EMERGENT PHENOMENA IN SCIENTIFIC PUBLISHING: A SIMULATION EXERCISE

JAKOB KAPELLER AND STEFAN STEINERBERGER

This paper is going to appear in Research Policy

Abstract. We model the academic production process understood as the creation, submission, evaluation and publication of papers: scientists produce manuscripts to the best of their abilities and try to publish them in academic journals, which rely on referees to judge the submissions. The resulting model is able to reproduce several properties of the journal-landscape but also illustrates that even under unrealistically optimistic assumptions the process of scientific publishing will give rise to several universal emergent phenomena for purely mathematical reasons: the efficiency of scientific publishing is delicate and very unstable.

1. Introduction

Determining the relative merit of a specific academic contribution is often a daunting task. Science has developed institutions – like academic journals – as well as accompanying routines – like peer review – to systematically address this task. Today these institutions play a dominant and influential role even though past research successfully documented their limitations: peer review, for instance, is known to lack robustness (Gans and Shepherd 1994) as well as objectivity (Bedeiean 2003) and might give rise to shrewed incentives (McDonnel and Kam 2010, Day 2015). Citation metrics are similarly contested, as they suffers from a general bias due to the skewed distribution of academic attention (Solla-Price 1965), entail substantial problems of internal validity (because citation counting measures relative impact as a proxy for quality; Amin and Mabe 2000), incorporate a series of conceptual biases (Kapeller 2010) and induce reactive behavior among authors, reviewers and editors (Reedijk and Moed 2006). In this paper we take a different perspective on evaluation in academia. We will assume that these problems simply do not exist and that evaluation procedures are valid, transparent, fair and as objective as possible: we aim to explore the properties of scientific discourse under the assumption that two main evaluative instruments in science – peer review and journal rankings – function rather objectively.

In assessing this question we employ a simple model simulating the academic production process: scientists produce manuscripts of different quality and try to publish these papers in journals. Journals, conversely, try to select those manuscripts with the highest quality for publication and rely on inputs from referees to make that judgement. The model is kept as simple as possible. Our main contribution is to show that even under overly optimistic assumptions, where decisions and journal-rankings are completely objective and distorting factors like opportunistic behavior or academic feuds are absent, the underlying structure of scientific publishing will inevitably exhibit idiosyncratic properties. Specifically, we show that scientific publishing can only be efficient in an *idealized* setting, where authors, referees and journals are perfectly objective and accurate. This idealized scenario turns out to be extremely unstable and already a tiny amount of noise fundamentally alters basic properties of the academic production process for the worse. We present our model in Section 2 and discuss the main results in Section 3. Summarizing, our paper aims to demonstrate that the way scientific publishing operates will give rise to a series of interesting and partly unexpected phenomena; some of these phenomena are harmful to scientific progress

Key words and phrases. Citation metrics, simulation, reviewing, manuscript quality, submission behavior.

despite the best intentions of authors, editors and referees and, hence, provide an additional point of departure for a critical assessment of the inner routines of academic institutions.

2. The model

2.1. **Introduction.** Our model tries to encapsulate the essence of the academic production process assuming (a) that only quality matters and (b) that the quality of academic products, like journals and manuscripts, is assessed as objectively and precisely as possible. Our model has a variety of parameters giving rise to a natural dichotomy: we distinguish between

> the *idealized scenario:* authors have perfect understanding of the quality of their work, which is submitted to the appropriate journal and judged by a referee who also has perfect judgement
> the *noisy scenario:* authors have an approximate understanding of the quality of their work, act strategically in the context of journal-submissions and are being judged by referees, who guess the true quality of the paper up to a small error

The idealized scenario behaves pretty much in the way one would imagine an idealized world of scientific publishing to behave: there exists a well-ordered journal landscape, authors submit to appropriate journals, papers are being published in journals whose reputation closely corresponds to the quality of the paper and rejection rates are rather low. Our focus is on *universal* emergent properties and qualitative phenomena in the noisy scenario. Universality here refers to the fact that we only care about phenomena that are *independent of the actual parameters* – clearly, as different parameters will give rise to different outcomes, these phenomena have to be described qualitatively rather than quantitatively. We shall now describe the model and will, for the sake of clarity, explicitly fix variables – however, we emphasize that the structures arising in the noisy scenario are stable under perturbing parameters and therefore quite independent of the type of error. Conversely, the idealized scenario is highly sensitive to even very slight changes.

2.2. **Scientists and Manuscripts.** We start with a fixed number of scientists $\mathrm{sci}_1, \dots, \mathrm{sci}_N$ (we use $N = 200$). These scientists come with different levels of skill, where skill is described by a real number and is chosen randomly (we use $\mathcal{N}(0.5, 0.2)$ normal distribution, where $\mathcal{N}(\mu, \sigma)$ denotes the Normal Distribution with mean μ and standard deviation σ). Every scientist sci_j now writes a paper: the quality $qual_j$ randomly fluctuates around the scientists' level of skill and is given by, say, a Gaussian fluctuation, we use

$$\mathrm{qual}_j = \mathrm{sci}_j(1 + \mathcal{N}(0, 0.2)).$$

The quality of the paper therefore depends on the skill of the scientist but even 'weak' scientists are capable of occasionally producing outstanding work. Conversely, outstanding scientists produce better papers on average and are more likely to produce truly outstanding work, however, they will also occasionally produce paper which are much worse than the intrinsic skill of their authors suggests. These N scientists will now submit their N papers to journals.

2.3. **Journals and Submissions.** There is a fixed number k of journals each of which is prepared to publish a fixed number of papers: we will model $k = 25$ journals each of which publishes the best 12 papers that are being submitted to said journal. This means that journals offer 300 slots for 200 papers and, in particular, the average journal will typically receive 8 submissions but is prepared for more. While this assumption is overly optimistic – and strongly unrealistic – it serves to ensure that our results are not a mere artefact of journal-scarcity. The prestige/reputation $\mathrm{jour}_1, \dots, \mathrm{jour}_k$ of the journals is given by a real number (that will eventually be compiled as an average of the quality of papers published there). Submission and evaluation of manuscripts is organized as follows.

(1) **Submission.** In a first step every scientist sci_j tries to estimate the quality of her paper. In the idealized scenario, authors have perfect insight. In the noisy scenario, the estimates

are still quite accurate but subject to a small error.

$$\mathrm{est}_j = \mathrm{qual}_j \qquad \text{(idealized scenario)}$$
$$\mathrm{est}_j = \mathrm{qual}_j(1+\mathcal{N}(0,0.2)) \qquad \text{(noisy scenario)}$$

The estimate in the noisy scenario is still very optimistic: the author is able to estimate the intrinsic value up to a typical error of 20%. The 'appropriate' place would now be the journal with reputation closest to the intrinsic quality qual_j of the paper. The scientists will now submit to the journal that is closest in reputation to $a_1 \cdot \mathrm{est}_j$, where a_1 is a numerical parameter. If the papers get rejected, they will submit to the journal with reputation closest to $a_2 \cdot \mathrm{est}_j$ and, in the case of a second rejection, to the journal closest to $a_3 \cdot \mathrm{est}_j$. There is a maximum of three submission and the specific parameters are

$$(a_1, a_2, a_3) = (1, 0.9, 0.8) \qquad \text{(idealized scenario)}$$
$$(a_1, a_2, a_3) = (1.1, 1, 0.9) \qquad \text{(noisy scenario).}$$

In the idealized scenario the authors are modest, try to submit to the appropriate journal and agree to a try a slightly lower tier in the case of rejection. We remark that the values a_2, a_3 in the idealized scenario have very little impact because the rejection rate is very low. In the noisy scenario, authors are slightly more ambitious at first. We assume in both scenarios that the reputation of a journal jour_j is universally agreed upon and available to all scientists (see 2.4).

(2) **Evaluation.** Journals aim to publish the best papers. Each journal requests a referee report for each submitted paper (a mathematical identity for Gaussians allows to replace multiple referees with Gaussian errors by one referee with a smaller Gaussian error, multiple referee reports are therefore also incorporated in the model, see below). Referees aim for a maximum of accuracy in quality assessment but succeed fully only in the idealized scenario, while they provide slightly distorted estimates in the noisy scenario. We assume the refereeing process to be double-blind: the referee does not know about the skill of the author but tries to judge the paper on its intrinsic merit and assigns a numerical value

$$\mathrm{ref}_j = \mathrm{qual}_j \qquad \text{(idealized scenario)}$$
$$\mathrm{ref}_j = \mathrm{qual}_j(1+\mathcal{N}(0,0.2)) \qquad \text{(noisy scenario)}$$

After having collected all the referee reports, the journal will accept the 'best' papers for which they have space (here a maximum of 12) following referees' reports. They accept submissions in three rounds but will cease to accept new manuscripts for publication after reaching the maximum number of articles to be published within a period.

2.4. **Journal re-evaluation.** The final step is a journal re-evaluation procedure. The reputation of a journal depends on the quality of the published papers within the last period as well as on a time-lag representing the quality of contributions published in the past. We thus replace the reputation jour_j of journal j by a weighted average of current reputation (80%) and the average intrinsic quality of papers published in the last round (20%)

$$\mathrm{jour}_j \leftarrow \frac{4 \cdot \mathrm{jour}_j}{5} + \frac{\text{average quality } \mathrm{qual}_j \text{ of papers published}}{5}$$

and will leave it unchanged in case no papers were published. This guarantees that a journal publishing papers that exceed its reputation in quality will rise in prestige and, conversely, reputable journals publishing weak papers will slowly lose their standing. Hence, we assume that journal rankings are completely efficient and objective in both scenarios. After the re-evaluation procedure has concluded, the entire process will begin from anew (without any changes in the values of sci_j – the skill of the scientists is fixed – and with the updated values for the reputation of the journals). The only unspecified quantity is the initial journal landscape (i.e. the initial distribution of prestige/reputation). It is well understood that in mathematical models of this type, the initial journal landscape is of no importance if the model is simulated for a sufficient amount of time (something that we also observed in our experiments). We start with an equidistant partition of the probability space mapped under the inverse cumulative distribution function of $\mathcal{N}(0, 0.2)$ because that is in first order what one would expect in the idealized scenario.

2.5. **Remarks.**

(1) The process of scientific publishing contains a lot of intrinsic randomness. We hasten to emphasize that our model *drastically* underestimates that degree of randomness, since all other variables, like journal quality, are based on undistorted observations of the "true" quality of contributions. We also completely ignore the possibility of personal feuds, research trends, competing research fields of different size and other factors that may affect the "objectivity" of scientific institutions. More realistic conditions could be implemented but the focus of our work is to show that the existence of curious phenomena, which arise already from a minimum of noise.

(2) In terms of of empirical plausibility, the model closely reflects subdisciplines of natural sciences moving at a fast pace with many research groups working on the same questions (if a paper gets rejected three times, a year has passed and the results are outdated).

2.6. **A short summary of results.** In the idealized scenario the universe of scientific publications is rather well-structured: the relative position of journals within rankings is stable and journals do indeed only publish papers that closely reflect the quality of their past issues.

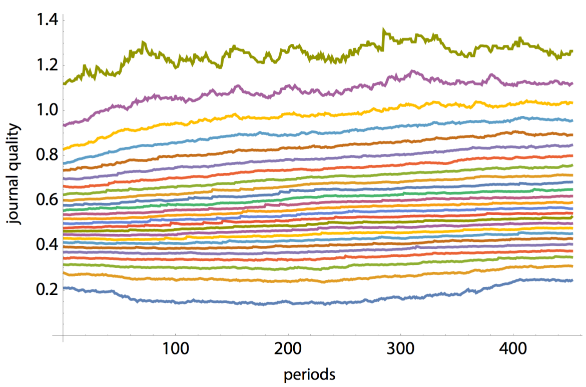

FIGURE 1. Quality of Journals over time (idealized scenario)

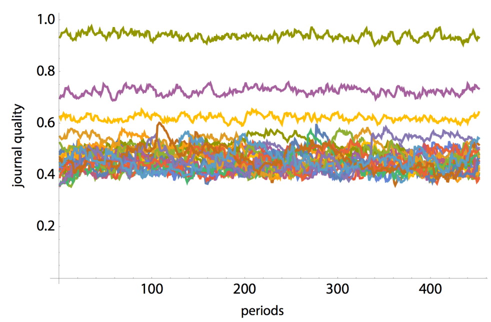

FIGURE 2. Quality of Journals over time (noisy scenario)

Our main contribution is to demonstrate that such an idealized scenario (Figure 1) is *very* unstable and easily perturbed: in the noisy scenario (Figure 2), we find several emerging phenomena that are stable under perturbation of parameters (and will always be found outside of the idealized scenario). We first summarize some of our main findings.

(1) **Top journals.** Our model predicts very few top journals, which publish almost exclusively excellent papers; their position at the top is stable over time.
(2) **Clustering.** Outside of the top journals there is a clustering of many different journals which are virtually indistinguishable; their position in the ranking is not stable.
(3) **Variation.** The quality of papers published in a journal that is not at the very top varies considerably; almost all journals tend to feature both surprisingly good and surprisingly bad contributions compared to their ranking.
(4) **Rejection rate.** Under ideal parameters, the probability of rejection decreases as the quality of a paper increases; however, under even slightly imperfect conditions, the likelihood of rejection is actually increasing as the quality of the paper increases (up until the paper is among the very best papers at which point it sharply decreases).
(5) **Journals as bottleneck.** The journal landscape deviates from the quality of papers being produced: this misfit acts as a bottleneck, which creates unnecessarily high rejection rates

and slows down the publication process, especially in the group of very good, but not stellar papers (say, top 20% but not top 5%).

3. Universally emergent phenomena I: Patterns of quality

We now discuss these phenomena in greater detail and supplement explicit examples; our emphasis is, of course, on the fact that these phenomena are stable under a different choice of probability distribution or perturbation of the actual numerical values. The transition from idealized to noisy is not a slow one: already a very small deviation from idealized parameters has a huge impact.

3.1. **Top journals.** Our first observation is that in the noisy scenario only the ranking of top-journals is stable, while the ranking of the remaining journal-population is rather volatile (see Figures 2-4). Rankings in the idealized scenario, on the other hand, are completely stable (see Figure 1). This phenomenon is easily explained: within our model, there is close correspondence between the quality of manuscripts and journals. Since scientists in the noisy scenario aim at publishing in higher-tier journals, the best journals will receive a lot of submissions – if the number of submissions is large, this can compensate for errors introduced in the refereeing process: simply put, the journal can afford to publish only those papers where both authors and referees believe the paper to be excellent.

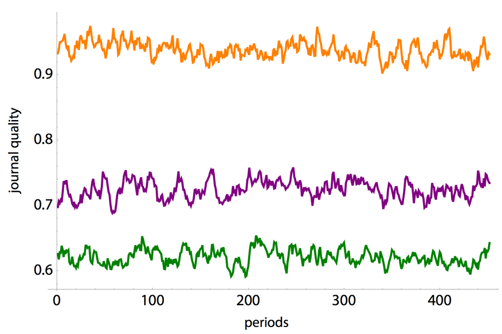

Figure 3. Quality of top 5 journals (noisy scenario)

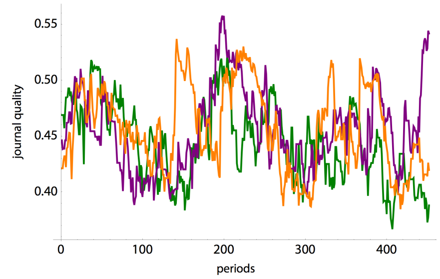

Figure 4. Quality of journals #11 - #13 (noisy scenario)

If a top journal receives a mediocre paper, it is certainly possible for the referee to misjudge the paper (from 'mediocre' to 'very good'); however, since the number of submissions to top journals is large, another referee is equally likely to misjudge another paper (from 'very good' to 'excellent' or even from 'stellar' to merely 'excellent'). We remark that our model implicitly captures the use of more than one referee report: a basic identity for the Gaussian distribution implies that if the error of one referee report is distributed as $\text{ref} \sim \mathcal{N}(\mu, \sigma)$, then an average of k independent referee reports is distributed as

$$\frac{\text{ref}_1 + \text{ref}_2 + \cdots + \text{ref}_k}{k} \sim \mathcal{N}\left(\mu, \frac{\sigma}{\sqrt{k}}\right).$$

Thus journals requesting more than one referee report can be modeled within our framework by changing a parameter. This first result is not surprising: most disciplines seem to have a very clear understanding which journals have the highest prestige – this is usually a very small number of journals, which have a tradition of having been outstanding in the past.

3.2. **Clustering.** The phenomenon of clustering, a large number journals that are not top journals being very similar to each other in terms of reputation, is of more interest (see Figures 2 and 4). Simply put, the clustering effect arises from both ambition (scientists wish to publish in prestigious outlets) as well as erroneous judgement of both author and reviewer: journals in 'the middle of the pack' do not have a clear enough profile to attract a number of submissions comparable to that of top journals and are thus to a greater extent exposed to random fluctuations in quality. However, in certain cases, if by pure chance one journal gets a number of outstanding submissions, it can manage to separate from the herd and establish an independent profile (or, conversely, lose its special status and become one among many).

3.3. **Variation.** It is not uncommon to judge the quality of a paper first by looking at the reputation of the journal it appeared in. In the idealized scenario, the quality of journal is indeed a very good proxy for the quality of the paper. In the noisy scenario, however, this relationship fails drastically: while top journals almost exclusively feature good or outstanding papers, all other journals will feature a surprising variety of papers; see Figures 6 and 7. for the distribution of quality that can be found in the *least ranked* journal in the idealized and noisy scenario, respectively. In the idealized scenario, the 'worst' journal features a selection of papers that is tightly clustered in the region of 'worst' papers; in the noisy scenario, the mean quality of papers published there goes up by a lot and starts featuring a tail – indeed, some of the papers published in the *least* ranked journal are actually *far above average* in quality.

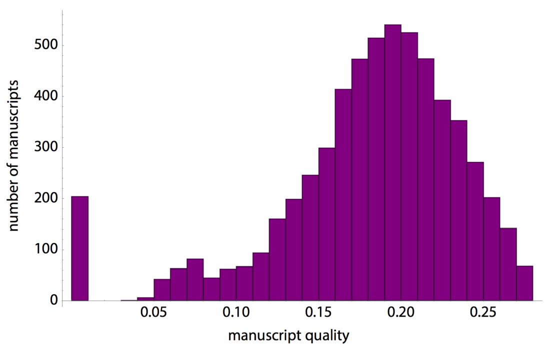

FIGURE 5. Quality of manuscripts published in the worst journal (idealized scenario)

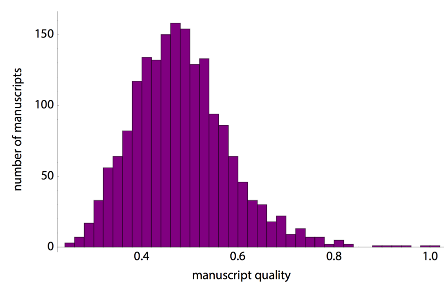

FIGURE 6. Quality of manuscripts published in the worst journal (noisy scenario)

3.4. **Empirical illustration of 3.1./3.2.** Many underlying assumptions in our model (e.g. the 'intrinsic quality of a paper/journal') have a platonic component that makes an empirical comparison difficult. However, in the case of our first two observations a rough comparison with empirical patterns is possible and carried out in Figure 5, which shows plots of Impact Factors for all Journals of three subject categories, for which a full series of data exists (plots on the left show the full population so obtained, while the plots on the right show the Impact Factor development for 9 journals ranked around the median journal). However, this comparison remains imperfect, since, as mentioned in our introduction, Impact Factors are a contested indicator of journal quality.

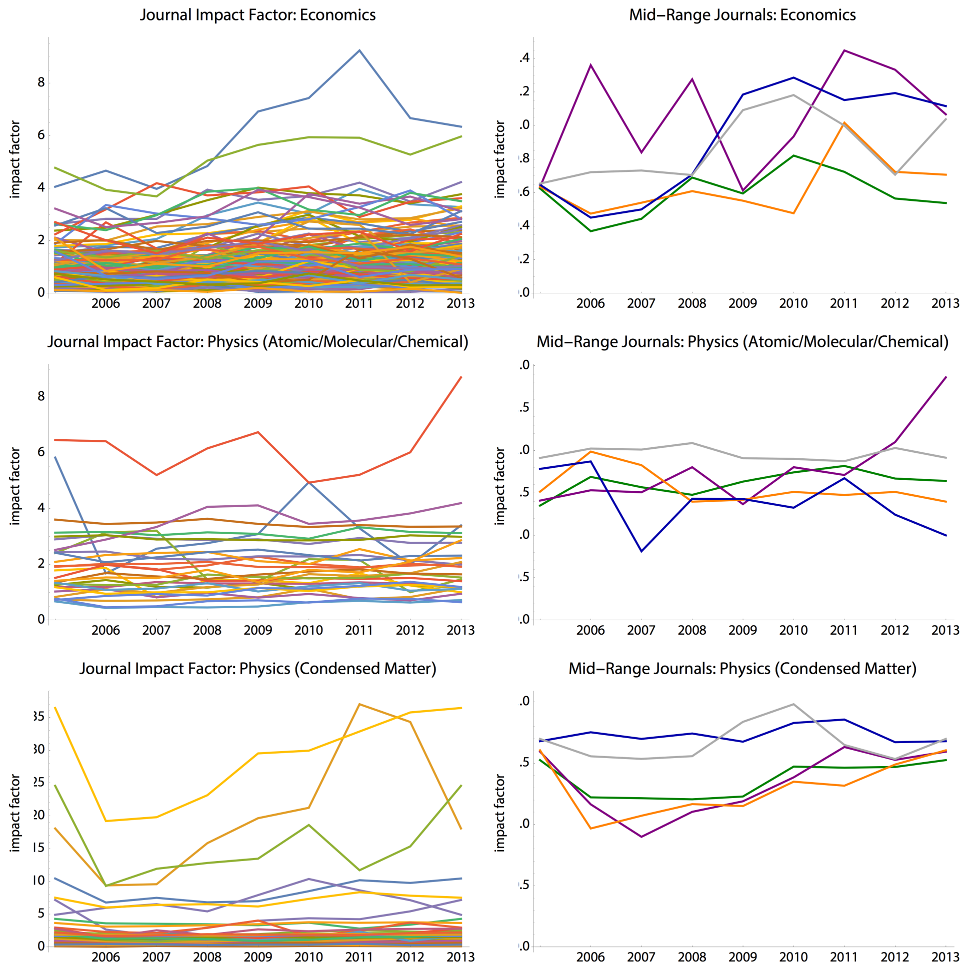

FIGURE 7. Impact Factors for journals in three subject areas (obtained from: *Web of Science*)

4. Universally emergent phenomena II: rejections and bottlenecks

4.1. **Rejection rate.** A very surprising phenomenon is the connection between rejection rate and quality of the paper (rejection is here understood in the sense that the paper remains unpublished after three attempted submissions). In the idealized scenario, we observe a high rejection rate for low-quality papers that sharply decreases after some minimal quality standards are met: the system works, low-quality papers are filtered out and research reaching a minimal standard of quality eventually gets published in a journal whose reputation closely mirrors the quality of the paper. However, much to our surprise, this pattern is not stable at all. Indeed, in the noisy scenario the clustering effect generates a lot of mediocre journals and the rejection rate for papers of less than average quality becomes negligible: there are a lot of appropriate journals to submit to and the chances of either one of them running out of space is small. We find that *the rejection probability is actually increasing as the quality of paper is increasing* until the quality of the paper reaches a very high quality at which point it starts to decay (because a 'bad' referee report would downgrade the paper from 'stellar and revolutionizing' to 'outstanding and important' in which case it is still likely to get accepted).

The NIPS experiment. One could argue that this paints a rather bleak picture of peer review as a way of judging scientific progress, however, a recent large-scale experiment had a comparably bleak outcome. NIPS (Neural Information Processing Systems) is one of the biggest and most prestigious conferences in Machine Learning (a field in computer science) and contributing to NIPS

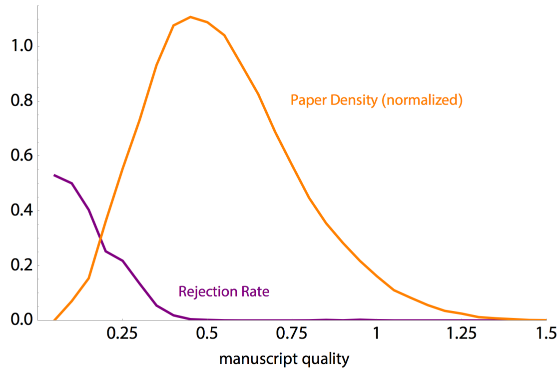

FIGURE 8. Paper Density and Corresponding Rejection Rates (idealized scenario).

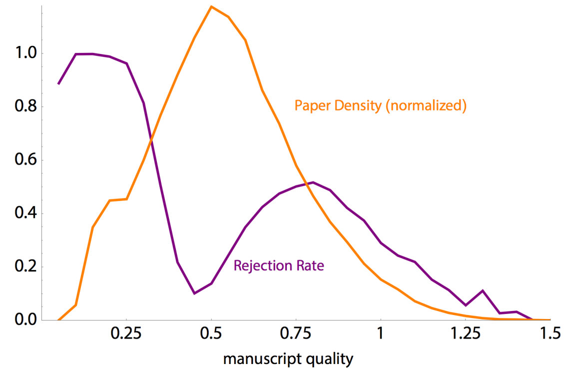

FIGURE 9. Paper Density and Corresponding Rejection Rates (noisy scenario).

is comparable in prestige to publication in an outstanding journal. Submissions are divided into two groups and two committees are tasked with deciding on acceptance/rejections. For NIPS 2014, (Cortes and Lawrence 2014) arranged for 10% of all submissions (a total of 166) to be reviewed by *both* committees. The drastic outcome was that a paper accepted by one committee had a likelihood of more than 50% of being rejected by the other committee. Given that the overall acceptance quota was 22.5%, a completely randomized decision would imply that an accepted paper would have a likelihood of 77.5% of being rejected by the other committee – the outcome is thus actually much more drastic than predicted by our model.

4.2. **Journals as Bottleneck.** It is clear that excellent journals publish excellent papers (because they can afford to be very picky); the converse fails and fails drastically: excellent papers get published in a wide variety of outlets or *do not get published at all.* As scientists' skills are distributed as $\mathcal{N}(0.5, 0.2)$ and the quality of papers are given by

$$q_j = \mathrm{sci}_j(1 + \mathcal{N}(0, 0.2)),$$

we have a good understanding of paper quality (in the sense that the probability distribution of the quality of papers could be explicitly computed and is, up to small errors, essentially Gaussian). In an ideal world, we would find a journal landscape that matches this distribution: for every paper there is a suitable outlet. However, as seen in examples above, this hardly ever occurs outside the idealized scenario. As a result, papers in, say, the top 20 percent but outside the top 5 percent face a profound lack of journals and often have to settle for lower-ranked journals. Conversely, papers at the lower end of the quality spectrum face an abundance of journals.

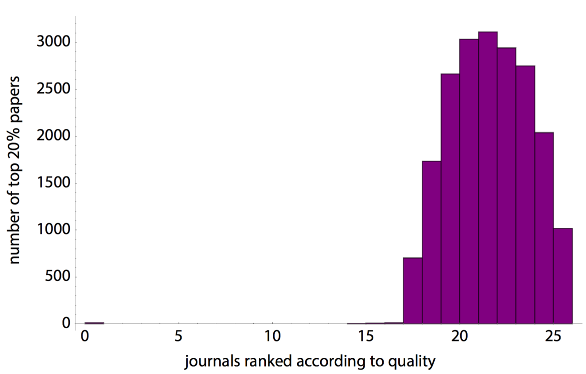

FIGURE 10. Where the best 20% of papers get published (idealized scenario).

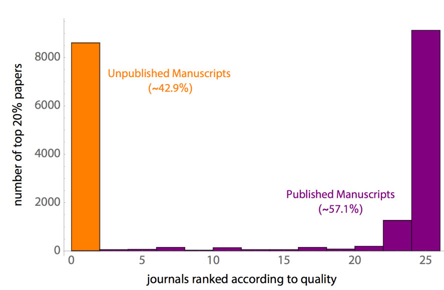

FIGURE 11. Where the best 20% of papers get published (noisy scenario).

Figures 10 and 11 show the trend: in the idealized scenario, most of the best 20% of all papers get published in the best 20% of journals. The noisy scenario is quite a bit more troubling: the papers still get published in top journals, however, the overall trend decreases the number of top journals and many of the papers (roughly $\sim 40\%$) do not get published at all. This dynamics acts as a bottleneck slowing down the publication process.

5. Robustness and stability: exploring variations of the model

The model, which assumes a variety of parameters, is remarkably robust. Our focus so far has been on simplicity of the model and there are various natural extensions that could be worth investigating (for example, the role of editorial desk-rejection, journals being able to dynamically adapt their publishing strategy, scientists undergoing a change in their skill level over time, multi-author papers, the formation of social networks, ...). The purpose of this section is to analyse some variations of the scenarios described above to assess to robustness of our model as well as the dependence of certain results on specific assumptions.

5.1. **Intensity of randomness.** So far, all random errors in the paper are given by $1+\mathcal{N}(0,0.2) = \mathcal{N}(1,0.2)$ – the main reason for this unified source of randomness is to avoid an unnecessary amount of additional variables that would obscure the actual model. Clearly, if the random variables is replaced by another one with a smaller variance, then the model moves towards the idealized scenario. However, that movement is gradual and not very sensitive to slight changes in the variance. The left frame of Figure 12 provides an example of journal reputation evolving when researchers are modest and both self-estimation and referee error are very tightly clustered around the true quality of papers with a multiplicative error behaving as $\mathcal{N}(1,0.05)$. We observe that the trend towards homogenization of journals happens at a slower rate but is equally inevitable.

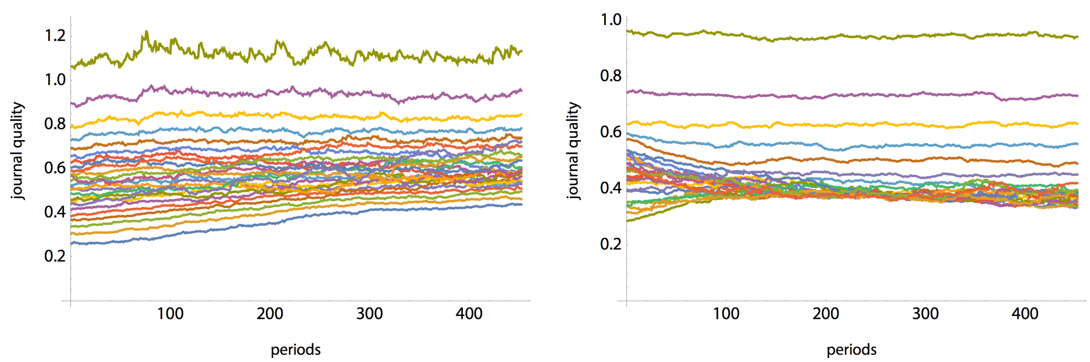

Figure 12. Evolution of journals with reduced self-estimation/referee error (left frame) and more conservative updates of journal quality (right frame).

In a similar vein the right frame of Figure 12 plots a variant of the noisy scenario with a much more conservative algorithm, when it comes to the update of journal quality. Specifically, it decreases the weight of the current issue from 20% to 5%, which leads to a tighter cluster in the middle but leaves our main observation of a noisy pattern intact.

5.2. **Eight variations.** There are three different sources of uncertainty in our basic model:

(1) scientists being unsure about the intrinsic value of their work ($\text{est}_j = \text{qual}_j(1+\mathcal{N}(0,2))$)
(2) scientists being unsure about where to submit their work (ambition) and
(3) referees being unsure about the quality of the submitted work ($\text{ref}_j = \text{qual}_j(1+\mathcal{N}(0,2))$)

The results hitherto described where obtained under the assumption of all three sources of randomness being present. In testing all eight possible combinations of the three factors being present or absent, we observe the following scenarios.

ERRORS IN ASSESSMENT	MODEST RESEARCHERS	AMBITIOUS RESEARCHERS
No Errors	The IDEALIZED Scenario	Journal patterns as in the IDEALIZED scenario, but with a downward trend, and no perverse effects with regard to rejection rates and bottlenecks
Only referees err	Behaves like IDEALIZED scenario	
Only authors err	Journal patterns as NOISY but no perverse effects with regard to rejection rates and bottlenecks	Behaves like the NOISY scenario
Both err		Noisy scenario

Figure 13. A landscape of possible scenarios.

The only feature of these results we have not yet encountered is that of a downard trend in journal quality with only a few journals surviving: this scenario consists of a drastic change in the journal landscape where most journals have their reputation tend to zero. Those with nonzero reputation have essentially nonzero variation in their reputation and are very different from each other. The main reason is that Ambition causes authors to submit a lot of papers to high-tier journals. Which fate occurs to a given journal depends on whether the journal has a good defense mechanism against large numbers of low-quality submissions. Curiously, this does not depend on having good referees! Indeed, the decisive factor is whether or not SELF-ESTIMATION occurs - if authors are uncertain about the quality of their work they can retain ambition without causing a decay within the journal-population. Uncertainty about the quality of one's work ensures that journals receive submissions of mixed quality instead of a bulk of very similar papers.

One of the entries in Table. 13 is easily explained: if authors have perfect understanding of the objective value of their work (no error in SELF-ESTIMATION) and submit to the appropriate journal (no AMBITION), then the model behaves as in the idealized scenario *independently* whether the referee has perfect or imperfect understanding. The reason is simply that all journals only receive submissions that are precisely at their level of reputation and any errors made by the referee are completely inconsequential because *all* submissions are at the same level of intrinsic value. Figure 13 highlights that different types of randomness have different impact on the behavior of the model. In detail:

(1) Ambition has the consequence that journals get many submissions that are below their level of reputation, which drives the observed perverse effects in the context of rejection rates and journals as bottlenecks (section 4). The phenomenon is less pronounced if scientists evaluate the value of their work with perfect accuracy, while in the case of self-estimation error journals are quickly overwhelmed with submissions and more likely to reject genuinely appropriate papers.
(2) Self-estimation not only aggravates the effects introduced by Ambition, but also creates the noisy patterns of journal quality observed and discussed in section 3. Hence, even in the absence of Ambition, errors in SELF-ESTIMATION are sufficient to partially collapse the idealized scenario.
(3) As a final and quite curious finding: the presence or absence of Referee-Estimation errors does not seem to have a profound structural impact, but rather may intensify or weaken the intensity of the unexpected outcomes observed in the noisy scenario.

6. Conclusion

In this paper we have demonstrated that in a perfect world, where estimates of scientific quality are always accurate and strategic behavior is absent, peer-review indeed is a viable tool suitable for objectively clustering academic research in outlets of different quality. However, we have also shown that even a *minimal* deviation from the idealized conditions drastically affects the outcomes of the academic production process. Already tiny misjudgements from authors/referees as well as minimal strategic considerations by authors lead to a clustering of journals and a high variability of quality among mediocre journals. Moreover, we observe a highly idiosyncratic development of the probability of rejection with respect to quality, where overall rejection rate increases with the quality of manuscripts for a majority of the population of scientific papers – this is tightly coupled to a mismatch between the quality of papers being produced and the distribution of the quality of journals. These results suggest a new dimension to the traditional criticism of practises in academic publishing: even in the absence of human fallibility, 'hot' topics, the pressure of grants and tenure, underlying phenomena emerge for purely mathematical reasons and are potentially harmful to the scientific process at large.

Acknowledgment. J.K. and S.S. were both supported by INET Grant #INO15-00038. S.S. was supported by an AMS-Simons Travel Grant and a Yale Provost Travel Grant. The authors gratefully acknowledge constructive criticism made by two anonymous referees.

References

[1] Amin, M., and Mabe, M. (2000). Impact Factors: Use and Abuse. Perspectives in Publishing, October 2000, 16. Bedeian, A. (2003). The manuscript review process: The proper roles of authors, referees, and editors. Journal of Management Inquiry, 12, 331-338.
[2] Binning, G., Rohrer, H., Gerber, Ch. and Weibel (1982), Applied Physics Letters 40, 178-180.
[3] Cortes, C., and Lawrence, N. (2015), The NIPS Experiment, http://inverseprobability.com/2014/12/16/the-nips-experiment,
[4] Day, T. E. (2015), The big consequences of small biases: A simulation of peer review, Research Policy 44 (6), 1266-1270.
[5] Gans, J., and Shepherd, G. (1994). How are the mighty fallen: Rejected classic articles by leading economists. Journal of Economic Perspectives, 8(1), 165-179.
[6] Higgs, PW (1966). Spontaneous symmetry breakdown without massless bosons. Physical Review, 145(4), 1156.
[7] Kapeller, J. (2010). Citation Metrics: Serious drawbacks, perverse incentives and strategic options for heterodox economics. American Journal of Economics and Sociology, 69(5), 1376-1408.
[8] Macdonald, S., and Kam, J. (2010). Counting footnotes: Citability in management studies. Scandinavian Journal of Management, 26, 189-203.
[9] Reedijk, J., and Moed, H. F. (2006). Is the Impact of Journal Impact Factors Decreasing? Journal of Documentation, 64, 183192.
[10] Solla-Price, D.J. (1965): Networks of Scientific Papers. Science, 149, 510-515.

(Jakob Kapeller) Department of Philosophy and Theory of Science, University of Linz, Altenbergerstrasse 69, Linz, Austria
E-mail address: jakob.kapeller@jku.at

(Stefan Steinerberger) Department of Mathematics, Yale University, 10 Hillhouse Avenue, New Haven, CT 06511, USA
E-mail address, Corresponding author: stefan.steinerberger@yale.edu