A theory of citations

Wojciech Olszewski*

June 2017

Abstract

We propose a model in which researchers maximize their number of citations in subsequent papers. The equilibrium is inefficient, because they distort their effort towards writing on popular topics. This inefficiency is smaller when citations in time-remote papers have a higher weight, and when citations in higher-ranked journals have a higher weight. We also discuss some other issues, for example, the relation between the quality of papers and the number of citations, and the relation between efficiency and the spread of citations across papers.

Preliminary and Incomplete

1 Introduction

Citations have always been playing an important role of accessing in adamemia. Papers with extremely high number of citations were always in some ways admired, and researchers were referring to citations in their evaluations of other researchers. In the last years, there has even been a surge of interest in using citations to assess researchers, academic journals, and various research institutions. Academic journals increased their emphasis on citation indices. References to citations became more common in evaluations. Studies of whether citation indexes can provide a compelling assessment of scientists and academic departments have been proposed (see Elison, 2010 and 2012).

Previous research is focused on suggesting various citation indexes, among which the h-index (Hirsch, 2005) is probably most commonly referred to. This includes axiomatic approach to characterize some existing indexes, or to search for a "right index". A recent article by Perry and Reny (2016) contains a summary of the literature on citation indexes.

We are interested in how the importance of citations affects research, and more specifically, the choice of research topics. In our model, researchers are better qualified to work (have an advantage in working) on

*Department of Economics, Northwestern University, Evanston, IL 60208 (e-mail: wo@northwestern.edu). I thank Aisleen Bohren, Natalia Fabra, Phil Reny, Joel Sobel and the audience at Northwestern University for very helpful comments, suggestions and discussion.

some topics than on others, but if this advantage is not large, and they expected to be cited less frequently, they may prefer not following their advantage and write on these other topics. This creates a spillover, and researchers even with larger advantage prefer not following this advantage and write on other topics. The outcome is socially inefficient.

The size of this inefficiency depends on various factors, such as the distribution of researchers talent to working on various topics, or the strength of citation incentives. Other motives, disregarded in the present paper, are also important, for example: pioneering research in some area, publishing record, or other forms of peer recognition.

The model provides an insight on numerous natural questions, both static and dynamic, in the context of citations: What would be the effect of assigning a higher weight to citations in time-remote papers? Or counting citations in more selective ways? It enables to discuss some other issues regarding the relation between the quality of papers and the number of citations, and the relation between efficiency and the spread of citations across papers. Or the dynamics of research motivated by citations in response to a revival in certain areas (for example, inspired by the availability of new types of data sets).

The model can certainly be extended in various manners. We discuss some important extensions in Section 7. Many other possible extensions are not addressed even in the discussion. For example, many interesting and important questions regarding optimal mechanism design seem to require exploring richer models. We believe, however, that the effects described in the present papers are robust, and will be present in any sensible model in which researchers are motivated by citations. Indeed, the driving force of our analysis is the premise that some researchers distort their effort from writing on topics in which their value added would be the highest towards writing on topics which are more popular in their profession.

2 Basic Model Report

A random i.i.d. number of researchers live (are active) in each period. A researcher may conduct research (write a paper) on one of two topics: topic X and topic Y. For example, if the population represents macroeconomists, each of whom may work on modeling the life-time income process, or regularities in the behavior of exchange rates (among a variety of other topics, which are excluded from the basic model for simplicity); and if the population represents microtheorists, each of them may work on auction design or behavioral economics.

Each researcher is characterized by a 3-dimensional type. First, she may be *partizan* or *strategic*. (In other words, she may be internally or externally motivated.) The researcher is of the former type with probability $1-p$ and of the latter type with probability p. In addition, she is characterized by the social value of her papers $(q_X, q_Y) \in [0,\infty)^2$. That is, if the researcher writes a paper on topic $i = X, Y$, this paper's social value is q_i. In particular, a researcher may have an advantage in writing on topic X ($q_X > q_Y$), or an advantage in writing on topic Y ($q_X < q_Y$). Her type is each researcher's private information. (In fact, it is irrelevant that only the researcher knows whether she is partizan or strategic; this component of her type could be commonly known.) We will from time to time call (q_X, q_Y) the researcher's type, if it follows from the context whether we mean a strategic or partizan researcher. The distribution of vectors (q_X, q_Y) has a full-support Lebesgue measurable density f, which is commonly known among researchers.

It is assumed that a paper on topic $i = X, Y$ by a researcher of type (q_X, q_Y) would (in expectation) be cited in each subsequent period q_i times if all researchers were writing on this topic, but would not be cited at all if all researchers were writing on the other topic. And proportionally, if a fraction M_X of researchers in a subsequent period writes on X, and a fraction M_Y of researchers in a subsequent period writes on Y, then the paper is expected to be cited $M_X q_X$ and $M_Y q_Y$ times, respectively.

This assumption means that citations are the right measure of social value. A paper is cited by a subsequent researcher only when she finds the existing paper useful in her research. And the papers that are more likely to be useful coincide with the papers of higher social value. Of course, one can consider other measures of social value, which may not be aligned with citations. Such measures would, however, introduce some exogenous inefficiency, since researchers would not be pursuing social goals, even in the absence of any strategic interaction. In turn, our basic model will exhibit only endogenous inefficiency coming from strategic interaction.

A partizan researcher is assumed to write on the topic with higher q, that is, on topic X if $q_X > q_Y$ and on topic Y if $q_X < q_Y$. The choice of topic when $q_X = q_Y$ will turn out irrelevant. A strategic researcher is assumed to choose the topic which generates a higher expected payoff. Her payoff is given by

$$(1 - \delta) \sum_{n=1}^{\infty} \delta^n \tau_n, \tag{1}$$

where δ stands for the common discount rate, and τ_n is the number of times the researcher is cited in period n. That is, a strategic researcher chooses a topic for her paper based on the beliefs regarding the fractions of researchers that will be working on each topic in subsequent periods.

Note that with a full-support density f and a number of researchers approaching infinity, the objective of maximizing the absolute number of citations is in approximation equivalent to maximizing the researchers' ranking with respect to the number of citations.

3 Equilibria

3.1 First-best outcome

The social optimum is attained when each researcher writes on the topic she has an advantage. This is topic X if $q_X > q_Y$, and topic Y if $q_X < q_Y$, and this social optimum would be attained if all researchers were partizan. To see why strategic researchers may have distorted incentives, consider a researcher maximizing (1) who happens to live in a society in which all other researchers are partizan. Which topic such a researcher would choose? Denote by μ_X and μ_Y the measures of sets $\{(q_X, q_Y) \in [0, 1]^2 : q_X > q_Y\}$ and $\{(q_X, q_Y) \in$

$[0,1]^2 : q_X < q_Y\}$, respectively. That is,

$$\mu_X = \int_0^{\infty} \left(\int_{q_Y}^{\infty} f(q_X, q_Y) dq_X \right) dq_Y,$$

and

$$\mu_Y = \int_0^{\infty} \left(\int_{q_X}^{\infty} f(q_X, q_Y) dq_Y \right) dq_X.$$

If $\mu_X = \mu_Y$, then strategic researchers would make the same choices as partizan researchers. But if $\mu_X > \mu_Y$, then the payoff of a strategic researcher of type (q_X, q_Y) from writing on topic i is $\delta q_i \mu_i$, which means that such a researcher will write on topic Y only when

$$\frac{q_Y}{q_X} > \frac{\mu_X}{\mu_Y} > 1.$$

That is, strategic researchers have incentives distorted towards writing on X. Symmetrically, strategic researchers have incentives distorted towards writing on Y when $\mu_X < \mu_Y$. Throughout the paper, we assume that $\mu_X > \mu_Y$. It will be convenient to normalize μ_X and μ_Y to the values such that $\mu_X + \mu_Y = 1$.

One can argue here that the market could easily remove the inefficiency by comparing citations divided by the average number of citations on a given topic. Applying this kind of measures seems to be present in practice while comparing across fields, such as macroeconomics and econometrics. One may even take into account the differences within fields, such as decision theory and game theory. However, a very fine classification of topics would rather be impractical, or even impossible. How, for example, one would classify a paper looking for an optimal mechanism in a setting in which agents have some behavioral biases?

3.2 Description of equilibria

In the previous subsection, we informally argued that (when $\mu_X \neq \mu_Y$) the efficient behavior is not an equilibrium. The game has, however, an inefficient equilibrium. We must first introduce formally our equilibium concept. A strategy prescribes a decision regarding the choice of topic for each type of each researcher, contingent on the past choices of other researchers. This implicitly includes the calendar time of making the decision. Strategies are assumed to be Lebesgue measurable, i.e., the set of types choosing each topic is Lebesgue measurable.

If a strategy is independent of the past choices of other researchers, then we call the strategy *history-independent*. History-independent strategies may, however, depend on calendar time. If in addition the strategy is independent of calendar time, then we call the strategy *stationary*.

In equilibrium, the prescribed strategies must give each type of strategic researcher an expected payoff which weakly exceeds the expected payoff from choosing the other topic, given that other researchers make the prescribed decisions. If in an equilibrium strategy profile is history-independent or stationary, then we call the equilibrium history-independent or stationary, respectively.

We study history-independent equilibria. The reason is that in the present model past choices of other researchers have no direct payoff implications on the players in a continuation game. So, they could only serve as some kind of sun spots, or players would be disciplining other players whose actions are irrelevant for them. Equilibria in history-dependent strategies would seem more reasonable in a richer model, in which players' productivity in each topic depends on the existing research on this topic.

We begin the analysis with two examples.

Example 1. In this example, we explore a special case of our model in which only one researcher lives in each period. Suppose that $p = 1/2$, i.e., a half of researchers are partizan, while the other half are strategic, and that the density f is equal to 4/3 on $\{(q_X, q_Y) \in [0,1]^2 : q_X > q_Y\}$, and is equal to 2/3 on $\{(q_X, q_Y) \in [0,1]^2 : q_X < q_Y\}$. Then, q_i, $i = X, Y$, represents the probability that a researcher who writes on topic i will be cited by a subsequent researcher who writes on the same topic. (Recall that the researcher will never be cited by subsequent researchers who write on the other topic.

We will explore only stationary equilibria. Denote by M_X an equilibrium probability that a researcher writes on topic X, and by $M_Y = 1 - M_X$ the equilibrium probability that a researcher writes on topic Y. Due to the presence of partizan researchers, both M_X and M_Y are positive. The types (q_X, q_Y) such that $\delta q_X M_X = \delta q_Y M_Y$, or equivalently, such that

$$q_Y = \frac{M_X}{M_Y} q_X$$

are indifferent between writing on X and writing on Y. The segment of points (q_X, q_Y) that satisfy this equation divides the square $[0,1]^2$ into two parts. All strategic researchers with type (q_X, q_Y) to the right of this segment write on topic X, and all strategic researchers with (q_X, q_Y) to the left write on topic Y.

Case 1 ($M_X/M_Y \leq 1$). In this case,

$$M_X = \frac{1}{2}\frac{4}{3}\frac{1}{2}\frac{M_X}{M_Y} + \frac{1}{2}\frac{4}{3}\frac{1}{2}.$$

Since $M_Y = 1 - M_X$, this equation says that

$$M_X = \frac{1}{3}\frac{M_X}{1 - M_X} + \frac{1}{3}.$$

It is easy to check that this quadratic equation has no solution. Therefore there is no equilibrium in this case.

Case 2 ($M_X/M_Y > 1$). In this case,

$$M_Y = \frac{1}{2}\frac{2}{3}\frac{1}{2}\frac{M_Y}{M_X} + \frac{1}{2}\frac{2}{3}\frac{1}{2}.$$

Since $M_X = 1 - M_Y$, this equation says that

$$M_Y = \frac{1}{6}\frac{M_Y}{1 - M_Y} + \frac{1}{6}.$$

It is easy to check that this quadratic equation has a unique solution

$$M_Y = \frac{3 - \sqrt[2]{3}}{6}.$$

Therefore there is a unique stationary equilibrium, up to the choices of types such that $q_X M_X = q_Y M_Y$, which have measure zero. This equilibrium is depicted in Figure 1.

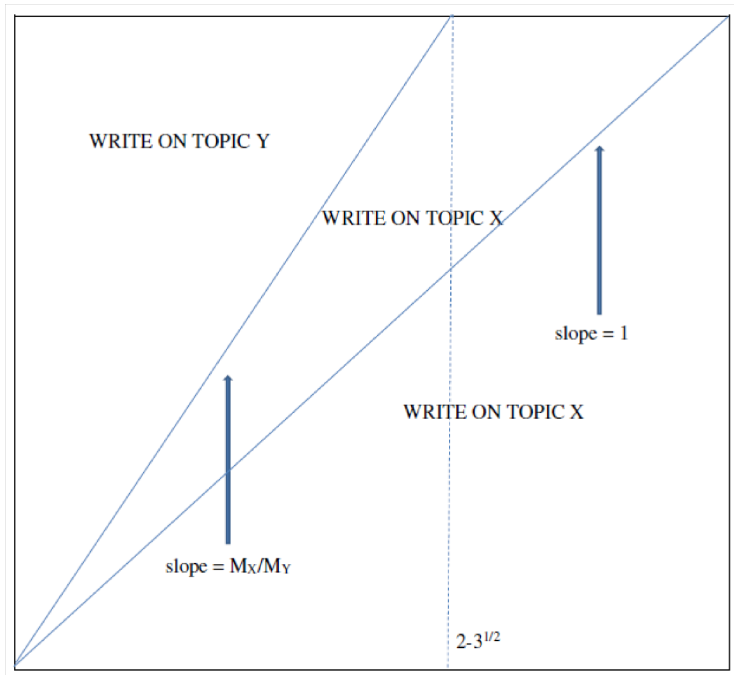

Figure 1. The unique stationary equilibrium strategies for the uniform distribution of types

Example 2. Suppose again that $p = 1/2$, i.e., a half of researchers are partizan, while the other half are strategic. Suppose also that the density f is equal to e^{-x-y} multiplied by 4/3 on $\{(q_X, q_Y) \in [0,1]^2 : q_X > q_Y\}$, and multiplied by 2/3 on $\{(q_X, q_Y) \in [0,1]^2 : q_X < q_Y\}$. Thus, $\mu_X = 2/3$ and $\mu_Y = 1/3$.

We will again explore only stationary equilibria. As in Example 1, denote by M_X an equilibrium fraction of researchers who write on topic X, and by M_Y the fraction of researchers who write on topic Y. The indifferent types (q_X, q_Y) are such that $\delta q_X M_X = \delta q_Y M_Y$, or equivalently, such that

$$q_Y = \frac{M_X}{M_Y} q_X.$$

The line of such points (q_X, q_Y) divides the quadrant $[0, \infty)^2$ into two parts. All strategic researchers with (q_X, q_Y) below this segment write on topic X, and all strategic researchers with (q_X, q_Y) above write on topic Y.

Consider the following two cases:

Case 1 ($M_X/M_Y > 1$). In this case, we have that

$$M_Y = \frac{1}{2}\frac{2}{3}\int_0^{\infty}\left(\int_0^{y\frac{M_Y}{M_X}} e^{-x}dx\right) e^{-y}dy + \frac{1}{2}\frac{2}{3}\frac{1}{2}.$$

The first component of the right-hand side represents the probability that a researcher is strategic and writes on topic Y, and the second component represents the probability that a researcher is partizan and writes on topic Y.

Integrating, we obtain that

$$\int_0^\infty \left(\int_0^{y \frac{M_Y}{M_X}} e^{-x} dx \right) e^{-y} dy = M_Y,$$

so that

$$M_Y = \frac{1}{3} M_Y + \frac{1}{6} = \frac{1}{4},$$

what is consistent with $M_X/M_Y > 1$.

Case 2 ($M_X/M_Y \leq 1$). In this case,

$$M_X = \frac{1}{2}\frac{4}{3} \int_0^\infty \left(\int_0^{x \frac{M_X}{M_Y}} e^{-y} dy \right) e^{-x} dx + \frac{1}{2}\frac{4}{3}\frac{1}{2},$$

which yields

$$M_X = \frac{2}{3} M_X + \frac{1}{3} = 1,$$

what contradicts the assumption that $M_X/M_Y \leq 1$.

Therefore, the model has a unique (stationary, and up to type sets of measure zero) equilibrium, in which 75% of researchers work on topic X, and the remaining 25% work on topic Y. Whereas in the efficient outcome, only 2/3 of researchers would work on topic X, and the remaining 1/3 would work on topic Y. The inefficient decision are made by strategic researcher with types (q_X, q_Y) between lines $q_Y = q_X$ and $q_Y = 3q_X$.

The settings from Examples 1 and 2 had unique stationary equilibria. Below, we provide a condition on the primitives of our model (Assumption 1) under which the model has a unique equilibrium, and this equilibrium is stationary. We will assume this condition through the main text. In Appendix, we will characterize equilibria in the general case when the condition may not satisfied. It will be convenient to introduce some notation up front, prior to stating the condition and our results.

Let M_i^t be the fraction of researchers who choose topic $i = X, Y$ in period t. This number is well defined for any history-independent strategies. When strategies are in addition stationary, then $M_i^t = M_i$ is constant over time. Given any (history-independent, not necessarily stationary) strategies of researchers living in periods $t, t+1, ...$, let

$$\overline{M}_i^t = (1 - \delta) \sum_{n=t}^{\infty} \delta^{n-t} M_i^n.$$

This is often called in the literature the occupation measure - in this case - of topic i from time t on. If the strategies are stationary, then $\overline{M}_i^t = M_i$. Finally, let

$$g_X(M_X) := p \Pr\{(q_X, q_Y) : q_X M_X > q_Y (1 - M_X)\} + (1 - p)\mu_X$$

and

$$g_Y(M_Y) := p\,\Pr\{(q_X, q_Y) : q_Y M_Y > q_X(1 - M_Y)\} + (1-p)\mu_Y.$$

Notice that functions g_X and g_Y are fully determined by the distribution of researchers' types.

Assumption 1. Function g_X has a unique fixed point.

If $M_X + M_Y = 1$, then $g_Y(M_Y) + g_X(M_X) = 1$. So, it would be equivalent to assume that function g_Y has a unique fixed point. Notice that $g_i(M_i)$ always has a fixed point M_i. Indeed, function g_X is continuous (because the distribution of types (q_X, q_Y) is continuous), and $g_X(0) > 0$ and $g_X(1) < 1$ (because of the presence of partizan researchers). For some distributions of researchers' types, g_X has multiple fixed points. But for many distributions of interest - including uniform, exponential (as we have seen in Examples 1 and 2) or normal truncated to the quadrant $q_X, q_Y \geq 0$, $g(M_i)$ - g_X has a unique fixed point.

The following proposition characterizes the equilibria of our basic model under Assumption 1. Its proof is relegated to Appendix.

Proposition 1 *The model has a unique equilibrium. This equilibrium is stationary. In the equilibrium, M_i is the fixed point of g_i. A researcher chooses topic X when $M_X q_X > M_Y q_Y$, and chooses topic Y when $M_Y q_Y > M_X q_X$; when $M_X q_X = M_Y q_Y$ (which event has probability 0), a researcher is indifferent between the two topics.*

4 Comparative statics

We begin with studying the effects of an increase in the fraction of partizan researchers in the population, and of an increase in the asymmetry across topics. The former increase is modeled by decreasing p. In the latter case, we wish to increase the ration of μ_X to μ_Y without introducing any changes in the relative density across types such that $q_Y > q_X$, or across types such that $q_Y < q_X$. Therefore, the latter increase will be modeled as multiplying the density f on the set $\{(q_X, q_Y) : q_Y > q_X\}$ by an $\alpha < 1$, and multiplying the density f on the set $\{(q_X, q_Y) : q_Y < q_X\}$ by $\beta > 1$ such that $\alpha\mu_Y + \beta\mu_X = 1$.

Proposition 2 *(i) If $p'' < p'$, then $M_Y'' > M_Y'$. (ii) For any $\alpha < 1$, $M_Y^\alpha/\mu_Y^\alpha < M_Y/\mu_Y$; in particular, $M_Y^\alpha < M_Y$.*

Proposition 2 (i) implies that an increase in the fraction of partizan researchers in the population leads to a more efficient outcome. Actually, there are more efficient decisions, and an average individual decision is more efficient. This happens for two reasons: (a) The fraction of partizan researchers is higher, and such researchers make efficient decisions; (b) A smaller set of strategic researchers with advantage in topic Y strategize by writing on topic X.

By Proposition 2 (ii), an increase in the asymmetry across topics reduces the set of types of strategic researchers with advantage in topic Y who write on topic Y. Proposition 2 (ii) shows even more, that the size of this set decreases as the fraction of the set of all researchers with advantage in topic Y, though this latter set decreases. This strategic effect reduces efficiency. But the direct effect of reducing the fraction of researchers with advantage in topic Y in the population enhances efficiency. So, the total effect on aggregate efficiency cannot be determined. For example, the strength of the direct effect depends on the ratio of the researchers who strategize and for whom $q_Y - q_X$ is large to the researchers who strategize and for whom $q_Y - q_X$ is small. In turn, the strategic effect is independent of this ratio.[1]

Proof. It follows from equation $g_Y(M_Y) = M_Y$ and Assumption 1 that the graph of g_Y intersects the diagonal from above to below at $M_Y < 1/2$. When p decreases, the graph of g_Y for $M_Y < 1/2$ moves up, because $\Pr\{(q_X, q_Y) : q_Y M_Y > q_X(1 - M_Y)\} < \mu_Y$ for $M_Y < 1/2$. So, the unique fixed point M_Y of g_Y becomes larger.

When the density f on the set $\{(q_X, q_Y) : q_Y > q_X\}$ is multiplied by an $\alpha < 1$, the graph of g_Y moves down for all M_Y. So, the unique fixed point M_Y of g_Y becomes smaller. Moreover, the values of g_Y for all $M_Y \leq 1/2$ are scaled down by α. Since $g_Y(1/2) = \mu_Y$ and $\alpha g_Y(1/2) = \mu_Y^\alpha$, and $M_Y = g_Y(M_Y)$, we would have $M_Y^\alpha/\mu_Y^\alpha = M_Y/\mu_Y$ if M_Y^α were equal to $\alpha g_Y(M_Y)$. However, $M_Y^\alpha < M_Y$, so $M_Y^\alpha = \alpha g_Y(M_Y^\alpha) < \alpha g_Y(M_Y)$, which implies that $M_Y^\alpha/\mu_Y^\alpha < M_Y/\mu_Y$. ■

4.1 Uniform across topics versus topic specific talent

It seems less clear ex ante, whether fields in which talent is uniform across topics are more efficient than fields in which talent is rather topic specific. We model fields in which talent is less uniform across topics by moving the mass of (q_X, q_Y) away from the diagonal. More precisely, for any pair of random variables (q'_X, q'_Y) and (q''_X, q''_Y) taking values in $[0, \infty)^2$, we say that (q''_X, q''_Y) is obtained from (q'_X, q'_Y) by moving mass away from the diagonal if $(q''_X, q''_Y) = (q'_X, q'_Y) + (\varepsilon_X, \varepsilon_Y)$ where $(\varepsilon_X, \varepsilon_Y)$ is a random variable with the following property: When the realization of (q'_X, q'_Y) is such that $q'_X > q'_Y$, then ε_X takes on only nonnegative values and ε_Y takes only nonpositive values. And when the realization of (q'_X, q'_Y) is such that $q'_X < q'_Y$, then ε_X takes on only nonpositive values and ε_Y takes only nonnegative values. The distribution of $(\varepsilon_X, \varepsilon_Y)$ can depend on the realization of (q'_X, q'_Y), but we will omit this relation in our notation. The concept is illustrated in Figure 2.

[1] We cannot say either whether there are more or less inefficient decisions, or whether an average individual decision is more or less inefficient.

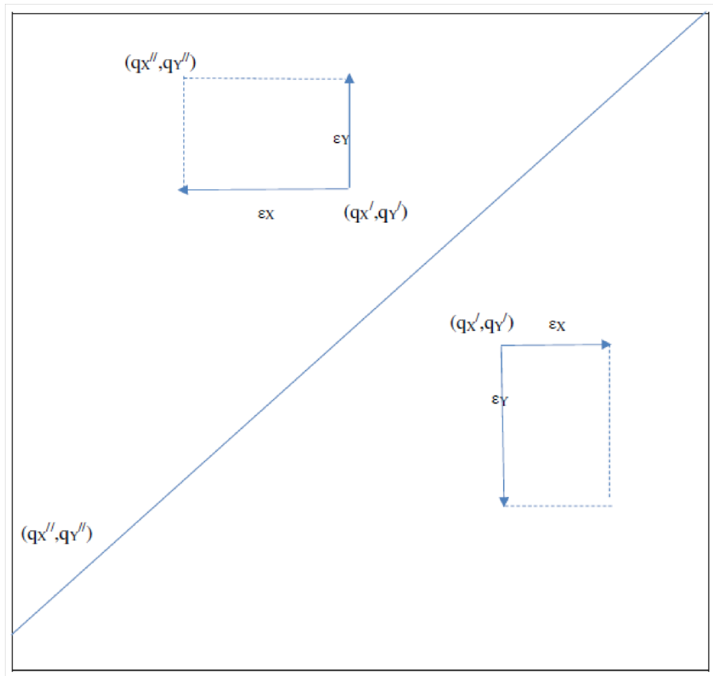

Figure 2. Moving mass away from the diagonal

This notion is a two-dimensional version of Diamond and Stiglitz (1974)'s simple mean-preserving spread.[2] Their concept defined for one-dimensional random variables with cdf's H and G requires that $G(t) \leq H(t)$ for all $t \geq t^*$ and $H(t) \leq G(t)$ for all $t \leq t^*$ for some t^*. It can be shown that (q''_X, q''_Y) is obtained from (q'_X, q'_Y) by moving mass away from the diagonal iff the probability of any set of the form $\{(q''_X, q''_Y) : q''_X > q^*_X$ and $q''_Y < q^*_Y\}$, where $q^*_X > q^*_Y$, is higher than the probability of the set $\{(q'_X, q'_Y) : q'_X > q^*_X$ and $q'_Y < q^*_Y\}$, and the probability of any set of the form $\{(q''_X, q''_Y) : q''_X < q^*_X$ and $q''_Y > q^*_Y\}$, where $q^*_X < q^*_Y$, is higher than the probability of the set $\{(q'_X, q'_Y) : q'_X < q^*_X$ and $q'_Y > q^*_Y\}$.

Proposition 3 *If (q''_X, q''_Y) is obtained from (q'_X, q'_Y) by moving mass away from the diagonal, then $M''_Y > M'_Y$.*

In an interpretation, there is more strategizing when talent is uniform across topics than when talent is topic specific. A larger number of researchers make inefficient decisions. Indeed, the set of types who write on the topic which is not their advantage is larger in the former case than in the latter case. And if (q'_X, q'_Y) makes an efficient decision, so does $(q''_X, q''_Y) = (q'_X, q'_Y) + (\varepsilon_X, \varepsilon_Y)$. Of course, this strategizing is inefficient. However, the total effect on aggregate efficiency is ambiguous, because moving mass away from the diagonal has a direct negative effect on the efficiency of individual decisions. Namely, the efficiency loss coming from

[2] Another version of simple mean-preserving spread, closer in form to the version used in the present paper was used in Klabjan et al. (2014).

the types who under both scenarios write on topic X but have an advantage in writing on topic Y is larger for (q''_X, q''_Y) than for (q'_X, q'_Y).

It is easy to see that the direct effect may be weaker or stronger to the strategic effect in terms of aggregate efficiency. For example, moving away from the diagonal may be nontrivial only within the region of types which strategize. In such a case, there is no strategic effect, but the direct effect is negative. Or moving away from the diagonal may take mass from the region of types which strategize to the region of types which do not do so. In such a case, the direct effect is "subsumed" with the opposite sign into the strategic effect.

Proof. Observe that $\Pr\{(q'_X, q'_Y) : q'_Y M_Y > q'_X(1 - M_Y)\} < \Pr\{(q''_X, q''_Y) : q''_Y M_Y > q''_X(1 - M_Y)\}$ for all M_Y. Indeed, for any given M_Y by adding $(\varepsilon_X, \varepsilon_Y)$ to (q'_X, q'_Y): (i) we move some mass from the region of types that choose X to the region of types that choose Y; (ii) we move some mass within the region of types that choose X, but (iii) we never move any mass from the region of types that choose Y to the region of types that choose X. This argument is illustrated in Figure 3. So, the graph of g_Y moves up for all M_Y when we replace random variable (q'_X, q'_Y) with random variable (q''_X, q''_Y). This yields $M''_Y > M'_Y$.

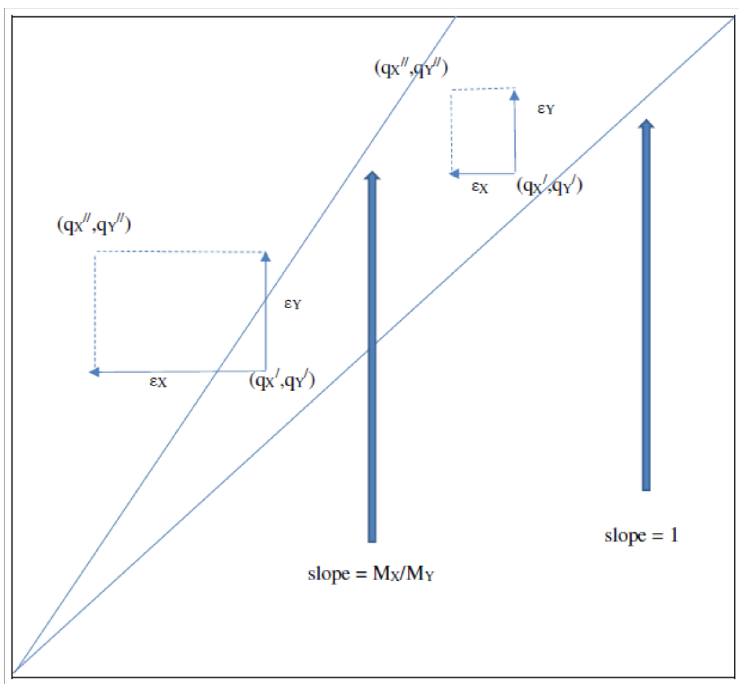

Figure 3. For the lower (q'_X, q'_Y), adding $(\varepsilon_X, \varepsilon_Y)$ results in the effect described in (i); and for the upper (q'_X, q'_Y), adding $(\varepsilon_X, \varepsilon_Y)$ results in the effect described in (ii)

■

5 Implications

5.1 Is a higher spread in citations across papers a signal of discipline's inefficiency?

Different fields, even within a discipline differ in the distribution of citations across papers. In some areas, the numbers are more even across papers than in others. Such data can be easily obtained in practice. In turn, it is much more difficult, if at all possible, to estimate the distribution of researchers' types. Therefore, it would be useful if we could make any conclusion regarding efficiency based on the structure of citations, with no reference to the distribution of researchers' types.

Some conclusions come out from our comparative statics. The strategic effects described in the previous section always increase the number of citations of the papers on topic X, and decrease the number of citations of papers on topic Y. This suggests that a higher spread in the number of citations is rather a signal of inefficiency. However, the differences in the spread of citations may come from various sources, and may be caused by direct effects. If two disciplines differ in the fraction of partizan versus strategic researchers, then the discipline with a higher spread of citations is in aggregate less efficient. Actually, there are more inefficient decisions, and an average individual decision is more inefficient. But if in some discipline people talented in doing research on one topic are also talented in doing research on other topics, the aggregate inefficiency may be smaller despite a higher spread of citations. In this case, we only conclude that a higher spread implies a higher number of inefficient decisions, but the inefficiency involved in an average decision may actually be lower. In addition, in some disciplines asymmetries across topics may be more substantial than in others, with some topics being more productive than others. In such cases, total effects may be driven by direct effects.

Thus, it seems fair only to conclude that if the relative productivity of topics in two disciplines is comparable, then the discipline with a higher spread of citations across papers is likely to have a higher number of inefficient research decisions regarding the choice of topics.

5.2 Are better papers cited more frequently?

Better papers are defined in the model as the papers having a higher q. So by definition, they have an advantage in terms of citations. They are more likely to be cited by each subsequent researcher who writes on the same topic. This does not yet mean that they are cited more frequently in equilibria. A strategic effect which may reduce, even overturn, this direct effect. Namely, there is an outflow of potential papers of any quality q on topic Y to papers of lower quality on topic X, and an inflow of papers of quality q on topic X from potential papers of higher quality on topic Y. These outflow and inflow are depicted in Figure 4. They both increase the number of citations per paper of any quality q, but the strength of these strategic effect may be different for different q's.

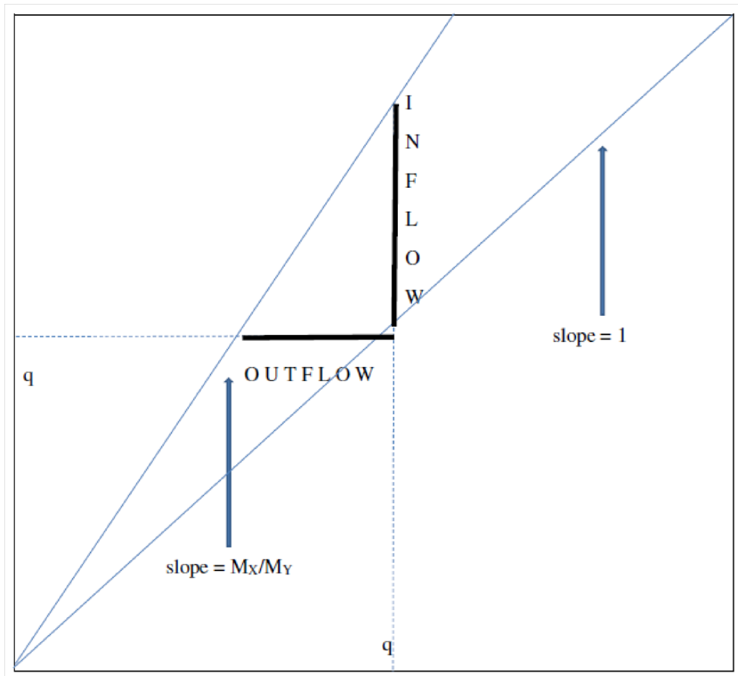

Figure 4.

The way in which these effects vary with q depends very much on the distribution of types, and no general conclusion can be derived. However, one can obtain some insight by looking at a density f that is symmetric across the diagonal, and is multiplied on the set $\{(q_X, q_Y) : q_Y > q_X\}$ by an $\alpha < 1$, and is multiplied on the set $\{(q_X, q_Y) : q_Y < q_X\}$ by the β such that $\alpha\mu_Y + \beta\mu_X = 1$. Since f is symmetric $\mu_Y = \mu_X$, so $\beta = 2 - \alpha$. Let $F(y \mid x)$ be the cdf that corresponds to density f conditional on x. Then, the average number of citations per paper of quality q is given by

$$q \cdot \frac{M_Y \alpha F\left(\frac{M_Y}{M_X} q \mid q\right) + M_X (2-\alpha) F(q \mid q) + M_X \alpha \left[F\left(\frac{M_X}{M_Y} q \mid q\right) - F(q \mid q)\right]}{\alpha F\left(\frac{M_Y}{M_X} q \mid q\right) + (2-\alpha) F(q \mid q) + \alpha \left[F\left(\frac{M_X}{M_Y} q \mid q\right) - F(q \mid q)\right]}.$$

In this case, the following condition will guarantee that the strategic effects reduce by more the number of citations per paper for higher values of q.

Condition 1 *For any $\underline{c} < \overline{c}$, the ratio*

$$\frac{F(\underline{c} q \mid q)}{F(\overline{c} q \mid q)}$$

increases in q.

This condition is implied by the MLR assumption, and is satisfied by many commonly used distributions, including exponential distributions and normal distributions truncated to the positive quadrant.

Proposition 4 *Suppose that Condition 1 is satisfied. Then, the fraction of papers on topic X, among all papers of quality q decreases with q.*

Since a paper of quality q on topic X is cited more frequently than a paper of quality q on topic Y, this result implies that the strategic effects reduce the number of citations per paper more for higher values of q. Put together with the direct effect that the papers of quality q are cited by each subsequent researcher more frequently, the relation between the number of citations and quality of papers is ambiguous. For exponential distributions and normal distributions, the direct effect dominates, but it is not difficult to construct distributions for which higher quality papers have lower average numbers of citations.

Proof. The fraction of papers on topic X among all papers of quality q is given by

$$\frac{(2-\alpha)F(q\mid q)+\alpha\left[F\left(\frac{M_X}{M_Y}q\mid q\right)-F(q\mid q)\right]}{\alpha F\left(\frac{M_Y}{M_X}q\mid q\right)+(2-\alpha)F(q\mid q)+\alpha\left[F\left(\frac{M_X}{M_Y}q\mid q\right)-F(q\mid q)\right]},$$

which is equal to

$$\frac{(2-2\alpha)+\alpha\dfrac{F\left(\frac{M_X}{M_Y}q\mid q\right)}{F(q\mid q)}}{\alpha\dfrac{F\left(\frac{M_Y}{M_X}q\mid q\right)}{F(q\mid q)}+(2-2\alpha)+\alpha\dfrac{F\left(\frac{M_X}{M_Y}q\mid q\right)}{F(q\mid q)}}.$$

Denote $F\left(\frac{M_X}{M_Y}q\mid q\right)/F(q\mid q)$ by $\phi(q)$, and by $F\left(\frac{M_Y}{M_X}q\mid q\right)/F(q\mid q)$ by $\psi(q)$. Since $M_X>M_Y$, by virtue of Condition 1, ϕ decreases in q, and ψ increases in q. Rewriting the formula for fraction of papers on topic X among all papers of quality q, we obtain

$$\frac{(2-2\alpha)+\alpha\phi(q)}{\alpha\psi(q)+(2-2\alpha)+\alpha\phi(q)}.$$

The derivative of this function is

$$\frac{\alpha^2\phi'(q)-\alpha(\beta-\alpha)\psi'(q)-\alpha^2\phi(q)\psi'(q)}{[\alpha\psi(q)+(\beta-\alpha)+\alpha\phi(q)]^2}<0.$$

■

6 Dynamic analysis

Up to now, we have been assuming that the distribution of types is constant over time. We will relax this assumption in the present section. Of course, there is a huge variety of ways in which the distribution can change over time. We would not be able to derive any general results. However, some important insight can be obtained by studying dynamic evolution of asymmetry across topics without any changes in the relative density of types which advantage in X, or types which have advantage in Y. We will assume that the distributions have constant fractions of partizan and strategic researchers, but the density f over types (q_X,q_Y) changes over time. More specifically, some constant over time density f is multiplied at time t by an α^t on the set $\{(q_X,q_Y):q_Y>q_X\}$; and on the set $\{(q_X,q_Y):q_Y<q_X\}$ is multiplied by β^t such that $\alpha^t\mu_Y+\beta^t\mu_X=1$.

6.1 Anticipated inflow of ideas on a topic

We will first study the dynamics of research in response to the news that an inflow of ideas is going to occur at some future date. For example, one may think about the invention of technology generating big data sets. In this case, we may yet not know in which way these data sets can be used, but anticipate that sooner or later, researchers will find ways of using these data sets to answer important questions.

Some insight will be obtained by studying the process such that $\mu_Y = \mu_X = 1/2$ and $\alpha^t = 1$ in all period except period $t = T > 1$, in which $\alpha^T = \alpha < 1$ (and so, $\beta^T = \beta = 2 - \alpha > 1$). Similar insight would be obtained by studying more general process such that α^t gradually, geometrically increases till time T, and gradually, geometrically decreases from time T on.[3]

Since researchers are forward-looking, the decisions of researchers living in periods $t = T, T+1, ...$ are fully efficient. They choose the topic whose q is higher. So, $M_X^T = \beta$, $M_Y^T = \alpha$, and $M_X^k = M_Y^k = 1/2$ for $k = T+1, T+2, ...$; $\overline{M}_X^T = (1-\delta)\beta(1/2) + \delta(1/2)$, $\overline{M}_Y^T = (1-\delta)\alpha(1/2) + \delta(1/2)$, and $\overline{M}_X^k = \overline{M}_Y^k = 1/2$ for $k = T+1, T+2, ...$ For researchers living in periods $1, ..., T-1$,

$$M_X^k = p \Pr\left\{(q_X, q_Y) : q_X \overline{M}_X^{k+1} > q_Y \left(1 - \overline{M}_X^{k+1}\right)\right\} + (1-p)(1/2),$$

$$M_Y^k = p \Pr\left\{(q_X, q_Y) : q_Y \overline{M}_Y^{k+1} > q_X \left(1 - \overline{M}_Y^{k+1}\right)\right\} + (1-p)(1/2);$$

and

$$\overline{M}_X^k = (1-\delta)M_X^k + \delta \overline{M}_X^{k+1},$$

$$\overline{M}_Y^k = (1-\delta)M_Y^k + \delta \overline{M}_Y^{k+1}.$$

Thus, M_X^k, M_Y^k, $\overline{M}_X^k$, and $\overline{M}_Y^k$ for $k = 1, ..., T-1$ are determined recursively. An argument analogous to that from the proof of Proposition 1 shows the uniqueness of equilibrium.

It can be easily proved by recursion that $M_X^k > 1/2$ (and $M_Y^k < 1/2$) for $k = 1, ..., T-1$. In an interpretation, the inflow of new ideas on topic X takes place in period T, but strategic researchers begin switching to topic X from period 1, that is from the time this inflow is anticipated. The next proposition shows that under some condition, this strategic effect is so strong that we observe over time a declining flow of papers on topic X.

Proposition 5 *Suppose that*

$$(1-\delta)p \Pr\{(q_X, q_Y) : [(1-\delta)\alpha + \delta]/[(1-\delta)\beta + \delta] < q_X < q_Y\} > \beta - 1/2.$$

Then, $M_X^k > M_X^{k+1}$ *for all* $k = 1, 2, ...$

[3] We conjecture, but have not proved formally that similar insight would be obtained also for stochastic processes such that α^T decreases to α randomly, and then increases to 1, also randomly.

Proof. The probability of writing on topic X in period T exceeds 1/2 by $\beta - 1/2$. In period $T-1$, the probability of writing on topic X is

$$p\Pr\{(q_X, q_Y) : q_X[(1-\delta)\beta + \delta] > q_Y[(1-\delta)\alpha + \delta]\} + (1-p)(1/2),$$

because $\overline{M}_X^T = (1-\delta)\beta(1/2) + \delta(1/2)$ and $\overline{M}_Y^T = (1-\delta)\alpha(1/2) + \delta(1/2)$.

This number exceeds 1/2 by

$$p[\Pr\{(q_X, q_Y) : q_X[(1-\delta)\beta + \delta] > q_Y[(1-\delta)\alpha + \delta]\} - 1/2].$$

Now, the result for $k = T-1$ follows from $1/2 = \Pr\{(q_X, q_Y) : q_X > q_Y\}$. By induction, the result extends to $k = 1, ..., T-2$. ■

So, it may happen that the number of papers on topic X in period 1, which may well precede the inflow of ideas on this topic, actually exceeds the number of papers on this topic when the inflow actually occurs. This is more likely to happen in fields in which talent is uniform across topics, because more mass of types (q_X, q_Y) is concentrated around the diagonal, and consequently in the region $\{(q_X, q_Y) : [(1-\delta)\alpha + \delta]/[(1-\delta)\beta + \delta] < q_X < q_Y\}$.

Finally, we modelled an inflow of ideas as an uniform increase of density across types (q_X, q_Y) such that $q_X > q_Y$. However, one can imagine inflows that consist of very good ideas on some topic, that is, the ideas with high q_X's. In such a case, the number of papers in periods preceding the inflow of new ideas is higher, but their average quality falls below the average quality of papers at the time the inflow actually occurs.

6.2 Counting citations in different periods with different weights

Many people believe that citations are a better signal of quality (social value) in the case of older that newer papers. There are several reasons behind this view. For example, immediate citations may be a result of current fashion, or the authors' position in the profession, and may be less correlated with quality. In addition, social value is better assessed only after some time. Our model is in fact based on this view.

In this section, we address the question whether it would be more efficient to assign lower weights to citations in papers that appear shortly after a given paper, and higher weights to citations in papers that appear long time after the given paper.

To make the analysis simple, we will consider only symmetric Markov processes, and will restrict attention to symmetric equilibria.[4] More specifically, let f be a full-support density over types (q_X, q_Y) symmetric across the diagonal. That is, $f(q_X, q_Y) = f(q_Y, q_X)$, in particular, $\mu_X = \mu_Y = 1/2$. For $t = 1, 2, ...$, let $\alpha^t = \alpha < 1$ or $2 - \alpha > 1$; note that $\alpha(1/2) + (2-\alpha)(1/2) = 1$. If $\alpha^t = \alpha$, then $\alpha^{t+1} = \alpha$ with probability $\theta > 1/2$ and $\alpha^{t+1} = 2 - \alpha$ with the complementary probability of $1 - \theta$. Symmetrically, if $\alpha^t = 2 - \alpha$, then $\alpha^{t+1} = 2 - \alpha$ with probability θ and $\alpha^{t+1} = \alpha$ with probability $1 - \theta$. The density f^t in period t on the set $\{(q_X, q_Y) : q_Y > q_X\}$ is equal to f multiplied by an α^t, and on the set $\{(q_X, q_Y) : q_Y < q_X\}$ is equal to

[4] These assumptions are inessential for our results, but greatly simplify the analysis.

f multiplied by $2-\alpha^t$. That is, if $\alpha^t = \alpha$, there are more researchers with advantage in topic X, and if $\alpha^t = 2-\alpha$, there are more researchers with advantage in topic Y.

We must first generalize some concepts and results of our basic model to the present dynamic setting. Similarly to Section 3, define $M_i^{t,j}$ be the fraction of researchers who choose topic i in period t when there are more researchers in period t with advantage in topic j. Unlike in Section 3, this fraction as the strategies may depend on the state j of the Markov process. These fractions are well defined for any history-independent strategies. Given any history-independent strategies of researchers living in periods $t, t+1, \ldots$, let

$$\overline{M}_i^{t,X} = (1-\delta)\sum_{n=t}^{\infty}\delta^{n-t}E_t M_i^{n,j_n},$$

where E_t denotes the expected value taken at time t of M_i^{n,j_n}, which depends on the state j_n of the Markov process in period n. That is, $\overline{M}_i^{t,X}$ is the occupation measure of topic i from time t on when in period t there are more researchers with advantage in topic X. One can analogously define $\overline{M}_i^{t,Y}$. Obviously, we have that $M_X^{t,j} + M_Y^{t,j} = 1$ and $\overline{M}_X^{t,j} + \overline{M}_Y^{t,j} = 1$ for $j = X, Y$.

Strategies are symmetric across topics, and will be simply called *symmetric* if whenever a researcher of type (q_X, q_Y) chooses topic X (topic Y) in state$j = X$, then a researcher of type (q_Y, q_X) chooses topic Y (topic X) in state $j = Y$. In this case, we also have that $M_X^{t,X} = M_Y^{t,Y}$, $M_Y^{t,X} = M_X^{t,Y}$, $\overline{M}_X^{t,X} = \overline{M}_Y^{t,Y}$, and $\overline{M}_Y^{t,X} = \overline{M}_Y^{t,X}$. For any stationary strategies, $M_i^{t,j}$ and $\overline{M}_i^{t,j}$ are independent of t. If strategies are symmetric and stationary, we denote $\overline{M}_X^{t,X} = \overline{M}_Y^{t,Y}$ by M^+ and $\overline{M}_Y^{t,X} = \overline{M}_X^{t,Y}$ by M^-. Of course, $M^+ + M^- = 1$.

Finally, let

$$\begin{aligned} g^+(M^+) &= (1-\delta)[p\Pr\{(q_X,q_Y) : q_X[\theta M^+ + (1-\theta)(1-M^+)] > q_Y[\theta(1-M^+) + (1-\theta)M^+]\} \\ &\quad +(1-p)(1-\alpha/2)] + \delta[\theta M^+ + (1-\theta)(1-M^+)] \end{aligned}$$

and

$$\begin{aligned} g^-(M^-) &= (1-\delta)[p\Pr\{(q_X,q_Y) : q_X[\theta(1-M^-) + (1-\theta)M^-] < q_Y[\theta M^- + (1-\theta)(1-M^-)]\} \\ &\quad +(1-p)(\alpha/2)] + \delta[(1-\theta)(1-M^-) + \theta M^-], \end{aligned}$$

where operator Pr refers to the distribution with density f multiplied by α on the set $\{(q_X, q_Y) : q_Y > q_X\}$, and multiplied by $2-\alpha$ on the set $\{(q_X, q_Y) : q_Y < q_X\}$.

Similarly to Section 3, we assume the following condition:

Assumption 2. Function g^+ has a unique fixed point.

If $M^+ + M^- = 1$, then $g^+(M^+) + g^-(M^-) = 1$. So, it would be equivalent to assume that function g^- has a unique fixed point. Notice that both g^+ and g^- always have a fixed point M_i, since each of them is continuous, exceeds 0 at 0 and falls below 1 at 1. The following proposition characterizes the symmetric equilibria under Assumption 2. Its proof is analogous to the proof of Proposition 1.

Proposition 6 *The model has a unique symmetric equilibrium. This equilibrium is stationary. In the equilibrium, M^+ is the fixed point of g^+. A researcher living in a period in which the state of the Markov process is X (is Y) chooses topic X when $M^+ q_X > M^+ q_Y$ (when $M^- q_X > M^- q_Y$), and chooses topic Y when the opposite inequality holds.*

We can now address our question. A way of modeling higher weight assigned to citations more distant in time is by an increase in δ.[5]

Proposition 7 *The unique fixed point M^+ of g^+ decreases when δ increases. This means that the unique symmetric equilibrium becomes more efficient.*

Proof. Rewrite equation $M^+ = g^+(M^+)$ as

$$\frac{[1 - \delta\theta + \delta(1-\theta)]M^+ - \delta(1-\theta)}{1-\delta} = p \Pr + (1-p)(1-\alpha/2),$$

where $\Pr := \Pr\{(q_X, q_Y) : q_X[\theta M^+ + (1-\theta)(1-M^+)] > q_Y[\theta(1-M^+) + (1-\theta)M^+]$. Notice that the right-hand side (RHS) of this equation is independent of M^+, while the left-hand side (LHS) is increasing in M^+. So, the graph of the LHS intersects the graph of the constant RHS at the unique fixed point M^+ from below to above.

Notice further that the RHS of this equation is also independent of δ, while the derivative of the LHS is positive in δ for all M^+. This means that the point M^+ at which the graph of the LHS intersect the graph of the RHS decreases when δ increases. ■

Proposition 7 captures the intuition that by assigning a higher weight to citations more distant in time we remove the incentives of following current trends or fashion, or simply choosing the topic that other do. Instead, we provide incentives for choosing what is likely to have more value, and this value can be known only in a longer time span.

One may also be tempted to address the more general question of characterizing the optimal intertemporal weight ratios. We find this question interesting and important, but obtaining any reasonable insight requires assuming a more concrete payoff function. We view (1) as a "first-order" approximation of researchers payoffs. In practice, researchers are not directly interested in maximizing citations of their papers, but in other goals which can be achieved by means of a higher number of citations. In particular, one should rather assume a limit on the value of a single citation, and that the marginal payoff of a citation diminishes with the number of previous citations.

[5] More precisely, let r be a (constant over time) ratio of the weight of a citation in n period after a given paper has appeared to the weight of a citation in $n+1$ periods after. Further, let w be the weight of immediate citations. Then, the payoff of a strategic researcher is

$$(1-\delta)\sum_{n=1}^{\infty} w\delta^n r^n \tau_n = \frac{w(1-\delta)}{(1-\delta')}(1-\delta')\sum_{n=1}^{\infty}(\delta')^n \tau_n,$$

where $\delta' = \delta r$. Normalizing this payoff by $w(1-\delta)/(1-\delta')$, which bears no loss of generality as only relative payoffs matter in the present analysis, we obtain that raising the ratio r is equivalent to raising the discount factor δ'.

We conjecture that in such a model there would exist an optimal moment from which we should count citations. The reason is that delaying this moment reduces inefficient choice of topics, but also reduces the incentives of conducting any research, at least in a model in which researchers face some cost of conducting research and have the option of not conduction any research at all.

6.3 Should quotations in higher-ranked journals have higher weight?

It is in the background of our analysis that the significance of ideas is difficult to determine at the time these ideas appear. This significance can be correctly evaluated only in a longer period of time. But at least some ideas are perceived or expected to be more significant than others already at the time they appear. A majority of scientists would probably agree that ideas can be partially, and not perfectly stratified quite early. For example, good peer reviews, possibly resulting in publications in historically higher-ranked journals, can be viewed as a measure of significance.

In addition, we assumed that only a researcher can evaluate the expected value q of the idea that she is working on. We believe that this assumption is right, but when a paper containing the idea already appears we obtain a signal regarding its expected value. In the present section, we will assume that this signal reveals the researcher's private information, that is, the expected value q of the idea becomes common knowledge. For example, it is perfectly revealed by peer reviews, and the ranking of journals in which the paper appears.

We are interested if the society can benefit, that is, enhance efficiency by increasing the weight assigned to citations in papers with higher q. In an interpretation, the question is whether by increasing the weight of citations in papers published in journals of higher reputation will enhance efficiency. Or would it be more efficient to refer to Web of Science instead of Google Scholar, as the former source seems more selective than the latter in counting citations.

Before presenting a formal analysis, we wish to mention two other methods of enhancing efficiency. First, it seems that if researchers were more selective in choosing citations, this would give other researchers incentives to write papers with higher q. Second, one might say that our analysis actually calls for relying on publication records instead of on the numbers of citations. Indeed, if strategic researchers were rewarded by the quality of journals in which their papers are published, they would make the same decisions as partizan researchers, and the outcome would be efficient.

We would, however, be more cautious with making this second conclusion. The quality of journals seems a better measure at the aggregate level than at the individual level. More precisely, the possibility of publishing in a specific journal in certain periods may strongly reflect the taste and views of the current editors, which makes it easier to publish certain kind of papers even when their social value is rather low (but this is revealed only much later). Over time, those tastes and views may well average out, which makes relying on citations with higher weights assigned to higher-ranked journals more effective to relying on publication records. In addition, the decisions regarding publication rely on the opinions of few experts. So, they seem to be weaker signals of quality, compared to citations over the entire future. This may, for example, affect the incentives of investing in valuable projects.

Finally, it should be mentioned that we disregard in our analysis the fact that journals are typically keen to publish papers with a higher expected number of citations, and their ranking depends on the citations of the papers that they publish. This fact, however, seem not to affect the analysis, since we are not differentiating between the social value and the expected number of citations.

Turning to our formal analysis, consider a special case of our model in which only one researcher lives in each period, and q stands for the probability of being cited, which coincides with the quality of journal in which the paper is published. Suppose that citations are weighted by the quality of papers in which cite them. More specifically, let $w : [0,1] \to [0,1]$ be a strictly increasing function, and redefine the payoff function (1) by assuming that a citation in a paper of quality q contributes $w(q)$ to the payoff. Up to now $w(q)$ was equal to 1 for all q. That is, the payoff is now given by

$$(1-\delta)\sum_{n=1}^{\infty}\delta^n \chi w(q_n),$$

where q_n is the quality of the paper written in period n; $\chi = 1$ if the paper written in period n is on the same topic as a given paper, and $\chi = 0$ if the paper written in period n is on the other topic.

Proposition 1 easily generalizes to the new payoff, but now

$$M_Y = \int_0^1 \left(\int_0^y w(y) f(x,y) dx \right) dy - p \int_0^1 \left(\int_0^{yM_Y/M_X} w(y) f(x,y) dx \right) dy \tag{2}$$

and

$$M_X = \int_0^1 \left(\int_0^x w(x) f(x,y) dy \right) dx + p \int_0^1 \left(\int_0^{yM_Y/M_X} w(x) f(x,y) dx \right) dy. \tag{3}$$

In the two formulas, the first term represents the expected payoff when all researchers with advantage in topic i write on topic i; this payoff is decreased for $i = Y$ and increased for $i = X$ by the flow of researchers with advantage in X who choose to write on topic X, which is represented by the second term in both formulas.

The total effect of weighting citations can be decomposed into three effects: The first effect reflects differences in the distribution of q's across topics. If, for example, there are many types with high q_Y but only few types with high q_X, then weighting citations naturally reduces inefficiency, because it increases the payoff to writing on topic Y. It may, in some extreme cases, even revert the direction of the flow of researchers, resulting in researchers with advantage in topic X writing on topic Y. We will not attempt to quantify this effect; instead we will assume it away, in order to describe two remaining effects.

More specifically, let g be a density on $[0,1]^2$ that is symmetric across the diagonal, that is, $g(x,y) = g(y,x)$, and let $f(x,y) = \alpha g(x,y)$ for $y > x$ and $f(x,y) = \beta g(x,y)$ for $y < x$, where $\alpha(1/2) + \beta(1/2) = 1$. In addition, normalize function w so that the value of the first terms in (2), and in (3) by symmetry, remain the same when w is replaced with 1. Finally, consider an auxiliary scenario, under which citations are weighted by w, but the papers on topic X of the researchers with advantage in Y had the same quality as their papers on topic Y. Or, in other words, the citations in their papers on both topics had the same weight.

Let M_X^o/M_Y^o be the equilibrium ratio of the probability of writing on X versus the probability of writing on Y when all citations are counted equally, let M_X^w/M_Y^w be the equilibrium ratio when citations are counted with weight w, and let M_X^a/M_Y^a be the equilibrium ratio under the auxiliary scenario. (The superscripts refer to 'original,' 'weighted,' and 'auxiliary'.) We decompose the change from M_X^o/M_Y^o to M_X^w/M_Y^w into the change to M_X^a/M_Y^a and then the change from M_X^a/M_Y^a to M_X^w/M_Y^w.

The effect of the former change depends on the quality of types with advantage in Y who write on X (under the original scenario with equal weights) compared to the quality of all types with advantage in Y. More precisely, if the distribution of q_Y contingent on types with advantage in Y who write on X first-order stochastically dominates the distribution of q_Y contingent on types with advantage in Y, then $M_X^a/M_Y^a \geq M_X^o/M_Y^o$.

This of course means that the auxiliary scenario is less efficient than the original scenario. And the other way round, if the distribution of q_Y contingent on types with advantage in Y first-order stochastically dominates the distribution of q_Y contingent on types with advantage in Y who write on X, then $M_X^a/M_Y^a \leq M_X^o/M_Y^o$, which means that the auxiliary scenario is more efficient than the original scenario.

The effect of the change from the auxiliary scenario to the weighted scenario is always negative in terms of efficiency (i.e., $M_X^a/M_Y^a \leq M_X^w/M_Y^w$), because the quality of papers of researchers with advantage in Y who write on X is lower compared to what it would be if they were writing on Y.

The intuition for the two effects can be explained as follows: In our model, more researchers have advantage in X than in Y, by assumption. This provides incentives for some researchers with advantage in Y for writing on X, namely, the researchers whose advantage in Y over X is not too high. This in turn magnifies the incentives for writing on X for researchers with even higher advantage in Y over X. This process must stop, however, since the researchers with sufficiently high advantage in Y will never work on X. This is a consequence of the presence of partizan researchers.

The incentives are magnified more (or less) for a strictly increasing function w than for a constant w if the quality of researchers with advantage on Y but writing on X is high (respectively, low) compared to the rest of the population. This is the first of the two described effects. In addition, the incentives are magnified less with a strictly increasing function w, because switching topics burns some value in terms of citations compared to a constant w. This is the second of the two described effects.

So, we conclude that putting a higher weight on citations in higher-quality papers enhances efficiency in disciplines in which researchers of relatively low quality are "shopping" for topics, because both effects work in the same direction. In turn, if researchers of relatively high quality are shopping for topics, then the two effects work in the opposite direction, and the total effect of putting a higher weight on citations in higher-quality papers is ambiguous.

7 Extensions

In the present paper, we opted for a simple model with only two topics, which captures some basic effects. In particular, we restricted attention to two topics. In the version of the model with more than two topics, stationary equilibria seem to induce a variety of patterns, and nonstationary equilibria seem to exhibit nontrivial dynamics. The model has also numerous possible extensions.

In our opinion, an important extension would have new topics arriving over time, possibly with other topics becoming obsolete. Numerous intriguing questions could be addressed in such an extension. For example, what types or researchers would me most keen on inventing new ideas which could not be classified to one of the existing categories? Or, what kind of dynamics would emerge; or would innovations be inefficient, and why? Designing such a setting requires a theory of topic arrivals, taking a position on their social value, and specifying the way in which the distribution of researchers' types evolves with topic arrivals. This is left for future papers.

Our theory treats papers are fairly independent units. It seems more realistic that new ideas build on the existing ideas. This is, for example, a standard approach in the recent literature on innovation. This is another promising direction of future research. The clustering that we have observed in our basic model would also come from ideas generating other ideas. Of course, this other kind of clustering seems rather a positive phenomenon, as it means studying topics where the ideas are. The negative strategic clustering would, however, not disappear, to the opposite it would be rather magnified.[6]

One may be interested in a stock-dependent distribution of researchers' types, i.e., a distribution that depends on the number of previously executed ideas on each topic. The social values of ideas may depend on these numbers too. For example, one may argue that the distribution of researchers' types and the social values of ideas are bell-shaped, with early research conducted by few pioneers having by itself little value, but inspiring subsequent research with more substantial value added, until the ideas become exhausted and their value marginal.

Finally, the execution of more socially valuable projects takes more time, which may suggest that it would be socially desirable to pay attention only to the most highly cited paper of an individual researcher.

8 References

Diamond, P. A. and J. E. Stiglitz (1974): "Increases in Risk and in Risk Aversion," *Journal of Economic Theory*, **8**, 337-360.

Ellison, G. (2010): "How Does the Market Use Citation Data? The Hirsch Index in Economics," National Bureau of Economic Research Working Paper 16419.

[6] One can also think of arguments that are more remote from our basic model. For example, it may be more interesting, and providing better incentives when a larger group of researchers is exploring related or similar problems. Or the other way round, it may be creating incentives for studying such problems inefficiently long, and only making marginal contributions.

Ellison G. (2012): "Assessing Computer Scientists Using Citation Data," Massachussetts Institute of Technology and National Bureau of Economic Research, mimeo.

Hirsch, J. E. (2005): "An Index to Quantify an Individual's Scientific Research Output." *Proceedings of the National Academy of Sciences* **102**(46), 16569-16572.

Klabjan, D., W. Olszewski, and A. Wolinsky (2014): "Attributes," *Games and Economic Behavior* **88**, 190-206.

M. Perry and P. J. Reny (2016): "How to Count Citations If You Must." *American Economic Review* **106**(9), 2722–2741.

9 Appendix

The following proposition characterizes the equilibria of our basic model in the general case, without Assumption 1.

Proposition 8 *In any equilibrium,*

$$\overline{M}_i^t = (1-\delta) g_i(\overline{M}_i^{t+1}) + \delta \overline{M}_i^{t+1}, \tag{4}$$

the sequence $(\overline{M}_i^t)_{t=1}^{\infty}$ monotonically increases or decreases, and converges to a fixed point M_i of function g_i. Conversely, any sequence $(\overline{M}_i^t)_{t=1}^{\infty}$ with these properties determines an equilibrium: A researcher living in period t chooses topic X when $\overline{M}_X^{t+1} q_X > \overline{M}_Y^{t+1} q_Y$, and chooses topic Y when $\overline{M}_Y^{t+1} q_Y > \overline{M}_X^{t+1} q_X$; when $\overline{M}_X^{t+1} q_X = \overline{M}_Y^{t+1} q_Y$ (which event has probability 0), a researcher living in period t is indifferent between the two topics.

An example of nonconstant sequence $(\overline{M}_i^t)_{t=1}^{\infty}$, which determines a nonstationary equilibrium is depicted Figure 5. For any g_i whose graph crosses the diagonal from below to above, one can similarly construct a continuum of nonstationary equilibria. In any nonstationary equilibrium, sequence $(\overline{M}_i^t)_{t=1}^{\infty}$ converges to a fixed point M_i of function g_i at which the graph of this function crosses the diagonal from below to above.

The proof of this proposition also includes the proof of Proposition 1.

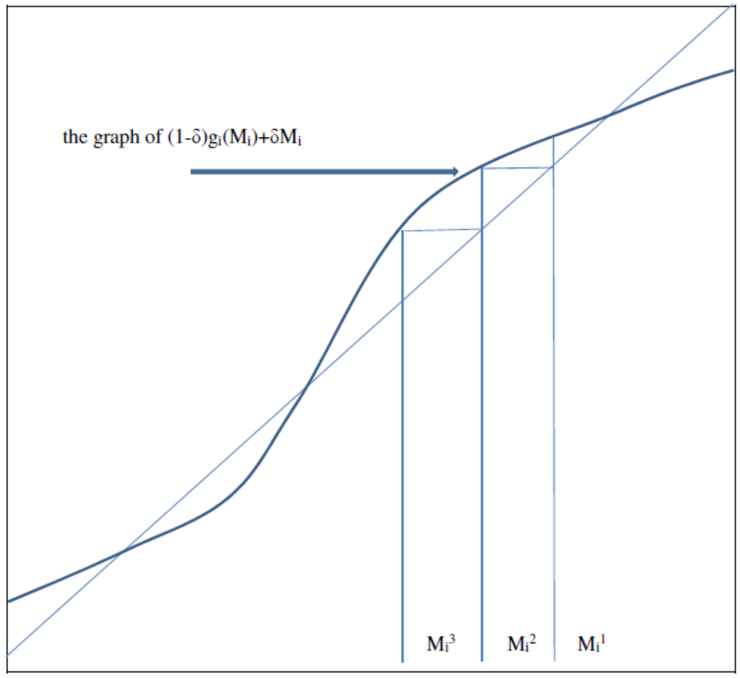

Figure 5. A construction of nonconstant sequence $(\overline{M}_i^t)_{t=1}^{\infty}$ satisfying the conditions of Proposition 8.

Proof: Notice that the payoff of a strategic researcher with type (q_X, q_Y) from writing on topic i is $\delta q_i \overline{M}_i^t N$, where N is the expected number of researchers per period. Therefore, the fraction of researchers who choose topic $i = X, Y$ in period t is given by $M_i^t = g_i(\overline{M}_i^{t+1})$. Thus, by definition, $\overline{M}_i^t$ is given by (4).

It follows immediately from definitions that: if $\overline{M}_i^{t+1} > \overline{M}_i^{t+2}$, then $M_i^t > M_i^{t+1}$ and $\overline{M}_i^t > \overline{M}_i^{t+1}$; and if $\overline{M}_i^{t+1} < \overline{M}_i^{t+2}$, then $M_i^t < M_i^{t+1}$ and $\overline{M}_i^t < \overline{M}_i^{t+1}$. Therefore, sequence $(\overline{M}_i^t)_{t=1}^{\infty}$ must be increasing or decreasing. In any case, the sequence must converge to a number between 0 and 1. By definition, so does sequence $(M_i^t)_{t=1}^{\infty}$, and the limit of these sequences must be a fixed point M_i of g_i.

This completes the proof of the first statement in Propositions 8. The converse is straightforward.

In a stationary equilibrium, $\overline{M}_i^t$ and $\overline{M}_i^{t+1}$ are equal to M_i. So, equation (4) reduces to $M_i = g_i(M_i)$. This, by Assumption 1, implies the uniqueness of stationary equilibrium. To complete the proof of Proposition 1, we must show that under Assumption 1 the model has no equilibrium other than the stationary equilibrium determined by the fixed points M_X and M_Y.

Suppose that $\overline{M}_i^{t+1}$ for some t is higher than the unique fixed point M_i. Since the graph of g_i intersects the diagonal only once from the left to the right, it follows from (4) that $\overline{M}_i^t < \overline{M}_i^{t+1}$, and so sequence $(\overline{M}_i^t)_{t=1}^{\infty}$ increases, and cannot converge to M_i. Similarly, if $\overline{M}_i^{t+1}$ for some t falls below the unique fixed point M_i, then it must be that $\overline{M}_i^t > \overline{M}_i^{t+1}$, and so sequence $(\overline{M}_i^t)_{t=1}^{\infty}$ decreases, and cannot converge to M_i. Thus, $\overline{M}_i^{t+1} = M_i$ for all t, and so does $\overline{M}_i^1$. As a result, the equilibrium coincides with the unique stationary equilibrium.