff, Marshall H. "Collaboration and the quality of economics research." Labour Economics 10.5 (2003): 597-608.

Mittal, Vikas, Lawrence Feick, and Feisal Murshed. "Publish and prosper: the financial impact of publishing by marketing faculty." Marketing Science 27.3 (2008): 430-442.

O'Keefe, Suzanne, and Ta-Chen Wang. "Publishing pays: Economists' salaries reflect productivity." The Social Science Journal 50.1 (2013): 45-54.

Oster, Sharon M., and Daniel S. Hamermesh. "Aging and productivity among economists." Review of Economics and Statistics 80.1 (1998): 154-156.

Oswald, Andrew J. "An Examination of the Reliability of Prestigious Scholarly Journals: Evidence and Implications for Decision-Makers." Economica 74.293 (2007): 21-31.

Page, Lawrence, et al. "The PageRank citation ranking: Bringing order to the web." (1999).

Sauer, Raymond D. "Estimates of the returns to quality and coauthorship in economic academia." Journal of Political Economy 96.4 (1988): 855-66.

Singh, Gangaram, Kamal M. Haddad, and Chee W. Chow. "Are articles in “top” management journals necessarily of higher quality?." Journal of Management Inquiry 16.4 (2007): 319-331.

Stanley, Tom D. "Beyond publication bias." Journal of Economic Surveys 19.3 (2005): 309-345.

Stephan, Paula E. The Economics of Science-Funding for Research. No. 12-2010. ICER-International Centre for Economic Research, 2010.

Zimmermann, Christian. "Academic rankings with RePEc." Econometrics 1.3 (2013): 249-280.

Academic R&D Spending Increased in FY 1992, NSF 93-324, December 1993, http://www.nsf.gov/statistics/rdexpenditures/datbrief/datbrief.htm

Higher Education R&A Expenditures Remain Flat in FY 2012, NSF 14-303, November 2013, http://www.nsf.gov/statistics/infbrief/nsf14303/

Figure 1: Equilibrium for the game with payoffs depending on journal's reputation.

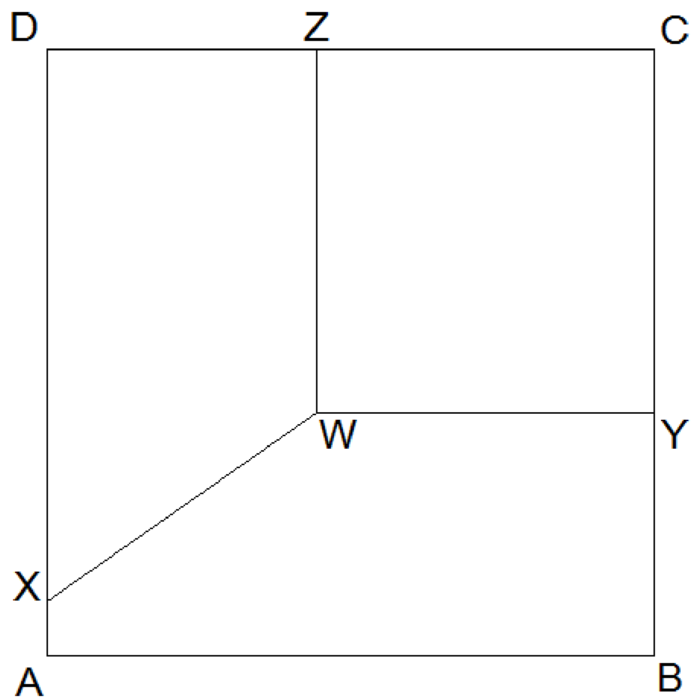

Figure 2: Equilibrium for the game with payoffs depending on research value, superimposed on Figure 1 for comparison.

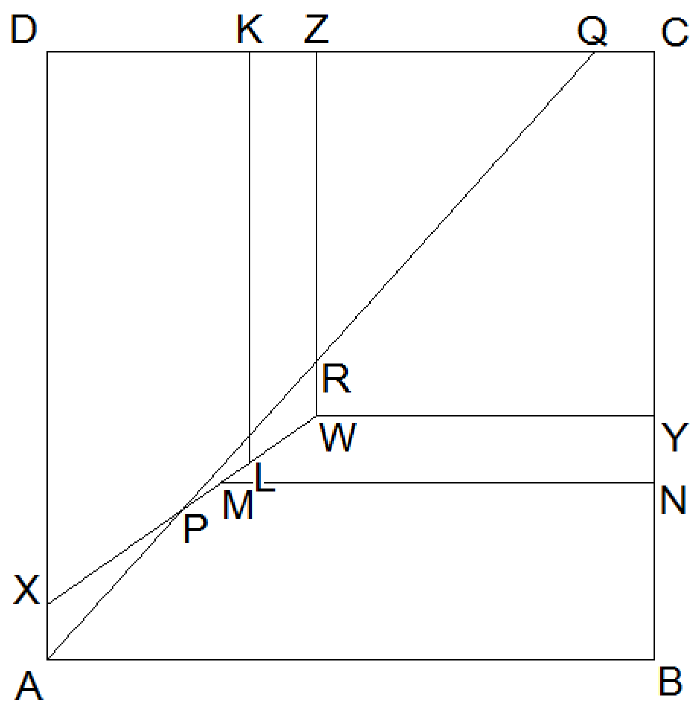

Table 1. Measures used as proxies and labels used as categories

Notion	**Empirical counterpart**
Value of a research project	Number of accumulated citations by the article

Ability of a researcher to create value	Average number of citations for a researcher
Relative ability of the researcher i in the field k	Average number of citations in a particular field obtained by a researcher divided by the overall average number of citations obtained by this researcher
A field of study	An area of research denoted by a distinct JEL code letter
Field's reputation	I am going to consider three measures: 1. Number of papers in the field that has been published in the top 10 journals (TOP10) 2. Number of papers in the field that has been published in the top 50 journals (TOP50) 3. Average impact factor of journals for papers published in the field (IF from now on)
Popularity of a field	Natural logarithm of the number of papers published in that field

Table 2. Basic statistics of variables for coauthor-article combinations.

Variable	**Mean**	**Std. Dev**	**Min**	**Max**
Number of citations	4.011	22.685	0	1598
"Researcher's ability"	4.011	12.583	0	888
Field A	0.0192	0.137	0	1
Field B	0.017	0.131	0	1
Field C	0.276	0.447	0	1
Field D	0.243	0.429	0	1
Field E	0.191	0.393	0	1
Field F	0.172	0.377	0	1
Field G	0.134	0.34	0	1
Field H	0.122	0.328	0	1
Field I	0.089	0.285	0	1
Field J	0.157	0.364	0	1
Field K	0.025	0.155	0	1
Field L	0.127	0.333	0	1
Field M	0.04	0.195	0	1
Field N	0.023	0.15	0	1
Field O	0.169	0.375	0	1
Field P	0.026	0.159	0	1
Field Q	0.08	0.272	0	1
Field R	0.063	0.243	0	1
Field Y	0.003	0.058	0	1
Field Z	0.021	0.144	0	1
Year	2007.7	4.33	1994	2013
Year 1994	0.002	0.046	0	1
Year 1995	0.004	0.067	0	1
Year 1996	0.008	0.089	0	1
Year 1997	0.01	0.101	0	1
Year 1998	0.013	0.114	0	1
Year 1999	0.021	0.144	0	1
Year 2000	0.027	0.161	0	1
Year 2001	0.03	0.171	0	1
Year 2002	0.031	0.173	0	1

Year 2003	0.033	0.179	0	1
Year 2004	0.041	0.198	0	1
Year 2005	0.052	0.221	0	1
Year 2006	0.057	0.232	0	1
Year 2007	0.068	0.252	0	1
Year 2008	0.084	0.278	0	1
Year 2009	0.095	0.294	0	1
Year 2010	0.102	0.302	0	1
Year 2011	0.101	0.301	0	1
Year 2012	0.134	0.341	0	1
Year 2013	0.086	0.281	0	1

Table 3. Basic statistics on Fields.

Field	**JEL**	**Average # of citations**	**SD in # of citations**	**# of papers**	**TOP -10**	**TOP -50**	**IF**
General Economics and Teaching	A	6.20	15.0	8029	130	180	2.42
History of Economic Thought, Methodology, and Heterodox Approaches	B	4.70	11.4	6545	97	118	2.23
Mathematical and Quantitative Methods	C	6.07	15.8	74122	102	1062	2.51
Microeconomics	D	6.49	16.3	69461	231	1511	3.51
Macroeconomics and Monetary Economics	E	9.74	35.2	50994	213	1272	3.21
International Economics	F	9.26	26.7	48338	110	816	2.64
Financial Economics	G	6.98	16.8	42941	380	982	3.26
Public Economics	H	6.62	14.7	34540	121	648	2.97
Health, Education, and Welfare	I	6.81	20.9	34364	140	553	2.64
Labor and Demographic Economics	J	7.74	17.3	52399	183	946	3.69
Law and Economics	K	6.66	15.6	9632	64	159	2.70
Industrial Organization	L	6.16	15.1	38092	122	589	2.29
Business Administration and Business Economics, Marketing, Accounting	M	5.15	9.91	19431	32	118	0.71
Economic History	N	7.20	29.3	9290	50	179	3.28
Economic Development, Technological Change, and Growth	O	8.07	29.8	48315	142	874	2.41
Economic Systems	P	7.52	30.4	8832	63	140	2.41
Agricultural and Natural Resource Economics, Environmental and Ecological Economics	Q	4.91	10.8	24336	61	245	1.66
Urban, Rural, Regional, Real Estate, and Transportation Economics	R	5.42	11.1	19267	23	162	1.29
Miscellaneous Categories	Y	4.22	7.44	1194	25	115	5.51
Other Special Topics	Z	5.82	17.5	10916	15	93	1.05

Table 4. Basic statistics of coauthor-field combinations.

Variable	Mean	Std. Dev	Min	Max
Fraction of research projects in "decision period"	0.141	0.228	0	1
Fraction of research projects in "learning period"	0.318	0.234	0.004	1
Researcher's ability in the field	0.985	0.758	0	15
TOP10	150.4	83.106	15	380
TOP50	838.5	414.25	93	1511
IF	2.734	0.693	0.713	5.517
Field size: Ln(number of articles in the field)	10.605	0.589	7.085	11.21
USA indicator	0.129	0.335	0	1

Table 5. Regression of equation (1).

variable	coefficient	Regular OLS		Heteroscedasticity robust	
		Std. error	p-value	Std.Error	p-value
Ability	0.967	0.005	0.000	0.100	0.000
A	-1.044	0.476	0.028	0.245	0.000
B	-1.343	0.500	0.007	0.342	0.000
C	-0.083	0.151	0.582	0.134	0.535
D	-0.311	0.156	0.047	0.155	0.045
E	0.029	0.173	0.866	0.181	0.871
F	0.120	0.181	0.507	0.192	0.531
G	-0.121	0.199	0.544	0.159	0.446
H	-0.300	0.201	0.136	0.139	0.031
I	0.494	0.235	0.036	0.359	0.169
J	-0.343	0.188	0.068	0.225	0.127
K	-0.860	0.418	0.040	0.305	0.005
L	-0.579	0.201	0.004	0.130	0.000
M	0.388	0.337	0.250	0.170	0.022
N	1.177	0.431	0.006	1.113	0.290
O	0.625	0.177	0.000	0.213	0.003
P	0.721	0.406	0.075	1.001	0.471
Q	0.149	0.245	0.542	0.142	0.294
R	-0.411	0.269	0.126	0.155	0.008
Y	-3.152	1.117	0.005	0.773	0.000
Z	-0.689	0.450	0.126	0.578	0.233
Year 1994	omitted				
Year 1995	0.851	1.701	0.617	4.052	0.834
Year 1996	-6.445	1.576	0.000	3.194	0.044
Year 1997	-6.183	1.539	0.000	3.230	0.056
Year 1998	-4.557	1.513	0.003	3.173	0.151
Year 1999	0.513	1.472	0.727	3.454	0.882
Year 2000	-1.736	1.459	0.234	3.235	0.592
Year 2001	-1.737	1.453	0.232	3.300	0.599
Year 2002	-7.740	1.452	0.000	3.159	0.014
Year 2003	-8.123	1.449	0.000	3.168	0.010
Year 2004	-7.825	1.441	0.000	3.172	0.014
Year 2005	-8.011	1.434	0.000	3.165	0.011
Year 2006	-8.354	1.431	0.000	3.175	0.009
Year 2007	-8.618	1.427	0.000	3.185	0.007

Year 2008	-8.244	1.423	0.000	3.174	0.009
Year 2009	-8.869	1.421	0.000	3.173	0.005
Year 2010	-9.203	1.420	0.000	3.184	0.004
Year 2011	-9.419	1.420	0.000	3.196	0.003
Year 2012	-9.630	1.417	0.000	3.202	0.003
Year 2013	-10.488	1.423	0.000	3.202	0.001
Constant	8.465	1.409	0.000	3.276	0.010

Note: dependent variable is number of citations of an article. $R^2 = 0.32$.

Table 6. Regression of equation (2).

			Field Exposure	Author's Ability	Field's Reputation	Field Popularity	Constant
TOP10	Full	Coefficient	0.274	4.987E-03	5.910E-05	0.032	-0.302
		SD_1	4.380E-03	1.339E-03	1.360E-05	1.935E-03	0.020
		p-value	0.000	0.000	0.000	0.000	0.000
		SD_2	6.160E-03	1.102E-03	1.480E-05	1.695E-03	0.017
		p-value	0.000	0.000	0.000	0.000	0.000
	US	Coefficient	0.259	8.037E-04	1.231E-04	0.024	-0.216
		SD_1	0.012	3.793E-03	3.940E-05	5.356E-03	0.054
		p-value	0.000	0.832	0.002	0.000	0.000
		SD_2	0.016	2.823E-03	4.270E-05	4.634E-03	0.045
		p-value	0.000	0.776	0.004	0.000	0.000
TOP50	Full	Coefficient	0.275	4.972E-03	1.030E-05	0.030	-0.275
		SD_1	4.380E-03	1.340E-03	4.710E-06	3.321E-03	0.032
		p-value	0.000	0.000	0.029	0.000	0.000
		SD_2	6.157E-03	1.102E-03	4.530E-06	2.809E-03	0.026
		p-value	0.000	0.000	0.023	0.000	0.000
	US	Coefficient	0.259	6.151E-04	2.580E-05	0.017	-0.142
		SD_1	0.012	3.794E-03	1.280E-05	8.949E-03	0.086
		p-value	0.000	0.871	0.044	0.061	0.098
		SD_2	0.016	2.809E-03	1.220E-05	7.457E-03	0.070
		p-value	0.000	0.827	0.035	0.025	0.043
IF	Full	Coefficient	0.274	4.960E-03	2.759E-03	0.034	-0.324
		SD_1	4.380E-03	1.340E-03	1.641E-03	1.943E-03	0.019
		p-value	0.000	0.000	0.093	0.000	0.000
		SD_2	6.159E-03	1.102E-03	1.571E-03	1.542E-03	0.015
		p-value	0.000	0.000	0.079	0.000	0.000
	US	Coefficient	0.258	4.887E-04	9.264E-03	0.027	-0.258
		SD_1	0.012	3.795E-03	4.532E-03	5.237E-03	0.051
		p-value	0.000	0.898	0.041	0.000	0.000
		SD_2	0.016	2.808E-03	3.980E-03	4.070E-03	0.039
		p-value	0.000	0.862	0.020	0.000	0.000

Note: SD_1 is standard deviation without clustering. SD_2 is standard deviation with clustering. Dependent variable is fraction of articles published by an author in a field after the cut-off date. For all specifications $R^2 = 0.09$.