The Economics of Citation

Jeong-Yoo Kim*
Kyung Hee University

Insik Min
Kyung Hee University

Christian Zimmermann
University of Connecticut

May 12, 2008

Abstract

In this paper, we study the citation decision of a scientific author. By citing a related work, an author can make his argument more persuasive. We call this the correlation effect. On the other hand, if he cites someone else's work, he may give an impression that he thinks the cited author more competent than himself. We call this the reputation effect. These two effects may be the main sources of citation bias. We empirically show that there exists citation bias in Economics by using data from RePEc. We also report how the citation bias differs across regions (U.S., Europe and Asia).

Key Words: citation bias, correlation effect, reputation effect, signal, strategy

JEL Classification Code: D81

*Corresponding author: This research was begun when the first author was visiting ISER, Osaka University in the winter of 2005. We are grateful to seminar audiences at Kookmin University, the University of Connecticut and participants in the applied microeconomics workshop held at Korea Foundation of Advanced Studies, the Autumn Conference of the Japanese Applied Economics Association held at Chuo University for helpful comments. We are also indebted to Kwanho Shin for useful suggestions in the primordial stage of this research. Mailing address: Department of Economics, Kyung Hee University, 1 Hoegidong, Dongdaemunku, Seoul 130-701, Korea, Tel: +822-961-0986, Fax: +822-966-7426, Email: jyookim@khu.ac.kr

1 Introduction

The scientific progress is achieved cumulatively by individual efforts of scientists. Scientists keep doing research even if they can hardly expect to get paid much for it. Presumably, to most of the scientists, the driving force of their research would not be monetary rewards, but receiving recognition for it. Since Shepard's Citations initiated as legal citators in 1873, ISI (Institute for Scientific Information) introduced various citation indices that have been used to measure a scientist's contribution[1] to his discipline. As a result, those indices have significantly influenced tenure, promotion and reappointment evaluations as well as other decisions in universities or research institutions, like merit pay or endowed chairs. These decisions are taken under the assumption that citations reflect the true quality of the researcher. What if there were some strategic aspects in citations? We investigate here whether there is some distortion in citation patterns.

For this purpose, we examine the correlation between an author's rank and the average rank of those he or she cites.[2] Figure 1 shows this. If there were no citation bias, the citation line would be horizontal. No matter who cites, the pool of cited works would be similar. However, a positive slope of the citation line drawn in Figure 1 suggests that there is a bias in the citation pattern. In particular, the figure shows that authors tend to cite other authors whose ranks are higher than themselves. The goal of this paper is to explain the phenomenon of such an upward bias in citation.

Our argument in explaining an upward citation bias starts from our fundamental view on citation, namely, "Citing is a *strategy*."[3] It is told that many scientific authors experience the embarrassing moment of finding their work not being cited in closely related works by others. Why have the latter authors failed to cite predating related works at the expense of embarrassing or even offending someone? There must be a gain from doing

[1]The word "contribution" is rather ambiguous in this context. Note that quality and influence cannot be identified, although they may be correlated. Then, it is not clear whether contribution refers to quality or influence.

[2]We use data from the RePEc (Research Papers in Economics), which is a decentralized database of working papers, journal articles and professional books. For more details of RePEc, see http://repec.org/ or Krichel (2000). Detailed variable descriptions are given in Table 1.

[3]The view that the academic world has been driven at least partly by strategic motivations seems to be shared by many researchers. See, for example, Zamora Bonilla (2005).

so. Scientific authors decide whether to cite a related work strategically by comparing the cost and the benefit of citing it. The decision is not entirely taken with honesty or scholarly conscience in mind.[4]

The benefit that an author can get from citing a related work is apparent. Above all things, it makes his argument more persuasive. Readers will believe that his argument is more likely to be correct or believable if it is supported by a closely related argument that was made independently by someone else. We call this the *correlation effect*, because the effect is mainly due to the correlation between the truth of the two arguments. Clearly, the correlation effect of citing is larger, that is, his argument will be perceived to be more convincing, if the related argument was advanced by a more competent author. For example, we say "Confucius said that $\cdots$," but we seldom say "My friend Charles said that $\cdots$," to try to convince others of his argument.[5]

This consideration may create some cost in citing a work by others. To elaborate, if an author cites someone else's work, it may give the impression that he thinks the cited author more competent than himself. This may make an author reluctant to cite the work by others, especially by less established authors. This cost of citing is generated mainly through damaging his reputation. So, we will call this the *reputation effect*. By omitting to cite a related work of less established authors deliberately, he can establish the reputation that he at least thinks himself more competent than the author he ought to cite but did not cite. Thus, an author's failure to cite someone else's related work has a vaulting effect in the sense that he intends to jump in reputation by using someone else as a vaulting tool. There are also minor costs of citing. An author cannot cite all the related works. It is burdensome both to the author and to readers. Moreover, it is costly to search for all the relevant works.

This paper consists of a theoretical part and an empirical part. In the theory part, we build a simple model to explain an author's citation decision. As we argued above,

[4]For example, Barry Palevitz (1997) writes his experience where he found a paper omitting to cite his work even though the paper is on a subject almost identical to that covered in his work and one of the authors knew about his work when they wrote the paper. The reader must surely have had similar experiences.

[5]We neglect here the strategy of citing journal editors or potential referees, something we cannot control for in our empirical work.

we identify two main effects, the correlation effect and the reputation effect. By the correlation effect, an author tends to cite only competent authors whose claims are likely to be correct, because citing a related claim by less competent authors may make his own claim look less likely to be true. Also, the reputation effect makes an author, particularly who is less reputed, even more selective in citing. This is because for an author whose academic ability is not yet widely known to cite a less competent author may give a bad signal about his ability. The two effects lead to citation bias.

In the empirical part, we show using data from RePEc that there does exist a citation bias in Economics. The most difficult part in this empirical research is to choose a proxy variable for the reputation of an author. For this purpose, we distinguish two individual ranking variables *RANK* and *RANK_NW*. The former refers to the overall rank of an author in RePEc using a set of 31 different criteria and the latter refers to his rank only determined by the number of authored works weighted by a simple impact factor. Thus, the variable *RANK_NW* does not take the number of citations into account. The variable RANK, which reflects the number of citations, is used as a proxy for an author's reputation.

Most strikingly, we obtain that the citation pattern of similarly ranked authors in terms of *RANK_NW* can be U-shaped with respect to *RANK*. This implies that the average rank of authors that an author cites may decrease as the author is less reputed, and then finally increase if the reputation of the author falls very low. This seems to support that the reputation effect exists, since it can be interpreted as the correlation effect dominated by the reputation effect for authors with intermediate reputation. Of course, this U-shaped citation pattern is not observed for all rank groups. For the top authors, only the declining part is observed.

As a rough proxy for an author's recognizability, we may alternatively use his seniority. We find a more severe citation bias among junior authors, that is, juniors are more selective in citations, which shows an evidence of the reputation effect. We also observe that the number of citations per article is significantly different across regions (U.S., Europe and Asia) conditional on the variable *RANK*. This can be viewed as another evidence of citation bias.

Our paper is organized as follows. In Section 2, we set up a model and provide a

theoretical analysis of an author's citation decision. To separate the correlation effect from the reputation effect, we consider two distinct cases when an author's ability is fully known to all other potential authors and when his ability is known only to a limited number of them. Section 3 contains the empirical analysis supporting the results derived in Section 2. Concluding remarks and some suggestions follow in Section 4.

2 Model

To explain citation bias, we consider the following model. A scientific author (author 1) makes a claim ω_1 in his writing. This claim can be either true or false. The (average) prior probability (or belief) that his claim is true is $\mu_1 \in (0,1)$. We can interpret μ_1 as the ability of the author. The author decides whether to cite a related claim ω_2 by another author (author 2). The average probability that author 2's claim is true is $\mu_2 \in (0,1)$.

We assume that the author is a risk-neutral Bayesian decision-maker, that is, he maximizes the posterior probability (or belief) that his claim is true. Thus, he decides to cite ω_2 if it increases the posterior probability that ω_1 is true. Let $P(\omega_1 = T \mid \omega_2 = T) = \alpha_T$ and $P(\omega_1 = F \mid \omega_2 = F) = \alpha_F$. We assume that $\alpha_T, \alpha_F > 1/2$, i.e., the two claims are correlated.[6] We also assume that α_T and α_F are common knowledge.

2.1 Complete Information

Consider the case that μ_1 and μ_2 are both common knowledge. If author 1 cites ω_2, the posterior belief that claim 1 is true is

$$P(\omega_1 = T \mid \omega_2) = P(\omega_1 = T \mid \omega_2 = T)P(\omega_2 = T) + P(\omega_1 = T \mid \omega_2 = F)P(\omega_2 = F).$$

Therefore, we have

$$E[P(\omega_1 = T \mid \omega_2)] = \alpha_T \mu_2 + (1 - \alpha_F)(1 - \mu_2).$$

[6]This assumption implies that we do not consider negative citations that provide contradictory views or evidence. Wright and Armstrong (2007) documents evidences that authors have a tendency against negative citations.

Since the expected probability that $\omega_1 = T$ with no citation is $E[P(\omega_1 = T)] = \mu_1$, he chooses to cite ω_2 if and only if

$$\mu_1 < \overline{\mu}_1 \equiv \alpha_T \mu_2 + (1 - \alpha_F)(1 - \mu_2). \tag{1}$$

For the following, we assume that $\overline{\mu}_1 \in (0, 1)$. Inequality (1) implies that a less capable author is more likely to cite another of given capability. The intuition is quite clear. A less capable author can increase the posterior belief that his claim is correct if he cites the claim by a reasonably competent author, whereas a more capable one only decreases the posterior belief by citing the claim. We call this the *correlation effect* of citation.

Rewriting inequality (1) leads to our result of selective citation in the case of complete information.

Proposition 1 *When μ_1 is publicly known, author 1 cites ω_2 if and only if $\mu_2 > \overline{\mu}_2 \equiv \frac{\mu_1 + \alpha_F - 1}{\alpha_T + \alpha_F - 1}$.*

Proof. Note that $\alpha_T + \alpha_F > 1$. Thus, it is clear that inequality (1) is equivalent to $\mu_2 > \overline{\mu}_2$. ‖

Proposition 1 suggests that an author cites only the claim made by competent authors. He is reluctant to cite an unreliable author's claim ($\mu_2 < \overline{\mu}_2$). The intuition behind this result is as follows. Given reasonably high α_T and α_F, if μ_2 is large, ω_2 is likely to be correct, which in turn implies that ω_1 looks correct by citing ω_2 because of high α_T. Similarly, if μ_2 is small, ω_2 is likely to be false, implying that citing ω_2 makes ω_1 look false because of high α_F.

Also, let us consider a specific case that $\alpha_T = \alpha_F \equiv \alpha$. If $\mu_2 > 1/2$, the citation benefit gets larger as α increases, so that author 1 is more willing to cite ω_2. In an extreme that $\alpha \approx 1$, he cites as long as the cited author's known ability is higher than his own. However, if $\mu_2 < 1/2$, the citation has a worse effect as α increases. The intuition is clear. As the two claims are more closely related, the truth of ω_2 is more likely to imply the truth of ω_1, while the falseness of ω_2 is more likely to imply the falseness of ω_1. When $\mu_1 = \mu_2 \equiv \mu$, inequality (1) holds if $\mu < 1/2$ but does not if $\mu > 1/2$, implying that an incompetent author ($\mu < 1/2$) always cites the claim by a comparable author, while a competent author does not.

2.2 Incomplete Information

To identify the second effect of citation, consider the alternative case that μ_1 is known only to a limited proportion of the public. Thus, we assume that a proportion λ of the population knows μ_1 for $\lambda \in (0, 1)$, while the rest does not know μ_1 but only knows its distribution $G(\mu_1)$, where $G(\mu_1)$ is defined over (0, 1).[7] We will call μ_1 the type of author 1. We retain the assumption that μ_2 is common knowledge.[8] One can imagine that author 2 is a widely known scholar, while author 1 is a junior scholar who has just entered academics.

Under incomplete information, the citation decision of an author may convey some meaningful information about μ_1. Since the citation decision depends on μ_1 in the model of complete information, the public may be able to infer the author's unknown ability from his citation decision. Taking this into account, an author with unknown ability may cite more selectively to pretend to be more capable. We call this the *reputation effect* of citation.

To show the reputation effect formally, we resort to the usual solution concept, the weak Perfect Bayesian Equilibrium (PBE), which requires the belief of the public to be updated from the prior belief according to Bayes' law whenever possible. Our interest is confined to the equilibrium outcome that some types of author 1 cite while other types do not.[9] In this equilibrium, there must be a type who is indifferent between citing or not citing ω_2 under incomplete information. Let this type be $\tilde{\mu}_1$. Then, we have

Proposition 2 *(i) Author 1 cites ω_2 if $\mu_1 \leq \tilde{\mu}_1(\lambda)$, while he does not if $\mu_1 > \tilde{\mu}_1(\lambda)$, (ii) $\tilde{\mu}_1(\lambda) < \overline{\mu}_1$, and (iii) $\tilde{\mu}_1(\lambda)$ is strictly increasing in λ.*

Proof. See the appendix.

[7] For example, it is usual that the ability of a freshly minted Ph.D. is known only locally.

[8] When μ_2 is unknown, $E(\mu_2 \mid \Omega)$ may be used as a proxy for μ_2 where Ω is observable characteristics of author 2, for example, his/her affiliation, gender, nationality etc. This may be another source of citation bias. Also, risk-averse authors should be less willing to cite an author with unknown μ_2 than the one whose μ is known. We will briefly discuss the empirical implication of this consideration in Table 3.

[9] This is a semi-separating equilibrium outcome. We will not consider the uninteresting pooling case where no types cite. In fact, a pooling equilibrium could be feasible if λ is small enough to make an author tend to take advantage of the reputation effect by not citing.

This proposition says that a more severe citation bias occurs due to the reputation effect. A less widely known author tends to be more reluctant to cite others. The intuition goes as follows. Citing has two effects. On one hand, it directly increases the credibility of his claim (correlation effect), but, on the other hand, it has an indirect signalling effect; adjusting the belief of his ability downwards (reputation effect). Thus, an author decides whether to cite or not by taking the two effects into account. So, the citing decision of an author with a very high μ_1 (and a very low μ_1 respectively) will never (hardly respectively) be affected by the incomplete information, but an agent with a medium range μ_1, especially close to $\overline{\mu}_1$, who would cite under complete information would rather opt not to cite under incomplete information if he takes account of the extra reputation effect.

In this model, an author's attempt to signal by omitting to cite deliberately gives the same reputation benefits across types, but is more costly to a type of lower μ_1 because he is giving up providing more convincing argument to informed readers. Due to a difference in this signaling cost, separation is possible.

3 Empirical Evidence

We use citation data from the RePEc. As of Februry 2007, the RePEc database holds close to 450,000 items of interest in Economics and related fields. In addition, 12,205 authors are registered through the RePEc Author Service,[10] each having contact information and a list of publications catalogued in RePEc. Finally, the Citations in Economics (CitEc) project[11] performs citation analysis on items in RePEc, which then allows to constitute rankings of all registered authors.

An author's overall rank is determined by taking a harmonic mean of his ranks in 31 different rankings based on citations, impact factors and paper downloads, removing the best and worst ranks.[12] From 12,205 registered authors, we collect the information

[10]See http://authors.repec.org/ or Barrueco Cruz, Klink and Krichel (2000).

[11]See http://citec.repec.org/ or Barrueco Cruz and Krichel (2005).

[12]The exact formulas for the variables are too complex to provide in the text. Those who are interested in the formulas may refer to http://ideas.repec.org/top/ or Zimmermann (2007).

given in Table 1.[13](Insert Table 1 here.) In Figure 1, we plot the *RANK_CITED* variable with respect to the author's rank (*RANK*). We exclude the authors whose *RANK_CITED* values are zero. It can indeed happen that none of the cited authors are registered, or that references could not be found for any of the author's works, especially if he has few of them. Thus 9,127 of 12,205 authors are considered in the simulation. We draw random 200 samples out of 12,205 authors and investigate the citation pattern of observed pairs for RANK_CITED and RANK values.[14] Figure 1 reveals that the citation pattern line is not horizontal, that is, the citation pattern is dependent on the author's rank (*RANK*), implying that citation bias does exist. (Insert Figure 1 here.) To show that the slope of the citation pattern line is significantly different from zero, we estimate the following regression equation;

$$RANK\_CITED = \beta_0 + \beta_1 \times RANK + e.$$

Here, the estimate for β_1 is 0.05 with a standard error of 0.002 and thus we can reject the hypothesis that $\beta_1 = 0$. Also, a positive slope of the citation pattern line is consistent with our theoretical result that authors tend to cite other authors with higher ranks than their own.

To examine the citation pattern from another angle, we draw 91 rank groups by assigning about 100 authors to each group according to their ranks. For each author, 1 is given if the *RANK_CITED* value is larger than the *RANK* value[15] and otherwise, 0 is given. Then, the average of the indicator values is computed for each rank group. The graphical result is reported in Figure 2. (Insert Figure 2 here.)

[13]Some suspect that there could be alternative explanations to support the upward citation pattern, and suggest us to check whether stratification could be another possible explanation for it. They argue, for example, that big names tend to touch on major, general subjects (e.g. highly abstract theoretical economics), while relatively incompetent authors tend to work only on minor or special ones (e.g. agricultural economics). However, we do not agree that only high rankers tend to be associated with general issues. Moreover, we believe that even if it is the case, the explanation does not seem to be consistent with the identified pattern. If the explanation were correct, we would obtain a curve which goes upward and then gets flat, because very low rankers also tend to cite only top rankers.

[14]Because a scatter plot does rarely help when the number of observations is 300 or more, we provide the scatter plot with the smoothed line based on randomly drawn samples. For the limitation of the scatter plot, see Acock (2006). The citation pattern line is plotted by using the Lowess smoothing method.

[15]This implies that the selected author's rank is higher than the average rank of his cited authors.

With no citation bias, the graph would decline smoothly. In Figure 2, however, the graph falls rapidly and we clearly observe that the averaged indicator values are recorded as zero from the 24th rank group.[16] This means at least that the authors in the middle range are unlikely to cite the authors with lower ranks than their own. Accordingly, Figure 2 is consistent with Proposition 2 saying that citation bias is more severe among less established authors if we interpret those authors with intermediate ranks as less established ones while interpreting the top ranking authors as established.

To test the citation bias solely due to the reputation effect, we need a proxy for the reputation of an author. We may think of several candidates for the proxy.

First, we pay attention to the difference between *RANK* and *RANK_NW*. We use variable *RANK_NW* as a proxy of the true ability of an author,[17] and variable *RANK* for a proxy of his overall ability including his reputation.[18] We group authors by *RANK_NW* assigning about 400 authors in each group, and take the upper 10% and 50% groups. In Figure 3, we present two regression-fitted lines denoted by the dashed line (upper 10%) and the solid line (upper 50%). (Insert Figure 3 here.) Interestingly, it is displayed that the dashed line shows the negative slope with respect to the proxy variable for an author's reputation, *RANK*. This suggests that a less reputed author is likely to cite high-ranking authors more selectively due to the reputation effect. From the solid curve, it is predicted that the authors up to the 4000th show a negative slope, while those of lower ranks than the 6000th show a positive slope. This can be also interpreted as their reputation effect almost balanced with the correlation effect at the minimum point. Overall, our theoretical result supports a U-shaped curve.[19]

[16] Approxately 2407th - 2518th ranked authors are allocated to the 24th rank group.

[17] An author's performance in terms of journal publication can be a reasonable proxy for his ability insofar as the refereeing process in academic journals is fair. See Kim and Park (2006) for the possibility of the unfair refereeing process especially in single-blinded journals.

[18] We can justify this choice of variable *RANK* for measuring the reputation as follows. As in the argument in footnote 8, risk-averse authors are reluctant to cite an author whose ability is not widely known. In fact, many authors seldom cite unfamiliar names. So, *RANK* of a less reputed author is likely to be lower than his *RANK_NW*.

[19] Table 2 shows that the observed shapes of two fitting curves in Figure 3 are supported by the regression model estimation. The quadratic regression model for the upper 50% indicates the positive and the negative significance for the squared term, and the linear model for the upper 10% indicates the negative significance for variable *RANK*. The squared term for *RANK* in the model with the upper 10%

Second, as an alternative proxy to the recognizability of an author, we use his seniority. More specifically, to distinguish the reputation effect from the correlation effect, we classify authors into two groups, seniors and juniors,[20] and then plot the relation between *RANK_NW* and *RANK_CITED* in Figure 4. While positive slopes of the fitted lines represent the bias due to the correlation effect, a lower fitted line for juniors than for seniors clearly show that there is a bias due to the reputation effect. In other words, juniors are more selective in their citations.

Finally, we add the empirical evidence of the discrimination effect informally discussed in footnote 8 by identifying bias towards citations of authors from prestigious institutions. In fact, an author's affiliation with a well known university helps getting his work widely recognized and frequently cited. Testing the citation bias that occurs due to the author's affiliation, we provide summary statistics in Table 3. (Insert Table 3 here.)

We find that citation bias exists, depending on the author affiliation. Authors affiliated with institutions from the USA or Canada are more likely to be cited than those in other continents. Of course, this may be due to their relatively higher rankings rather than due to citation bias. To examine the citation bias controlled by the rank of authors, we propose the following regression model. Compared to the previous estimation model, we replace the *RANK* variable with the *RANK_NW* variable to avoid the simultaneity problem between *AVE_CITING* and *RANK*;

$$AVE\_CITING = \beta_0 + \beta_1 RANK\_NW + \beta_2 AFFI2 + \beta_3 AFFI3 + e_i,$$

where *AFFI2* (Europe) and *AFFI3* (Others) are dummy variables for the affiliation regions. Considering that 50.1% of 11,599 new number authors have no cited records in the works of other authors, a Tobit model is employed as the estimation approach. The coefficients and standard errors are reported in Table 4. (Insert Table 4 here.)

In this regression, we find that the *RANK_NW* variable is negatively significant. After controlling the author rank, the region dummy variables are still negatively significant at a 5% level. Therefore, the empirical result supports the hypothesis that authors with US or Canada affiliations are more likely to be cited than authors with other regional affiliations.

authors is estimated to be insignificant. (Insert Table 2 here.)

[20]Here, we define junior authors by ones whose publication was within 3 years.

4 Conclusion and Caveats

In this paper, we provided a theoretical model of citation and tested the results empirically. Overall, the empirical results presented in this paper support the hypothesis that there is either individual-based or group (geography)-based citation bias. In particular, we find evidence for the correlation effect, namely that authors prefer to cite better ranked authors to make their claim more legitimate. We also find evidence for the reputation effect, whereby authors cite more selectively to avoid a signal of incompetence when there is uncertainty about their competence.

We acknowledge, however, that authors may also take consideration of other factors, for example psychological or political one in deciding to cite. An author may cite someone's work simply because he is a colleague or because he used to be the author's advisor/student. Or, he may not cite a work just for the reason that he does not like the author personally. Although some citations are an outcome of such personal considerations, the inherent nature of the citation should not be to give a favor to someone, but to cite his work because it is relevant.

One important feature in the citation decision that we neglected to mention in this paper is the network effect in a broad sense. It is often reported that a small group of scholars give mutual favors by citing each other. Also, some physicists[21] recently have identified a hub structure in scientific citation networks and explained it by using preferential attachment. The explanation roughly says that a newcomer in a network (a newly written paper) is more likely to link to an article with more links, that is, more likely to cite an article who is more often cited. In the citation network they found, each node represents a paper, not an author.[22] Indeed, the network structure would be roughly preserved even if each node represents an author instead of an article. Then, our theory of citation bias based on the correlation effect and the reputation effect could provide a rationale for the preferential attachment in this specific context of the citation network. If each node is identified with an author, the preferential attachment, which is very crucial

[21] See, for example, Jeong *et al.* (2003).

[22] In Jeong *et al.* (2003), for instance, a node is associated with a paper published in 1988 in Physical Review Letters.

to a hub structure, can be also interpreted as herding in an economic term,[23] going like "an author tends to cite someone else simply because many people cite him." This may be another source of citation bias.

Finally, it is not easy to establish whether other citation strategies are significant, especially that of adapting citations to the intended outlet: citing editors or potential referees, even being asked by referees to cite them. One could argue that better established authors would give less in to such games or that editors in better journals may not allow such behavior, but this is only anecdotal evidence we cannot verify without data set.

To conclude, there is a significant citation bias in academic journals. The academic tradition of evaluating an author in terms of *RANK* incorporating the number of citations clearly aggravates the bias. On this ground, we believe that *RANK_NW* should be more often used to evaluate an author's performance than *RANK* to mitigate the citation bias.

Appendix

Proof of Proposition 2:

(i) Let I be the set of μ_1 who does not cite in equilibrium. By the definition of $\tilde{\mu}_1$, we have

$$E[P(\omega_1 = T \mid \omega_2)] = \alpha_T \mu_2 + (1 - \alpha_F)(1 - \mu_2) = V(\tilde{\mu}_1),$$

where $V(\mu_1) = \lambda\mu_1 + (1 - \lambda)E(\mu_1 \mid I)$. Then, since $V(\mu_1)$ is increasing in μ_1, it is clear that $E[P(\omega_1 = T \mid \omega_2)] < V(\mu_1)$ for all $\mu_1 > \tilde{\mu}_1$ and that $E[P(\omega_1 = T \mid \omega_2)] > V(\mu_1)$ for all $\mu_1 < \tilde{\mu}_1$.

(ii) By the definition of $\overline{\mu}_1$, we have $E[P(\omega_1 = T \mid \omega_2)] = \overline{\mu}_1$. This implies that

$$\overline{\mu}_1 = \lambda\tilde{\mu}_1 + (1 - \lambda)E(\mu_1 \mid I). \tag{2}$$

Note that $E(\mu_1 \mid I) > \tilde{\mu}_1$, because $I = \{\mu_1 \mid \mu_1 > \tilde{\mu}_1\}$. Therefore, it follows that $\overline{\mu}_1 > \tilde{\mu}_1$.

(iii) Total differentiation of (2) directly shows the monotonicity of $\tilde{\mu}_1(\lambda)$ with respect to λ.

[23] See Banerjee (1992), and Bikhchandani, Hirshleifer and Welch (1992) for informational explanations of herding.

References

[1] Acock, A. C. (2006), A Gentle Introduction to Stata, Stata Press Publication, College Station, TX

[2] Banerjee, A. (1992), 'A Simple Model of Herd Behavior,' Quarterly Journal of Economics, 107, pp. 797-817.

[3] Barrueco Cruz, J., Klink, M. and Krichel, T. (2000), 'Personal Data in a Large Digital Library,' In: Proceedings 4th European Conference on Research and Advanced Technology for Digital Libraries, Lisboa.

[4] Barrueco Cruz, J. and Krichel, T. (2005), 'Building an Autonomous Citation Index for Grey Literature: RePEc, The Economics Working Papers Case,' The Grey Journal: An International Journal on Grey Literature, 1 (2) pp. 91-97.

[5] Bikhchandani, S., Hirshleifer, D. and Welch, I. (1992), 'A Theory of Fads, Fashion, Custom, and Cultural Change as Informational Cascades,' Journal of Political Economy, 100, pp. 992-1026.

[6] Jeong, H., Neda, Z. and Barabasi, A-L. (2003), 'Measuring Preferential Attachment in Evolving Networks,' Europhysics Letters, 61, pp. 567-572.

[7] Kim, J.-Y. and Park, J. (2006), 'On Prejudice,' Scottish Journal of Political Economy, 53, pp. 505-522.

[8] Krichel, T. (2000), 'Working towards an Open Library for Economics: The RePEc project,' presented at the "PEAK 2000 Conference: The Economics and Use of Digital Library Collections", http://openlib.org/home/krichel/papers/myers.html

[9] Palevitz, B. (1997), 'The Ethics of Citation: A Matter of Science's Family Values,' The Scientist, 11, pp. 8.

[10] Wright, M. and Armstrong, J.S. (2007), 'The Ombudsman: Verification of Citations: Fawlty Towers of Knowledge?' MPRA Paper No. 4149, available at http://mpra.ub.uni-muenchen.de/4149/

[11] Zamora Bonilla, J.P. (2005), 'Scientific Studies and the Theory of Games,' forthcoming in Perspectives on Science.

[12] Zimmermann, C. (2007), 'Academic Rankings with RePEc,' University of Connecticut, working paper.

<Table 1: variable description>

variable	***Description***
RANK	Author's overall rank
RANK_NW	Author's rank determined by his number of works weighted by the simple impact factor of their series
RANK_CITED	Average rank of authors cited in this authors' works: when several authors are ranked for a cited work, the highest rank is taken
NW_CITING	The number of works citing this author
NW_WORKS	The number of this author's publications
AVE_CITNG	*NW_CITING / NW_WORKS*
AFFI	Author's affiliation: for multiple affiliations, the first affiliation is chosen

< Table 2: Quadratic and linear estimation >

Quadratic regression model: upper 50% group

variable	coefficient	standard error	t-value	p-value
RANK	-0.556	0.180	-3.08	0.002
$(RANK)^2$	4.19e-05	1.71e-05	2.45	0.015

Linear regression model: upper 10% group

variable	coefficient	standard error	t-value	p-value
RANK	-0.105	0.034	-3.07	0.002

<Table 3: Summary statistics>

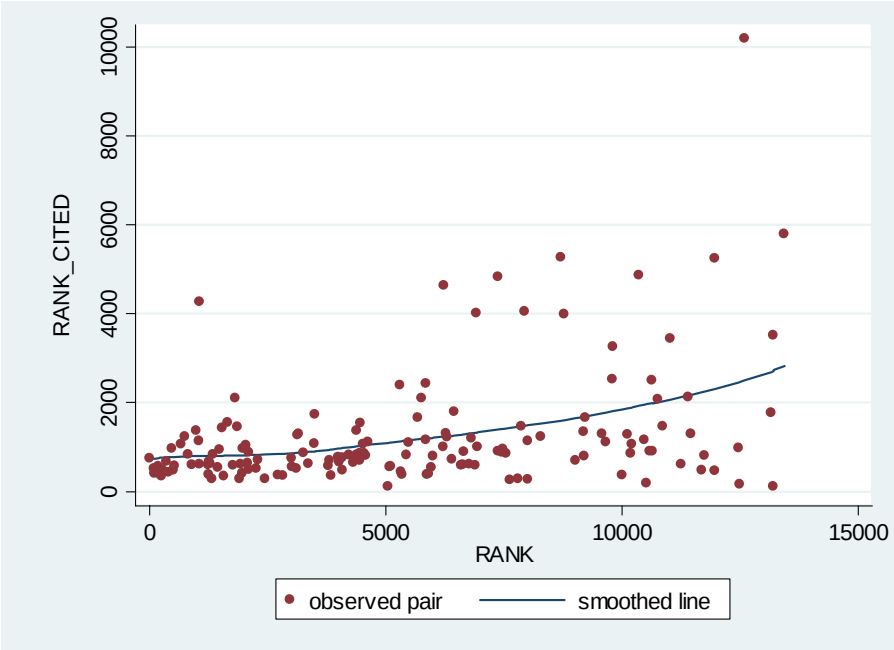
Affiliated region	***obs***	***NW_CITING***	***AVE_CITING***
USA & Canada	3,743	52.1	1.27
Europe	6,394	11.42	0.39
Others	1,462	5.98	0.27

Note: 11,599 of 12,205 authors are considered and authors with no explicit affiliation are excluded.

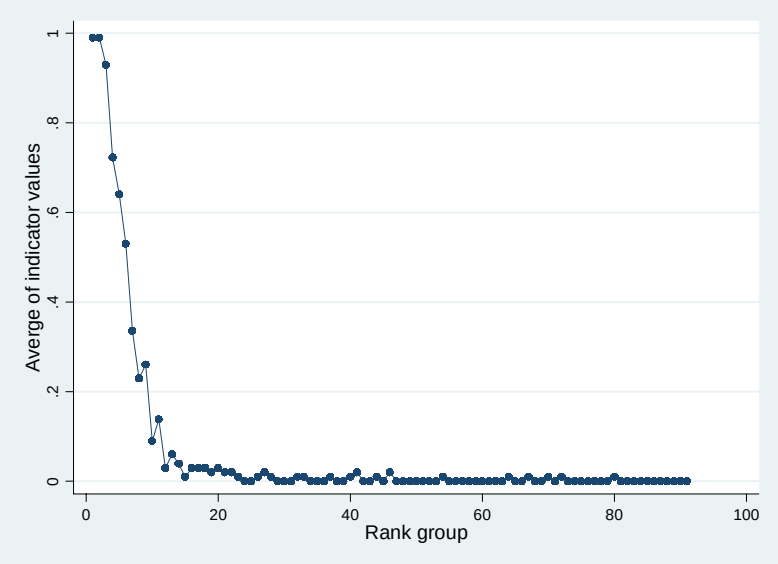
<Table 4: Tobit model estimation>

variable	*coefficient*	*standard error*	*t-value*	*p-value*
RANK_NW	-0.00066	0.00001	-60.76	0.000
AFFI2	-0.438	0.0648	-6.76	0.000
AFFI3	-0.452	0.105	-4.29	0.000
logL	-15536.89			

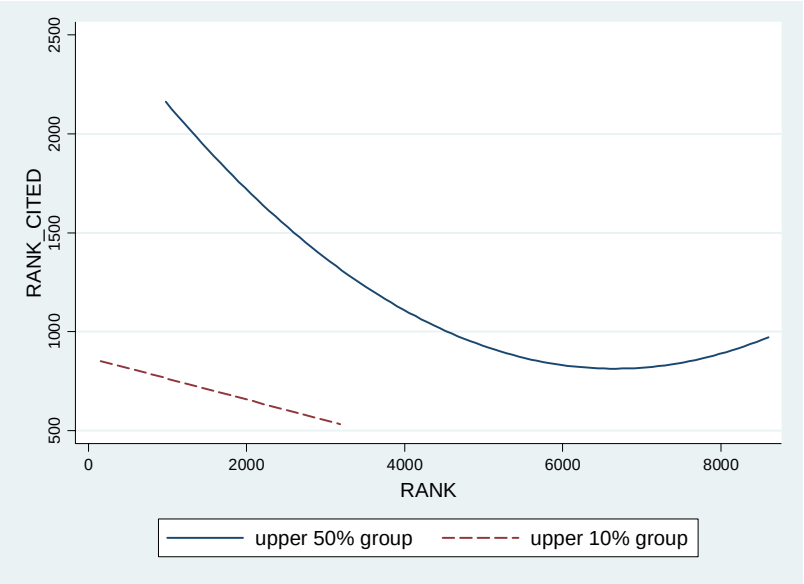
<Figure 1: *RANK_CITED* vs. *RANK*>



<Figure 2: Average of indicator values for each rank group>



<Figure 3: upper 10% and 50% *RANK_NW* groups>



<Figure 4: Seniors vs. Juniors>