Graz Economics Papers – GEP

GEP 2017–11

The distribution of article quality and inefficiencies in the market for scientific journals

Philipp Kohlgruber and Christoph Kuzmics

October 2017

Department of Economics
Department of Public Economics
University of Graz

The distribution of article quality and inefficiencies in the market for scientific journals*

Philipp Kohlgruber† Christoph Kuzmics‡

October 3, 2017

Abstract

We build an oligopoly model of the market of scientific journals that allows us to relate the (in-)efficiency of this market to the exogenous distribution of article quality. Journal quality is endogenously determined by the submission choices of scientists. The efficiency of any stable equilibrium depends crucially on the exogenous distribution of article quality, especially on the fatness of the upper tail. For the empirically plausible Pareto distribution the market is inefficient even in the limit as the number of publishers tends to infinity.

Keywords: oligopoly, natural monopoly, efficiency, price competition, endogenous product differentiation

JEL classification numbers: C72, C73, D43, L13, L15, L82

1 Introduction

We build an oligopoly model of the market of scientific journals that allows us to relate the (in-)efficiency of this market to the exogenous distribution of article quality in the presence of many journals and many publishers.[1][2] As Bergstrom (2001) pointed out journal quality is the endogenous result of a coordination problem between writers of academic articles and their readers. Scientists would like to publish in journals that are widely read (which automatically have a high reputation) and libraries would like to acquire journals that

*We would like to thank Michael Greinecker for helpful comments and suggestions.

†University of Graz, Austria, philipp.kohlgruber@edu.uni-graz.at

‡University of Graz, Austria, christoph.kuzmics@uni-graz.at

[1]The model is based on a model due to Shaked and Sutton (1982) and specifically adapted to the market for scientific journals. Other specific theoretical models for the scientific publishing industry have been built by McCabe and Snyder (2005), McCabe, Snyder, et al. (2007), Jeon and Rochet (2010), McCabe and Snyder (2015), McCabe and Snyder (2016). These papers mostly focus on the trade-offs between offering open access while charging authors submission fees and free submission while charging readers.

[2]There is a large number of empirical studies into the market for scientific publishing, who all indicate potential inefficiencies in this market. See e.g. Bergstrom (2001), Tenopir and King (2000), and Larivière, Haustein, and Mongeon (2015).

publish the best articles. This is a key feature of our model. Another key feature is the assumption of heterogenous article quality. We are interested in how the publishers' journal pricing decision depends on the exogenously given distribution of article quality and how this affects overall welfare in this market.

We first provide a technical result that provides necessary conditions for an equilibrium in such a market. While we cannot provide a complete closed form solution we can provide bounds on prices that suffice to, in turn, provide bounds on welfare.

We then show that a market with this key feature may or may not be inefficient. Whether or not it is inefficient depends on the exogenously given distribution of article quality. For instance, if article quality is drawn from a uniform distribution on a given interval, then as the number of publishers tends to infinity, all journal prices tend to zero (the assumed marginal cost of making the journal accessible to one more library) and consumer welfare tends to the first best. If article quality follows an exponential distribution, journal prices of top journals do not tend to zero as the number of journal and publishers tends to infinity. In this limit the market also exhibits absolute welfare losses compared to the first best. The ratio of welfare losses to first best welfare tends to zero, however, also in this case. Note, however, that the presence of non-zero prices in the limit does already indicate a form of inefficiency in that the money universities then spend on journal access could instead have been spent on hiring more researchers. We do not model this aspect directly. Finally, if article quality follows an appropriate Pareto distribution, then not only are journal prices significantly positive in the limit as journal and publisher numbers tend to infinity, but even the relative welfare loss is positive in this limit. In this limit many libraries are not able to acquire the top quality journals *and* the journals that these libraries do acquire are of significantly lower quality. For a sample of economics journal articles, the latter case seems to be the most appropriate. Thus, our finding reinforces the view, held for instance by Bergstrom (2001), that the market for scientific journals is indeed inefficient and should be regulated.

The above statements are true for all equilibria that are stable (in an evolutionary sense). The market always also has a highly efficient but unstable equilibrium, in which all journals are of equal quality, and are equally frequently read and purchased, and all scientists submit their articles randomly to all journals. This is unstable, as a slight change in the readership of different journals will, because of the coordination motive, lead to increased submissions to the more read journals, which in turn will lead to libraries wanting to acquire these journals more, which will in turn lead to more submissions to these journals, and so on, until we reach a coordinated and stable equilibrium in which journals are clearly ranked in terms of quality.

The remainder of the paper is organized as follows. Section 2 provides a formal account of the model. Section 3 provides the main results and Section 4 concludes with a discussion of the implications of our findings. Appendix A discusses provides an empirical analysis of the distribution of article quality. The proof of the main technical result is provided in a series of lemmas in Appendix B.

2 Model

There are four kinds of agents in our model, three of them are strategic. In what follows we state who they are, what their strategies are, and what their payoffs are.

2.1 Agents

There is a continuum of *authors*. Every author writes one paper. An author is characterized by the exogenously given uni-dimensional quality of her paper. The quality of papers across authors is distributed according to some cumulative distribution function F with density f, defined on the set of positive real numbers.

There are n (scientific) journals. Every journal has an editor, who can also be thought to obtain help from referees. This editor is non-strategic and does a perfect job of identifying the quality of any paper submitted to her journal. If the proportion of submitted papers to this journal is less than or equal to $1/n$ of all papers in the market she accepts all of them. If this proportion is higher she uses a minimum threshold level of quality in such a way that she accepts exactly a fraction $1/n$ of all papers in the market. Let H_i denote the cumulative distribution function of the quality of the papers in journal i. Any journal i's quality, denoted $z_i = \zeta(H_i)$, is then given by some aggregate statistics of the quality of the papers it publishes. This function ζ is such that if H_i first-order stochastically dominates H_j then $\zeta(H_i) > \zeta(H_j)$. As prime examples we consider ζ to identify the mean or median of the distribution of paper quality.

There are n *journal publishers* and every journal publisher owns exactly one journal.

There is a continuum of *libraries*, the sole purchasers of journals. Libraries are characterized by a parameter θ, their type, that captures their idiosyncratic trade-off between the quality of a journal and its price. This parameter θ, across libraries, is distributed according to a cumulative distribution function G with density g on the set of positive real numbers.

2.2 Strategies

Authors know the quality of their paper and choose which journal to submit their paper to. For any journal i, the journal publisher of journal i has no influence on journal quality (beyond what her editor already does for her) and chooses a price p_i for her journal. Libraries observe the quality of all journals, buy at most one journal, and choose which one they buy.[3]

[3] One should think of there being a separate market for scientific journals for every subfield of science. Nevertheless, in reality libraries typically buy more than one journal in each subfield of science. The assumption that every library is buying at most one journal is made for simplicity. We conjecture that allowing libraries to purchase more than one journal would make the market even less competitive.

2.3 Payoffs

A library, characterized by parameter θ, receives a utility of $\theta z_i - p_i$ if it purchases journal i. Let $z = (z_1, z_2, ..., z_n)$ and $p = (p_1, p_2, ..., p_n)$ denote the vector of journal qualities and price, respectively. The libraries' choices, for a given pair (z, p) of journal quality and price vectors, determine the demand for any journal i, denoted by $D_i(z, p)$.

The journal publisher of any journal i has zero costs and receives profits of $p_i D_i(z, p)$.

Authors receive a payoff of $D_i(z, p)$ if the publish their paper at journal i.[4] They receive a payoff of -1 if their paper is not published. This implies that authors prefer to have a paper published at a journal with zero demand over not getting the paper published at all.

2.4 Equilibrium

The solution concept we employ is Nash equilibrium and an evolutionary refinement of Nash equilibrium. We call a pair (z, p) of journal qualities and prices an *equilibrium* if, given this pair, libraries are buying a journal that maximizes their payoff, inducing demand $D_i(z, p)$ for each journal i, journal publishers are maximizing profits, and authors submit their paper to the journal with the highest $D_i(z, p)$ among all journals that would accept their paper.

We call (z, p) a *stable equilibrium* if it is an equilibrium and if small changes to z, the vector of journal qualities, with resulting changes to $D_i(z, p)$ for all journals i, do not change the authors' optimal submission decision. Note that this is an evolutionary notion of stability, not a strategic notion.

3 Results

Every such game has a highly competitive but unstable equilibrium. This equilibrium is as follows. Every author submits her paper to journals by randomizing equally between all journals. Every journal has exactly the same quality. All journals have zero prices (even if there are only two). This is by the usual homogeneous good Bertrand price competition logic. Libraries choose to acquire a journal by randomizing equally among all journals. Thus, the equilibrium demand for all journals is the same. This equilibrium is, however, unstable in exactly the same way as the mixed equilibrium is unstable in the two-player two-action coordination game. To see this note that almost any small change in journal qualities will mean that journals have different quality levels; keeping prices fixed at zero, this will lead to zero demand for low quality journals and high demand for the top-quality journals; this, in turn, changes authors' optimal submission decision. We focus on stable equilibria from now on.

[4] Attema, Brower, and Exel (2014) demonstrate empirically that economists value publications in top journals highly and that they are willing to make substantial sacrifices for such publications.

3.1 Stable Equilibria

Let $\bar{z}_\alpha$ denote the α-quantile of the distribution of paper qualities according to cumulative distribution function F. Thus, $\bar{z}_0 = 0$ and $\bar{z}_1$ equal to the upper bound of the support of F, which could be infinity. We can assume, without loss of generality, that journal qualities are ordered by the journal number, i.e., that $z_1 \geq z_2 \geq ... \geq z_n$.

Lemma 1 *Any stable equilibrium (z, p) satisfies $z_1 > z_2 > ... > z_n > 0$, $p_1 > p_2 > ... > p_n > 0$ and has the following properties:*

1. *An author submits her paper to journal i if her paper quality $z \in \left(\bar{z}_{\frac{n-i}{n}}, \bar{z}_{\frac{n-i+1}{n}}\right)$ and only if $z \in \left[\bar{z}_{\frac{n-i}{n}}, \bar{z}_{\frac{n-i+1}{n}}\right]$.*
2. *The distribution of paper quality of journal i, H_i, is given by the restriction of F to the interval $\left(\bar{z}_{\frac{n-i}{n}}, \bar{z}_{\frac{n-i+1}{n}}\right)$, and journal i's quality is given by $z_i = \zeta(H_i)$.*
3. *The demand for journal 1 is given by*
$$D_1(z,p) = 1 - G\left(\frac{p_1 - p_2}{z_1 - z_2}\right).$$
4. *The demand for journal $i \in \{2, ..., n-1\}$ is given by*
$$D_i(z,p) = G\left(\frac{p_{i-1} - p_i}{z_{i-1} - z_i}\right) - G\left(\frac{p_i - p_{i+1}}{z_i - z_{i+1}}\right).$$
5. *The demand for journal n is given by*
$$D_n(z,p) = G\left(\frac{p_{n-1} - p_n}{z_{n-1} - z_n}\right) - G\left(\frac{p_n}{z_n}\right).$$
6. *Journal prices satisfy the following system of equations:*
$$\begin{aligned} 1 - G\left(\frac{p_1 - p_2}{z_1 - z_2}\right) - g\left(\frac{p_1 - p_2}{z_1 - z_2}\right)\frac{p_1}{z_1 - z_2} &= 0 \\ G\left(\frac{p_{i-1} - p_i}{z_{i-1} - z_i}\right) - G\left(\frac{p_i - p_{i+1}}{z_i - z_{i+1}}\right) - & \\ g\left(\frac{p_{i-1} - p_i}{z_{i-1} - z_i}\right)\frac{p_i}{z_{i-1} - z_i} - g\left(\frac{p_i - p_{i+1}}{z_i - z_{i+1}}\right)\frac{p_i}{z_i - z_{i+1}} &= 0 \\ G\left(\frac{p_{n-1} - p_n}{z_{n-1} - z_n}\right) - G\left(\frac{p_n}{z_n}\right) - & \\ g\left(\frac{p_{n-1} - p_n}{z_{n-1} - z_n}\right)\frac{p_n}{z_{n-1} - z_n} - g\left(\frac{p_n}{z_n}\right)\frac{p_n}{z_n} &= 0, \end{aligned}$$
where the second equation holds for all $i \in \{2, ..., n-1\}$.

The proof of Lemma 1 is given in the appendix. For the remainder of this paper we assume that G, the distribution over library types, is uniform on the unit interval. We also assume that journal quality is determined by the median paper quality of all the papers published in this journal. For this case we get the following corollary.

Corollary 1 *Let G be the uniform distribution on $[0,1]$. Let $\zeta(H)$, for any distribution H, be the median of H. Then any stable equilibrium has characteristics as in Lemma 1 with prices satisfying*

$$\begin{aligned} p_1 &= \frac{p_2}{2} + \frac{z_1 - z_2}{2} \\ p_i &= \frac{1}{2}\left(\frac{z_i - z_{i+1}}{z_{i-1} - z_{i+1}} p_{i-1} + \frac{z_{i-1} - z_i}{z_{i-1} - z_{i+1}} p_{i+1}\right) \\ p_n &= \frac{p_{n-1} z_n}{2 z_{n-1}}, \end{aligned}$$

where the second equation holds for all $i \in \{2, ..., n-1\}$.

The corollary follows immediately from using the uniform distribution for G in Lemma 1 and setting $\zeta(H)$ to be the median of a given distribution H.

3.2 Journal Prices and Welfare

We typically identify an efficient market as one in which prices are equal to marginal cost. We can also measure utilitarian welfare directly. In this section we look at journal prices and welfare in stable equilibria.

While it is not straightforward to determine all journal prices by solving the system of equations given in Corollary 1, we can provide upper and lower bounds on some prices.

Proposition 1 *Let G be the uniform distribution on $[0,1]$. Let $\zeta(H)$, for any distribution H, be the median of H. Then in any stable equilibrium we have, for any journal $i \in \{2, ..., n\}$,*

$$p_i < \frac{1}{2} p_{i-1},$$

and

$$\frac{1}{2}(z_1 - z_2) < p_1 < \frac{2}{3}(z_1 - z_2).$$

Proof: Consider journal $i \in \{2, ..., n-1\}$. By Lemma 1 we have that $p_i \leq p_{i-1}$. By Corollary 1 we have

$$p_i = \frac{1}{2}\left(\frac{z_i - z_{i+1}}{z_{i-1} - z_{i+1}} p_{i-1} + \frac{z_{i-1} - z_i}{z_{i-1} - z_{i+1}} p_{i+1}\right).$$

Together this implies that

$$p_i < \frac{1}{2}\left(\frac{z_i - z_{i+1}}{z_{i-1} - z_{i+1}} p_{i-1} + \frac{z_{i-1} - z_i}{z_{i-1} - z_{i+1}} p_{i-1}\right) = \frac{1}{2} p_{i-1}.$$

Now consider journal n. By Corollary 1 we have

$$p_n = \frac{p_{n-1} z_n}{2 z_{n-1}}.$$

By Lemma 1 we have that $z_n < z_{n-1}$. Together this implies that $p_n < p_{n-1}/2$. This concludes the proof of the first claim of this proposition.

To prove the upper bound in the second claim we use the just established inequality $p_2 < p_1/2$ and the first equation in Corollary 1, $p_1 = p_2/2 + (z_1 - z_2)/2$. Together this implies that $p_1 < p_1/4 + (z_1 - z_2)/2$, and thus, $p_1 < 2(z_1 - z_2)/3$. For the upper bound in the second claim we use the fact that $p_2 > 0$ by Lemma 1 and equation $p_1 = p_2/2 + (z_1 - z_2)/2$ to obtain that $p_1 > (z_1 - z_2)/2$. QED

The following proposition provides bounds for the demand for the top journal, journal one.

Proposition 2 *Let G be the uniform distribution on $[0,1]$. Let $\zeta(H)$, for any distribution H, be the median of H. Then in any stable equilibrium we have that the demand for journal one satisfies*

$$\frac{1}{2} < D_1(z,p) < \frac{2}{3}.$$

Proof: By Lemma 1, point 7, applying the assumption that G is uniform on $[0,1]$, we obtain $D_1(z,p) = p_1/(z_1 - z_2)$. Using the second claim in Proposition 1 establishes the result. QED

We now consider utilitarian welfare. Unconstrained optimal welfare in this market model would be achieved by having all libraries acquiring the top journal, journal one, and given by

$$W^* = \frac{1}{2} z_1.$$

Equilibrium welfare is given as follows. Using the same notation as in Lemma A5 in the appendix, let $\theta_i = (p_i - p_{i+1})/(z_i - z_{i+1})$ for all $i \in \{1, 2, ..., n-1\}$ and let $\theta_n = p_n/z_n$. For convenience let also $\theta_0 = 1$, $\theta_{n+1} = 0$, and $z_{n+1} = 0$. Then utilitarian welfare given equilibrium (z,p) is given by

$$W^E = \sum_{i=1}^{n+1} \int_{\theta_i}^{\theta_{i-1}} \theta z_i d\theta.$$

The absolute equilibrium welfare loss is then given by

$$\Delta W = W^* - W^E = \sum_{i=1}^{n+1} \int_{\theta_i}^{\theta_{i-1}} \theta (z_1 - z_i) d\theta,$$

and the relative equilibrium welfare loss is given by $(\Delta W)/W^*$.

Proposition 3 *Let G be the uniform distribution on $[0,1]$. Let $\zeta(H)$, for any distribution H, be the median of H. Then in any stable equilibrium absolute welfare loss is bounded from below by*

$$\Delta W > \frac{1}{18}(z_1 - z_2),$$

and relative welfare loss bounded from below by

$$\frac{\Delta W}{W^*} > \frac{1}{9}\frac{z_1 - z_2}{z_1}.$$

Proof: We have

$$\begin{aligned}
\Delta W &= \sum_{i=1}^{n+1}\int_{\theta_i}^{\theta_{i-1}} \theta(z_1 - z_i)d\theta = \int_{\theta_1}^{\theta_0} \theta(z_1 - z_1)d\theta + \sum_{i=2}^{n+1}\int_{\theta_i}^{\theta_{i-1}} \theta(z_1 - z_i)d\theta = \\
&= \sum_{i=2}^{n+1}\int_{\theta_i}^{\theta_{i-1}} \theta(z_1 - z_i)d\theta > \sum_{i=2}^{n+1}\int_{\theta_i}^{\theta_{i-1}} \theta(z_1 - z_2)d\theta > \\
&> \int_{0}^{\theta_1} \theta(z_1 - z_2)d\theta > (z_1 - z_2)\int_{0}^{\theta_1} \theta d\theta > \\
&> (z_1 - z_2)\frac{1}{2}\theta_1^2 > (z_1 - z_2)\frac{1}{2}\frac{1}{9},
\end{aligned}$$

where the first step is by definition, the fourth follows from $z_i < z_2$ for all $i > 2$ by Lemma 1, and the last follows from the fact that $\theta_1 = 1 - D_1(z, p)$ and then using the lower bound in Proposition 2. The remaining steps follow from simple algebra. QED

3.3 Three Examples

In this section we look at three examples for F, the distribution of paper quality, and apply the previous results in these examples. The three distributions we use here are the uniform distribution on the interval $[0, 1]$ with $F(z) = z$ for $z \in [0, 1]$, the exponential distribution with parameter $\lambda > 0$ with cumulative distribution function $F(z) = 1 - e^{-\lambda z}$ for $z \geq 0$, and a Pareto-distribution cumulative distribution function $F(z) = 1 - (1/(1 + z))^2$. Table 1 summarizes the main findings for these three examples.

		uniform	exponential	Pareto
z_1	$=$	$\frac{2n-1}{2n}$	$\frac{\ln(2n)}{\lambda}$	$\sqrt{2n} - 1$
$z_1 - z_2$	$=$	$\frac{1}{n}$	$\frac{\ln 3}{\lambda}$	$\sqrt{2n}\left(1 - \sqrt{\frac{1}{3}}\right)$
p_1	$\geq$	$\frac{1}{2}\frac{1}{n}$	$\frac{1}{2}\frac{\ln 3}{\lambda}$	$\frac{1}{2}\sqrt{2n}\left(1 - \sqrt{\frac{1}{3}}\right)$
p_1	$\leq$	$\frac{2}{3}\frac{1}{n}$	$\frac{2}{3}\frac{\ln 3}{\lambda}$	$\frac{2}{3}\sqrt{2n}\left(1 - \sqrt{\frac{1}{3}}\right)$
ΔW	$\geq$	$\frac{1}{18}\frac{1}{n}$	$\frac{1}{18}\frac{\ln 3}{\lambda}$	$\frac{1}{18}\sqrt{2n}\left(1 - \sqrt{\frac{1}{3}}\right)$
$\frac{\Delta W}{W^*}$	$\geq$	$\frac{1}{9}\frac{2}{2n-1}$	$\frac{1}{9}\frac{\ln 3}{\ln(2n)}$	$\frac{1}{9}\frac{1-\sqrt{\frac{1}{3}}}{1-\sqrt{\frac{1}{2n}}}$

Table 1: Table summarizing characteristics of the market for scientific journals for three different distributions of individual paper quality.

For the case of the uniform distribution, as an example of all distributions with bounded support, journal prices and equilibrium welfare loss relative to the first best can be shown to tend to zero as the number of journals (and publishers) tends to infinity. For the case of the exponential distribution, as an example of unbounded support with a fairly "thin" upper tail, journal prices (of the top journals) remain positive in the limit and there are absolute welfare losses relative to the first best even in the limit, but the relative welfare loss tends to zero. For the Pareto distribution, as an example of a distribution with a "fat" upper tail, journal prices as well as absolute and relative welfare loss remain significantly positive even in this limit where the number of journals (and publishers) tends to zero. This latter case is the empirically most plausible as is demonstrated in Appendix A.

4 Discussion

4.1 Regulation

The analysis of our model makes a clear case for regulating the scientific journal publishing industry. We show that the market for scientific publishing has elements of a "natural monopoly" through the endogenous determination of quality differentiation among journals, which is determined by the scientific writers and not the journals themselves. As the marginal cost of providing journal access to one more library is known (and is essentially zero), regulating the industry would probably be very easy. At the EU level, for instance, one could simply force all publishers to sell their journals for a fixed price somewhere near zero, but high enough for the publishers to make some profit.[5] Perhaps one could allow them to charge prices that non-for-profit journals charge, such as many society journals.

The question of regulating the scientific publishing industry was raised at least once before, by the British Office of Fair Trading (OFT)[6] who in 2002 started an investigation into the so-called scientific technical and mathematics (STM) journal market. The OFT, see Office of Fair Trading (2002), agreed with Bergstrom (2001) that publishers were not able to deliver any convincing reason for the strong price increases in the market. Nevertheless, the OFT saw no necessity to step in. The market, supposedly, was experiencing strong changes around the turn of the millennium. Technological change, supposedly, would increase the importance of electronic journals and decrease marginal costs in the market to a negligible level. The OFT therefore expected that the market will self-regulate, and no further investigation took place.

4.2 Bundling and the "Big Deal"

In our model each publisher only owns one journal, an assumption that is clearly wrong. If we change our model to allow some publishers to own more than one journal, then clearly the market would be even less competitive. Moreover, if some publishers did own more than one journal, these publishers could engage in additional second degree prices discrimination by, for instance, offering bundles. This is something publishers actually do, and they would also have incentives to do so in our model. For more on bundling of

[5] In France the NGO couperin.org has tried this with some success.

[6] The Office of Fair Trading was responsible for protecting consumer interests throughout the UK. It closed on 31.03.2014 and its responsibilities have been passed to a number of different organizations.

journals see e.g. Frazier (2001), McCabe et al. (2004), Frazier (2005), Bergstrom (2010), and Bergstrom, Courant, McAfee, and A.Williams (2014). All this will likely make the market even less competitive.

4.3 Partial quality control by journals

While editors and publishers of a journal cannot completely control its quality because of the mentioned coordination motive, they could control how many articles they want to publish in any given year. Effectively they have some control over the lower bound on quality subject to this lower bound being lower than the upper bound that is imposed on them by other journals and the authors' behavior. We conjecture that allowing journals in our model to control this lower bound would lead to even more inefficiency. But to assess this properly the model would have to be adapted in several aspects. Not only would we need to expand the strategy space of publishers to include this partial quality control, but we would also need to endow libraries with preferences that reflect not only a trade-off between quality and price but also the size of journals.

A On the distribution of paper quality

How does one measure the quality of a scientific article? According to Bergstrom, West, and Wiseman (2008, p. 236), "[t]here is only one adequate approach to evaluating the quality of an individual paper: read it carefully, or talk to others who have done so." This is the approach that is used, more or less, by editors and referees when they assess an individual paper. It is also used, to some extent, for tenure decisions in top universities. From the point of view of a non-expert (in the particular field of the given paper) outside observer, however, this approach is not practical. As a rough proxy for quality the community on the whole has turned to looking at the number of citations that a paper accumulates over time.

The idea of constructing citation indexes goes back to Garfield (1955). See Garfield (2006) for a history of this citation-counting based measuring approach. For a list of shortcomings of this approach see e.g., Ritzberger (2008, p. 404-405); see also Seglen (1997). Despite these shortcomings or imperfections citation-counting has been used not only to assess the quality of individual papers, but also the quality of researchers (e.g., the "Handelsblatt-Ranking" of economics researchers in the German speaking world), and of journals based, for instance, on the fixed-point approach of Palacios-Huerta and Volij (2004), West, Bergstrom, and Bergstrom's (2010) Eigenfactor approach, or simply the basic Journal Impact Factor based directly on the original idea by Garfield (1955).[7]

If we take citation-counting as our best proxy for paper quality, what distribution of paper quality does this imply? To assess this, we look at data obtained from the Web of Science of all 13,537 economics papers published in the year 1999 and the number of citations that each of these papers had by the end of February 2013. There is nothing special about

[7]See e.g., Monastersky (2005) for a critical assessment of the increasing influence of citation-counting in hiring, tenure decisions, and the awarding of grants.

1999 and 2013, and we conjecture that we could obtain similar findings with data for other years.

Figure 1 plots the empirical frequency of citations of these papers and contrasts this with the density functions of the exponential distribution and the Pareto distribution (with shape parameter equal to two), where the remaining parameters in both distributions are chosen such as to exactly match the theoretical and empirical mean. The sample mean of citations in our data is 17.43 (with a median of 4). The appropriate exponential distribution, thus, has a density given by $f(z) = \lambda e^{-\lambda z}$ with $\lambda = 1/17.43$ to match the sample mean. The appropriate Pareto distribution has a density given by $f(z) = 2x_m^2/(z + x_m)^3$ with mean $x_m = 17.43$.

Neither the exponential nor the Pareto-distribution can match the empirical frequency of 30.81% of zero citations in our data set. But note two things. First, even papers with zero citations probably have some at least ex-ante value that is simply not measured well by citation-counting. Second, the more important part of the distribution for our results is the upper tail-behavior.

From Figure 1 it is apparent that, especially for moderate to high citations, the Pareto distribution is a much better fit than the exponential distribution.

B Proof of Lemma 1

We here provide a series of lemmas that finally combine to prove Lemma 1.

For any pair (z, p) define $A(z, p) = \{i \in \{1, ..., n\} \mid D_i(z, p) > 0\}$ as the set of *active journals.*

Lemma A2 *Consider an equilibrium (z, p). Consider two journals $i, j \in A(z, p)$. Then $z_i > z_j$ if and only if $p_i > p_j$.*

Proof: A library of type θ strictly prefers to buy journal i over journal j if and only if $\theta z_i - p_i > \theta z_j - p_j$ or if and only if

$$\theta > \frac{p_i - p_j}{z_i - z_j}.$$

As G, the distribution of θ, admits a density it must be atomless. Therefore, for both journals to have positive demand we must have that $(p_i - p_j)/(z_i - z_j) > 0$. Thus, if $z_i > z_j$ we must have $p_i > p_j$ and vice versa. □

Lemma A3 *Consider an equilibrium (z, p). Consider two journals $i, j \in A(z, p)$ with $i \neq j$. Suppose $z_i = z_j$. Then $p_i = p_j = 0$ and for all journals $l \in A(z, p)$ with $z_l \leq z_i$ we have $p_l = 0$ and $z_l = z_i$.*

Proof: A library of type θ strictly prefers to buy journal i over journal j if and only if $\theta z_i - p_i > \theta z_j - p_j$. If $z_i = z_j$, any library of any type θ strictly prefers to buy journal i over journal j if and only if $p_i < p_j$. For both journals to have positive demand we

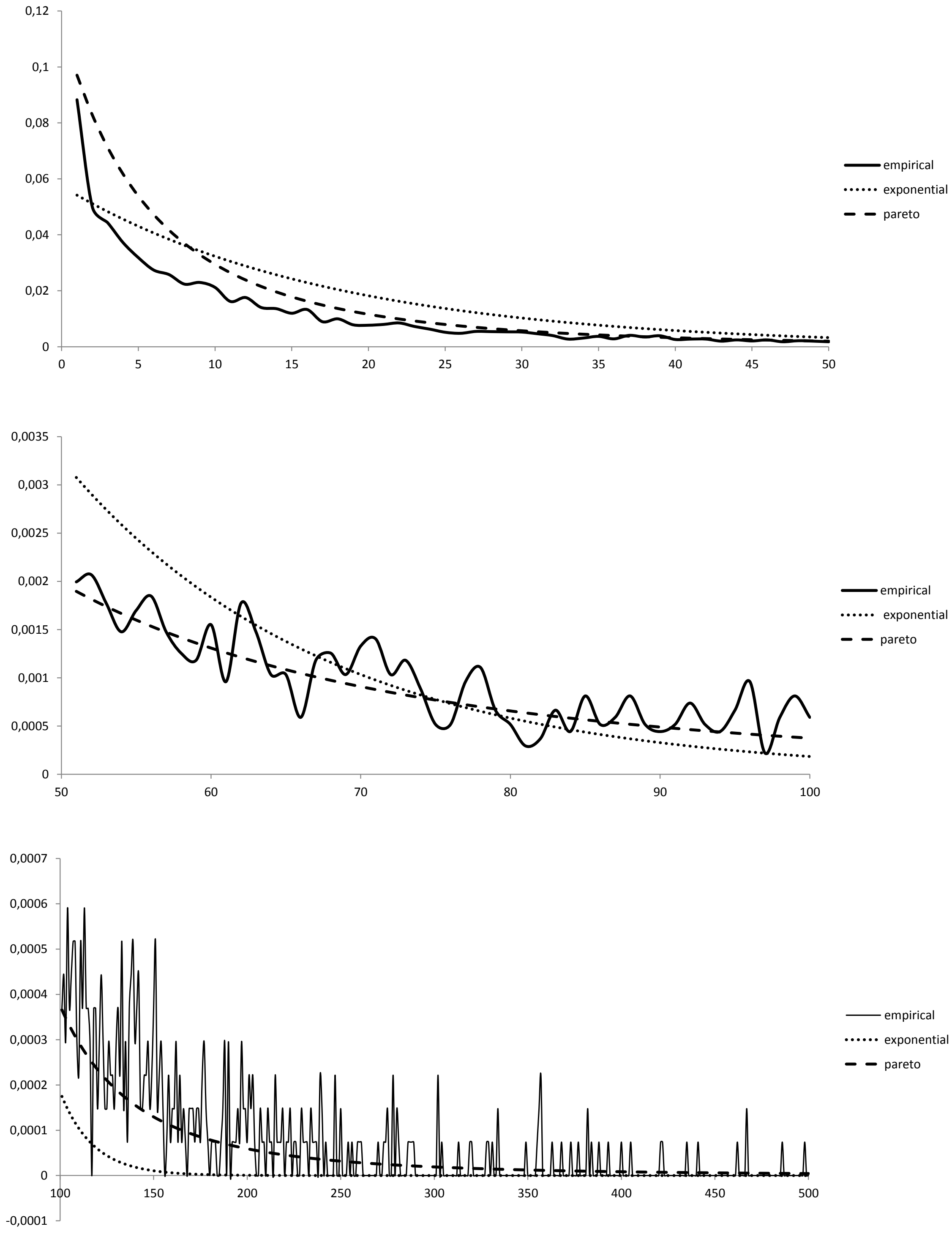

Figure 1: Citations of economics papers published in 1999. Comparison of the empirical frequency distribution and the exponential and Pareto distribution.

must have $p_i = p_j$. But then by the logic of homogeneous good Bertrand competition we must have $p_i = p_j = 0$. The remaining statement of this lemma immediately follows from Lemma A2. □

Lemma A4 *Consider an equilibrium (z, p) such that there are two journals $i, j \in A(z, p)$ with $i \neq j$ and $z_i = z_j$. Then this equilibrium is not stable.*

Proof: By Lemma A3 we must have $p_i = p_j$. For authors with a given paper quality z to submit to both journals i and j we must have that they are indifferent and, thus, that $D_i(z, p) = D_j(z, p)$. Let $\epsilon > 0$ and consider a vector of journal qualities z' such that $z'_l = z_l$ for all $l \neq i, j$, $z'_i = z_i + \epsilon$, and $z'_j = z_j - \epsilon$. Then $D_i(z', p) \geq D_i(z, p)$ and $D_j(z', p) = 0$. Then, given the vector of journal qualities z' all authors would strictly prefer to submit to journal i over j. This violates the definition of a stable equilibrium. □

Lemma A5 *Consider an equilibrium (z, p) such that $A(z, p) = \{1, 2, ..., k\}$ with $2 \leq k \leq n$ with $z_1 > z_2 > ... > z_k$. Then there are cutoffs θ_i, $i = 0, 1, ..., k$, such that $\theta_0 = \infty > \theta_1 > \theta_2 > ... > \theta_{k-1} > \theta_k > 0$ and a library of type θ buys journal $j \in A(z, p)$ if $\theta \in (\theta_j, \theta_{j-1})$ and only if $\theta \in [\theta_j, \theta_{j-1}]$. Moreover for $i \in \{1, ..., k-1\}$, $\theta_i = (p_i - p_{i+1})/(z_i - z_{i+1})$ and $\theta_k = p_k/z_k$.*

Proof: Consider any three journals i, j, l with $1 \leq i < j < l \leq k$. We need to show that there is an interval of library types such that libraries of a type below the lower bound of this interval prefer journal l over the other two, libraries with a type within the interval prefer journal j over the other two, and libraries of a type above the upper bound of this interval prefer journal i over the other two.

A library of type θ strictly prefers journal i over journal j if and only if $\theta > \theta_{ij} = (p_i - p_j)/(z_i - z_j)$, strictly prefers journal i over journal l if and only if $\theta > \theta_{il} = (p_i - p_l)/(z_i - z_l)$, and strictly prefers journal j over journal l if and only if $\theta > \theta_{jl} = (p_j - p_l)/(z_j - z_l)$. By assumption we have $z_i > z_j > z_l$ and by Lemma A2 we have $p_i > p_j > p_l$. Thus, we have that $\theta_{ij}, \theta_{il}, \theta_{jl} > 0$.

First, we show that $\theta_{ij} > \theta_{jl}$. Suppose not. Suppose $\theta_{ij} < \theta_{jl}$. There are three cases to consider. Case 1: Suppose $\theta_{il} > \theta_{jl}$. Then a type $\theta \in (\theta_{jl}, \theta_{il})$ prefers l over i, i over j, and j over l, a contradiction to the fact that any library by having a well-defined utility function must have transitive preferences. Case 2: Suppose $\theta_{il} \in (\theta_{ij}, \theta_{jl})$. Then every type θ's best choice is either l or i, never j, a contradiction to the assumption that $j \in A(z, p)$. Case 3: Suppose $\theta_{il} < \theta_{ij}$. Then a type $\theta \in (\theta_{il}, \theta_{ij})$ prefers i over l, l over j, and j over i, a contradiction to the fact that any library by having a well-defined utility function must have transitive preferences.

Second, we show that $\theta_{il} \in (\theta_{jl}, \theta_{ij})$. Suppose not. There are two cases to consider. Case 1: Suppose $\theta_{il} > \theta_{ij}$. Then a type $\theta \in (\theta_{ij}, \theta_{il})$ prefers i over j, j over l, and l over i, a contradiction to the fact that any library by having a well-defined utility function must have transitive preferences. Case 2: Suppose $\theta_{il} < \theta_{jl}$. Then a type $\theta \in (\theta_{il}, \theta_{jl})$ prefers i over l, l over j, and j over i, a contradiction to the fact that any library by having a well-defined utility function must have transitive preferences.

Thus, we have $\theta_{ij} > \theta_{jl}$ and any type $\theta < \theta_{jl}$ prefers journal l over the other two, any type $\theta \in (\theta_{jl}, \theta_{ij})$ prefers journal j over the other two, and any type $\theta > \theta_{ij}$ prefers journal l over the other two. □

Lemma A6 *Consider an equilibrium* (z, p) *such that* $A(z, p) = \{1, 2, ..., k\}$ *with* $1 \leq k \leq n$ *with* $z_1 > z_2 > ... > z_k$. *Then* $k = n$.

Proof: By Lemma A2 and Lemma A5 we have that $p_k > 0$. Consider a journal $i \notin A(z, p)$. This means that $D_i(z, p) = 0$. Note that all active journals together publish a fraction $k/n < 1$ of all papers. The authors' incentives are such, however, that they submit their paper to a journal even if its demand is zero. In fact, every journal receives a fraction of $1/n$ papers and all papers have positive quality with probability one. Thus, journal i has positive quality z_i, and it could improve its profits from zero to a positive amount by charging a small positive price. QED

Proof of Lemma1: A stable equilibrium (z, p) must satisfy that all active journals have distinct qualities by Lemma A4. Lemma A6 then implies that all journals must be active and we have that $z_1 > z_2 > ... > z_n$. Lemma A2 implies that $p_1 > p_2 > ... > p_n$. Points 4-6 and that $z_n > 0$ and $p_n > 0$ then follow from Lemma A5. Point 1 follows from the fact that $z_1 > z_2 > ... > z_n$. Points 2 and 3 follow directly from point 1. Point 7 is then derived from the first order conditions for profit maximization of the n publishers. QED

References

Attema, A. E., W. B. Brower, and J. V. Exel (2014): "Your right arm for a publication in AER?," *Economic Inquiry*, 52(1), 495–502.

Bergstrom, C. T., J. D. West, and M. A. Wiseman (2008): "The Eigenfactor metrics," *Journal of Neuroscience*, 28(45), 11433–11434.

Bergstrom, T. C. (2001): "Free labor for costly journals?," *Journal of Economic Perspectives*, 15(4), 183–198.

——— (2010): "Librarians and the terrible fix: Economics of the Big Deal," *Serials*, 23(2).

Bergstrom, T. C., P. N. Courant, R. P. McAfee, and M. A.Williams (2014): "Evaluating big deal journal bundles," *Proceedings of the National Academy of Sciences*, 111(26), 9425–9430.

Frazier, K. (2001): "The librarian's dilemma: Contemplating the costs of the "big deal."," *D-Lib Magazine*, 7(3).

——— (2005): "What's the big deal?," *The Serials Librarian*, 48(1-2), 49–59.

Garfield, E. (1955): "Citation indexes to science: a new dimension in documentation through association of ideas," *Science*, 122, 108–111.

——— (2006): "The history and meaning of the journal impact factor," *Jama*, 295(1), 90–93.

JEON, D.-S., AND J.-C. ROCHET (2010): "The pricing of academic journals: A two-sided market perspective," *American Economic Journal: Microeconomics*, 2(2), 222–255.

LARIVIÈRE, V., S. HAUSTEIN, AND P. MONGEON (2015): "The oligopoly of academic publishers in the digital era," *PloS one*, 10(6), e0127502.

MCCABE, M. J., ET AL. (2004): "Information goods and endogenous pricing strategies: The case of academic journals," *Economics Bulletin*, 12(10), 1–11.

MCCABE, M. J., AND C. M. SNYDER (2005): "Open access and academic journal quality," *The American Economic Review*, 95(2), 453–458.

——— (2015): "Does online availability increase citations? Theory and evidence from a panel of economics and business journals," *Review of Economics and Statistics*, 97(1), 144–165.

MCCABE, M. J., AND C. M. SNYDER (2016): "Open Access as a Crude Solution to a Hold-up Problem in the Two-Sided Market for Academic Journals," Discussion paper, National Bureau of Economic Research.

MCCABE, M. J., C. M. SNYDER, ET AL. (2007): "Academic journal prices in a digital age: A two-sided market model," *The BE Journal of Economic Analysis and Policy*, 7(1), 1–37.

MONASTERSKY, R. (2005): "The number thats devouring science," *Chronicle of Higher Education*, 52(8), 14.

OFFICE OF FAIR TRADING (2002): "The market for scientific, technical and medical journals," Discussion paper, OFT, A statement by the OFT.

PALACIOS-HUERTA, I., AND O. VOLIJ (2004): "The measurement of intellectual influence," *Econometrica*, 72(3), 963–977.

RITZBERGER, K. (2008): "A ranking of journals in economics and related fields," *German Economic Review*, 9(4), 402–430.

SEGLEN, P. O. (1997): "Why the impact factor of journals should not be used for evaluating research.," *BMJ: British Medical Journal*, 314(7079), 498.

SHAKED, A., AND J. SUTTON (1982): "Relaxing price competition through product differentiation," *The review of economic studies*, pp. 3–13.

TENOPIR, C., AND D. W. KING (2000): *Towards electronic journals: realities for scientists, librarians, and publishers.* Special Libraries Association, Washington, DC.

WEST, J. D., T. C. BERGSTROM, AND C. T. BERGSTROM (2010): "The Eigenfactor™ Metrics: A network approach to assessing scholarly journals," *College & Research Libraries*, 71(3), 236–244.

Graz Economics Papers

For full list see:
http://ideas.repec.org/s/grz/wpaper.html
Address: Department of Economics, University of Graz,
Universitätsstraße 15/F4, A-8010 Graz

11–2017 **Philipp Kohlgruber and Christoph Kuzmics**: The distribution of article quality and inefficiencies in the market for scientific journals

10–2017 **Maximilian Goedl**: The Sovereign-Bank Interaction in the Eurozone Crisis

09–2017 **Florian Herold and Christoph Kuzmics**: The evolution of taking roles

08–2017 **Evangelos V. Dioikitopoulos, Stephen J. Turnovsky and Ronald Wendner**: Dynamic Status Effects, Savings, and Income Inequality

07–2017 **Bo Chen, Yu Chen and David Rietzke**: Simple Contracts under Observable and Hidden Actions

06–2017 **Stefan Borsky, Andrea Leiter and Michael Paffermayr**: Product Quality and Sustainability: The Effect of International Environmental Agreements on Bilateral Trade

04–2017 **Stefan Borsky, Alexej Parchomenko**: Identifying Phosphorus Hot Spots: A spatial analysis of the phosphorus balance as a result of manure application

03–2017 **Yu Chen, Yu Wang, Bonwoo Koo**: Open Source and Competition Strategy Under Network Effects

02–2017 **Florian Brugger**: The Effect of Foreign and Domestic Demand on U.S. Treasury Yields

01–2017 **Yu Chen**: Simplifiability of Contingency Plans with Applications to Optimal Mechanism Design

13–2016 **Jan de Haan, Rens Hendriks, Michael Scholz**: A Comparison of Weighted Time-Product Dummy and Time Dummy Hedonic Indexes

12–2016 **Katja Kalkschmied**: Complementary Institutional Elements and Economic Outcomes

11–2016 **Dieter Balkenborg, Josef Hofbauer, Christoph Kuzmics**: The refined best reply correspondence and backward induction

10–2016 **Karl W. Steininger, Wolf D. Grossmann, Iris Grossmann**: The economic value of dispatchable solar electricity: a Post-Paris evaluation

09–2016 **Sofie R. Waltl**: Estimating aggregate quantile-specific gross rental yields for residential housing in Sydney

08–2016 **Karl Farmer**: Financial integration and house price dynamics in equilibrium modeling of intra-EMU and global trade imbalances

07–2016 **Gabriel Bachner, Birgit Bednar-Friedl**: Counterbalancing the Effects of Climate Change Adaptation on Public Budgets: Factor Taxes, Transfers, or Foreign Lending?

06–2016 **Margareta Kreimer, Ricardo Mora**: Segregated Integration: Recent Trends in the Austrian Gender Division of Labor

05–2016 **Karl Farmer, Bogdan Mihaiescu**: Credit constraints and differential growth in equilibrium modeling of EMU and global trade imbalances

04–2016 **Daniel Eckert, Frederik Herzberg**: The birth of social choice theory from the spirit of mathematical logic: Arrow's theorem in the framework of model theory

03–2016 **Christoph Kuzmics, Jan-Henrik Steg**: On public good provision mechanisms with dominant strategies and balanced budget

02–2016 **Florian Gauer, Christoph Kuzmics**: Cognitive Empathy in Conflict Situations

01–2016 **Thomas Aronsson, Olof Johansson-Stenman, Ronald Wendner**: Redistribution Through Charity, and Optimal Taxation when People are Concerned with Social Status

09–2015 **Sofie R. Waltl**: Variation across price segments and locations: A comprehensive quantile regression analysis of the Sydney housing market

08–2015 **Karl Farmer, Bogdan Mihaiescu**: Pension systems and financial constraints in a three-country OLG model of intra-EMU and global trade imbalances

07–2015 **Javier López Prol, Karl W. Steininger**: Photovoltaic self-consumption regulation in Spain: profitability analysis and alternative regulation schemes.

06–2015 **Miriam Steurer, Caroline Bayr**: Quantifying Urban Sprawl using Land Use Data

05–2015 **Porfirio Guevara, Robert J. Hill, Michael Scholz**: Hedonic Indexes for Public and Private Housing in Costa Rica

04–2015 **Sofie R. Waltl**: A Hedonic House Price Index in Continuous Time

03–2015 **Ronald Wendner**: Do Positional Preferences for Wealth and Consumption Cause Inter-Temporal Distortions?

02–2015 **Maximilian Gödl, Christoph Zwick**: Stochastic Stability of Public Debt: The Case of Austria

01–2015 **Gabriel Bachner**: Land Transport Systems under Climate Change: A Macroeconomic Assessment of Adaptation Measures for the Case of Austria