Efficiency in rewarding academic journal publications.
The case of Poland

Wojciech Charemza[*] Michał Lewandowski[†] Łukasz Woźny[‡]

[*] Faculty of Business and International Relations, Vistula University, Warsaw, Poland

[†] Institute of Econometrics, Warsaw School of Economics, Warsaw, Poland

[‡] Department of Quantitative Economics, Warsaw School of Economics, Warsaw, Poland

February 2021

ACKNOWLEDGEMENT

Financial support of the Ministry of Science and Higher Education of Poland for the Vistula University is gratefully acknowledged. We thank Jędrzej Siciński and Tomasz Jurkiewicz from the University of Gdańsk for their help in data collection, Jason Jones from Aptech Systems Inc for his substantial programming tips, Svetlana Makarova and the participants of the CFE 2020 Conference for their helpful comments and suggestions. We are responsible for all the deficiencies in this work.

Abstract

We consider the efficiency of a mechanism for incentivising publication in academic journals where a research supervisory body awards points for papers that appear in quality publications. Building on the principal-agent literature with hidden types, we assume that such a body wants to maximise the expected prestige of academic disciplines. It sets up a reward system so that researchers who are aiming to maximise their own rewards also maximise the objective function of the research supervisory body, through their submission decisions. The model is calibrated to the reward scheme introduced within the Polish higher education reform in 2018, for which a series of policy recommendations is given.

Keywords: academic publications; efficient mechanisms; optimal categorisation

JEL classification: (possible:)

I23: Higher education, research institutions

C55: large data sets

O31: Innovation and inventions: Processes and incentives

C53: Forecasting and prediction models: simulation methods

..

Highlights:

- Researchers, in their publication strategy, aim at maximisation of expected rewards
- The centralised system of rewards aims at maximisation of academic prestige of a country
- Rewards system introduced in Poland in 2018 is losing some efficiency mainly due to breaks in the monotonicity between rewards and journals' prestige
- Another source of inefficiency is overambition of induced allocation, at the expense of allocation to middle-ranking journals

1 Introduction and motivation

A natural, though a controversial, way of measuring a published academic article's quality is to examine its citations and sometimes, for applied research, to check whether its findings have been implemented in practice. Such ex-post evaluation is impractical for many purposes, as it can often take several years of waiting before citations appear, or practical effects of implementation are visible. The practicalities of organising research and evaluating the authors' career prospects require a system of ex-ante assessment of the quality of academic publications rather than ex-post assessment. Ideally, such a system should give an immediate evaluation of the possible quality of an academic article when it is published. There have been numerous attempts, mainly informal or semi-official, to introduce such assessments, normalise and formalise them. An appropriate system should have at least some rules for aggregation so that an appointment or promotion board could compare candidates against each other (for a study for the US and Canada see McKiernan et al., 2019). Various informal and clandestine systems are widespread in many countries and academic institutions. The evaluation usually assesses the quality or prestige of the academic journals in which the papers are published. This prestige is usually measured using an index for how frequently the published articles are cited, like CiteScore, Journal Impact Factor, SNIP, SJR and others. In this paper, we will refer to such measures as JIMs (Journal Impact Measures). As it is difficult to compare thousands of different journals, they are often graded into ranks (classes) or awarded 'stars'. This is common in many countries and at numerous universities. It is usually done informally or implicitly, sometimes by individual universities and in a few cases by entire higher education systems. Even where this practice is officially denied and criticised (see, e.g. Stephan, 2012), it is frequently used as an informal base for promotions or job interviews, or in similar ways. However, the criteria are often vague or may not even be disclosed, though attempts have been made to predict the grades given to journals, for example by Hudson (2013).

Rankings of journals are already used informally in many countries and for many disciplines, but their unofficial and sometimes semi-clandestine nature makes it quite hard to get clear and accurate information about them. To the best of our knowledge, the first complex and transparent case where such a reward scheme was explicitly defined and officially introduced for all academic disciplines and all universities was the Polish reform of higher education that was passed in 2018 and implemented in 2019. In the official legal act, the Polish Ministry of Science and Higher Education published a list of over 30,000 academic journals for 44 academic disciplines, with journals grouped in six categories and identical points awarded to journals in each class. Universities funding of research depends, in a nonlinear way, on the number of points their research accumulates. Consequently, most Polish universities offer substantial bonuses to their academic staff for publishing in highly favoured journals.

It is interesting to note that attempts to centralise allocation of papers to journals by establishing a research supervisory body awarding bonuses for publications are particularly popular in countries where researchers' internal mobility is relatively low, like Russia, Poland, Romania, Turkey and some other countries. Countries where the market for researchers resembles a contestable market in its ease of entry, low cost of mobility, open competition for academic jobs, and high or unregulated salaries do not use such mechanism, at least not explicitly (see MacLeod and Urquiola, 2021b). However, for countries where there is little mobility, or academic salaries are inflexible, the practice of grading journals can be a reasonable way of rewarding researchers, while also increasing the academic reputation of the country or the university.

Given this background, we construct an allocation model that we subsequently apply to evaluate the rationale and efficiency of this scheme. The paper has four specific goals, which are (*i*) to explain the rationale that researchers apply in deciding to send their papers to journals of different qualities where the chances of acceptance are different; (*ii*) to evaluate how efficient the allocation strategy that the reward system encourages is; (*iii*) to identify and quantify the possible distortionary effects of the reward scheme; and (*iv*) to deliver policy conclusions.

The reward allocation system applied in the model is loosely based on the principal-agent idea (see MacLeod and Urquiola, 2021a). Agents in such a mechanism, who in this case are the researchers, aim to maximise their rewards here in the form of points for their publications. The principal, which is the Research Supervisory Body (RSB), sets up the reward (points) system. The system is constructed to maximise the academic prestige of the organisation or the country through their publications.

After the introduction, the paper briefly explains the institutional framework of the Polish higher education reform in Section 2 and describes the database. The general assumptions of the model are given in Section 3. In Section 4 we discuss the settings for the simulation experiment that is designed to evaluate the allocative efficiency versus automatic ranking, which uses the quantiles of a series of journals ranked by their prestige as the benchmark. Section 5 concludes and provides some policy recommendations. The Appendix contains the aggregate results from evaluating the efficiency of allocation under the Polish ministerial scheme for 44 academic disciplines.

2 Outline of the publication reward system in Poland

The new higher education law in Poland was passed in 2018 and implemented from 2019. The legal act stated that the criteria for the reward scheme were to improve (*i*) the prestige of research conducted at Polish academic institutions, (*ii*) the financial benefits of the research, and (*iii*) the social benefits of the research. These criteria indirectly imply that the objective role of the RSB might not necessarily be solely to maximise the prestige of particular academic disciplines; nevertheless in this paper we focus on this aim.

The document specifies that for each of the 44 disciplines, publication in an academic journal is to be rewarded by ministerial points (MPs) that will depend on the quality of the journal. The rewards will be of 200, 140, 100, 70, 40 and 20 points. A separate system of ministerial points was created to rank the quality of monographs, chapters in edited volumes and, for some disciplines, conference presentations. These additional rewards are ignored here, and we concentrate solely on publication in academic journals. For the approach based on the conference presentations see, e.g. Gorodnichenko et al. (2021).

The initial stage of setting up the points system was to rank the admissible journals in the preliminary lists from the highest to the lowest using one or more of five journal impact measures (JIMs): CNCI, JIF, CiteScore, SJR and SNIP. The journals were then divided into six quantile categories, with the threshold quantiles of 97%, 90%, 75%, 50% and 25%.This was done separately for each discipline. Which JIM to use for classifying the journals in each discipline was not specified, but was left to panels of experts. This initial list contains some local journals that are not classified by any of the five JIMs, and these journals are called here the zero-JIM journals. Within this initial classification, there is a one-to-one correspondence between the JIM and the journal's rank on the list, with zero-JIM journals placed at the bottom.

However, in the next stage, the monotonicity of the correspondence between the rank and the JIM is broken. This happened firstly because of the principle of homogenising the rewards for interdisciplinary journals that are claimed by more than one discipline. As the overall level of the JIM is likely to be different for different disciplines, the same journal might, be ranked differently in the initial classification by different disciplines and assigned different numbers of points. As the number of points must be strictly on the 200/140/100/70/40/20 scale and must be unique for each journal, an algorithm was applied to decide on the category of the multidisciplinary journals. This means the reform assumes that the prestige of all journals, measured by their JIMs, is identical for all disciplines. We call this the assumption of *homogeneous prestige*, in contrast to *heterogeneous prestige*, which is where journals might have different rewards for different disciplines.

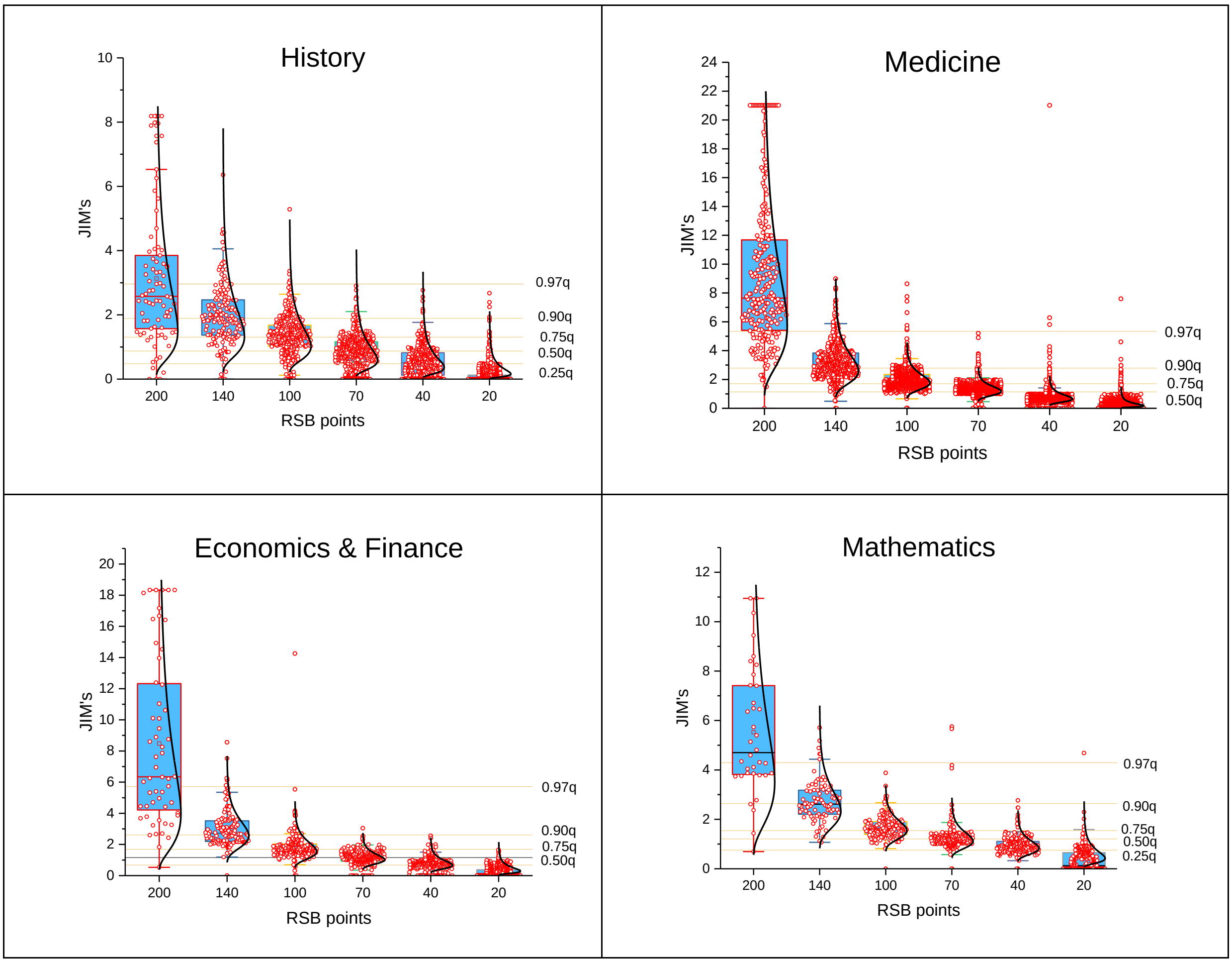
Some non-graded zero-JIM journals, mainly of local circulation and reputation, have also been added to the list, while some lower-graded journals have been upgraded and, it seems, some higher-graded journals have been downgraded. The reasons for this are not clear, but there are five possible additional explanations. Firstly, it might be the result of efforts to meet objectives (*ii*) and (*iii*) of the reform by increasing the financial and social effects of the research. Secondly, it might come from a conscious effect to promote some new journals expecting that their reputation and citability will increase in the future. Thirdly, it might be a consequence of realising that the initial quantile scale was too ambitious or not ambitious enough, at least for some disciplines, and so was ineffective at stimulating the research. Fourthly, there might be mistakes in the classification made by the experts by considering journals with false JIMs, including predatory, or semi-predatory journals for example. Finally, it might be that local interests were considered, such as promoting journals that are popular with local authors or that have Poles on their editorial boards. The final total number of journals for which points are awarded in the 44 disciplines is 32,323.

We present synthetic results for all 44 disciplines, but we concentrate on four representative disciplines: history, medicine, economics with finance, and mathematics. Breaks in the JIM-rank monotonicity for these disciplines are illustrated in Figure 1. The box-plots shown in this figure are centred around a median, and each box gives the 0.25-0.75 interquartile range.

In this paper, we define JIMs as the maximum of all the available journal impact measures and their variations: CNCI, JIF, CiteScore, SJR and SNIP. We consider all the variants of these measures published between 2015-2018, giving one and five-year indices. Overall we consider 32 different measures. Each measure is then scaled by its mean for comparability. We realise that this might be controversial, as each journal impact measure has its own advantages and disadvantages (see, e.g. Stern, 2013, Callaway, 2016, Bertoli-Barsotti and Lando, 2017 and, particularly, Bornmann and Wohlrabe, 2019, who exensively evaluated various ways of standardising JIM's) but the way we do it seems to be the best compromise. After all, this is vaguely similar to how the Polish Ministry of Science and Higher Education experts assessed the quality of journals).

Figure 1 indicates frequent breaks of monotonicity, which go both ways, as some journals with low JIMs are re-classified to a higher class than in the quantile classification, while some journals are re-classified downwards. Although it is possible that there is an occasional mistake in our data, these should be rare enough to allow us to conclude that the break in the monotonicity of the rewards is evident.

Figure 1: Distributions of JIMs by ministerial points



Legend: Data published by the Ministry of Higher Education of Poland on 18 December 2019, matched with JIMs obtained from various sources. The level of some quantiles of the ordered JIMs series is marked by horizontal lines.

Table 1 gives the descriptive characteristics of the journals selected for each discipline. It shows substantial differences between the disciplines, both in the number of journals available and in the assessment criteria used. Particularly notable is that the journals selected for history have the lowest JIMs of the four disciplines and at the same time, most journals with no impact measure. Medicine has the greatest number of journals, the highest JIMs, and at the same time, the lowest proportion of journals without any impact measure. Economics and mathematics show similar characteristics, with a higher number of zero-JIM journals for economics and finance.

Table 1: General information on journals selected for four disciplines

	History	Medicine	Econ & Fin	Mathematics
Number of journals	3,027	8,659	1,755	1,224
Number of zero-JIMs	1,012	950	477	218
Average relative JIM	0.59	1.09	0.96	0.95
Number of disciplines/ number of journals	3.5	4.6	3.5	3.4

Legend: *Number of journals* is the total number of journals claimed by each discipline; *Number of zero-JIMs* is the number of journals claimed by each discipline that are not in any of the following databases: CNCI, JIF, CiteScore, SJR SNIP, or appear there with zero impact measures*; Average relative JIM* is the average of the maximum score for each journal of all the available JIMs from the databases, where each JIM is divided by its average across the disciplines. The last row gives the average number of disciplines claiming the same journal.

3 The model

3.1 Main assumptions

The principal assumptions of the allocation model applied here are the following:

A1. In each discipline, there is a set J of journals.Each journal j in this set has an intrinsic and exogenous probability of accepting an article. These probabilities are unconditional in the sense that they do not depend on the author's abilities. They are understood in a rather wide sense and include self-rejections, which are non-submissions by researchers because of critical evaluations of their own work. These probabilities, denoted p_j , $j=1,2,...,\bar{j}$, are ordered from the hardest to the easiest journal to reach, that is $0<p_1<p_2<...<p_{\bar{j}}<1$.

A2. The prestige of journals, measured by JIMs and denoted $\phi_j>0$, is a monotonous function of the probabilities of acceptance, where the lower p_j is, the higher ϕ_j is. That is, ϕ_j increases as the probability of acceptance decreases, so that $\phi_1 \geq \phi_2 \geq ... \geq \phi_{\bar{j}-1} \geq \phi_{\bar{j}} > 0$.

A3. The RSB groups journals into K classes and awards points r_k , $k=1,2,...,K$, for publication in each class. The rewards are in descending order so that $r_1 > r_2 > ... > r_K > 0$.

The set of indices of the journals that fall into class k are denoted J_k. It is assumed that: $J_k \cap J_m = \varnothing$ for any $k \neq m$, $1 \leq k, m \leq K$; and $\bigcup_{k=1}^{K} J_k$ consists of all the journals assigned to the discipline. Note that there might not be monotonicity with respect to p_j in the RSB classification, that is, it might happen that $\max_{j \in J_k}\{p_j\} > \min_{j \in J_m}\{p_j\}$ for $m > k$. In other words, the RSB might put a journal with lower prestige in a higher category than a journal with higher prestige. However, if the quantile classification is adopted so that journals in the highest quantile group get the highest reward, such monotonicity is preserved.

A4. In each discipline there are N researchers, each intending to maximise their reward from publishing one, single-authored, paper and submitting it to a journal once.

A5. The researchers have different abilities or types. The abilities a_i, $i = 1, 2, ..., N$, are distributed between 0 and 1, with 1 as the highest ability.

A6. Own abilities are known to the researchers, but not to the RSB.

A7. Each researcher knows p_j in a set of potential target journals G_i and is able to formulate the probability, conditional on their ability, of acceptance to the j th journal, denoted $p(a_i, p_j)$, monotonously increasing in a_i and p_j, and such that:

For each a_i and p_j, it holds that $p(0, p_j) = 0$, $p(a_i, 0) = 0$; $p(1, p_j) = 1$, $p(a_i, 1) = 1$;

Further in the text, we consider the case where the researcher does not know p_j, but only its subjective probability weight.

A8. Each researcher decides to submit the paper to the journal which, given the ability a_i, maximises the expected reward. The researcher i aims to submit to a journal indexed by λ_i, which will allow the researcher to achieve the maximum expected reward R_i, that is:

$$R_i = \max_{j \in G_i} \left\{ r_{k(j)} p(a_{i,} p_j) \right\} \quad , \tag{1}$$

where $k(j)$ denotes the category assigned to journal j and:

$$\lambda_i = \arg\max_{j \in G_i} \left\{ r_{k(j)} p(a_{i,} p_j) \right\} \quad .$$

It is worth noting that in practice the maximum number of journals for which the conditional probabilities of acceptance have to be computed in the in the maximisation problem above is equal to K. If more than one journal of interest is selected from each class, the researcher, by the rationality principle, will prefer the one where the probability of acceptance is the greatest. This means it would not be rational to consider submitting to a journal with a lower probability of acceptance and equal reward, so such journals can be dismissed outright without entering (1).

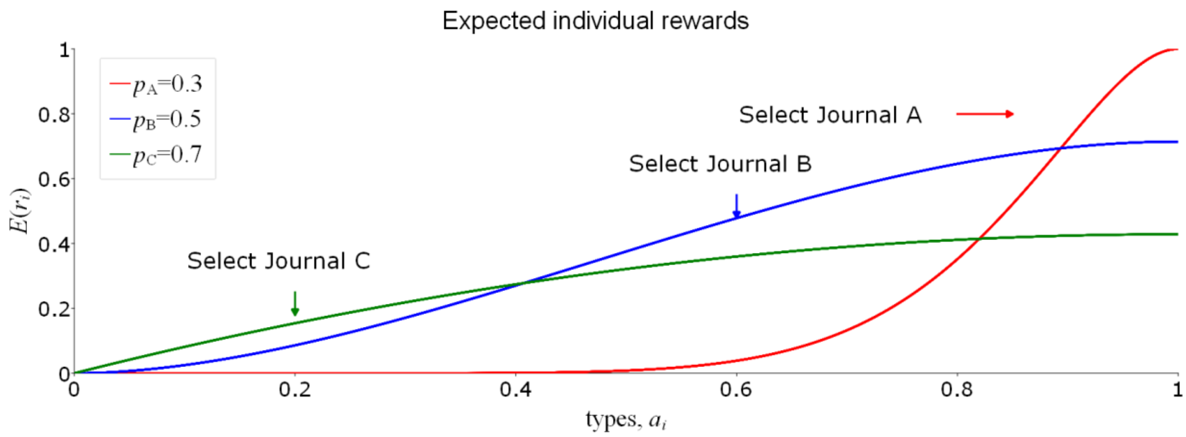
A9. The *RSB* (the Ministry) wants to maximise the total expected prestige of the discipline, Ψ , which is a sum of the expected prestiges of individual researchers, Ψ_i:

$$\Psi = \sum_i \Psi_i = \sum_i \phi_{\lambda_i} p(a_i, p_{\lambda_i}) \qquad (2)$$

This can be aimed by the RSB by deciding on the split of journals into classes k and setting the rewards r_k for each class. Evidently, the first-best solution, that is, setting individual rewards for each researcher, that so that $r_{\lambda_i} = \phi_{\lambda_i}$ is not achievable, as this would setting thousands of classes, and then changing them frequently. It would also be open to manipulation.

This set of assumptions is generally in line with earielr empirical findings regarding the strategy of allocations of academic papers to journals by the researchers (see, e.g. Śpiewanowski and Talavera, 2020). An illustration of the maximisation decision in (1) is given in Figure 2. It plots $E(r_i) = r_{k(j)} p(a_i, p_j)$ against the abilities of researchers normalised between 0 and 1 in the case where there are three journals A, B and C, each with different probabilities of acceptance so that $p_A = 0.3$, $p_B = 0.5$ and $p_C = 0.7$. The rewards r_k are equal to 1 for publication in A, 0.714 for publication in B, and 0.429 for publication in C. That means the rewards are equal to the probabilities of rejection scaled by the highest of them, that is $(1-p_\kappa)/(1-p_A)$, where $\kappa = \{A, B, C\}$. Selecting C gives the maximum expected reward for researchers whose ability is below 0.41 and selecting B for researchers whose ability is between 0.41 and 0.89, while researchers whose ability is above 0.89 can maximise the expected reward by aiming for A.

Figure 2: Illustration of the allocation scheme in Step 2



3.2. Calibration and simulation assumptions

The model given in Section 3.1 can be evaluated under various, more specific, assumptions. We can examine the effect of assuming different distributions of abilities, or formulating the conditional probabilities of acceptance in different ways, or we may perhaps apply different efficiency criteria. We limit our interest to how breaking the monotonicity of the relation between the rewards and JIM's ranking affects the prestige of research conducted at the academic institutions.

For our case study: we use information from the list of over 30,000 journals and the reward points assigned to each of them, as published on 18 December 2019 by the Ministry of Higher Education of Poland.[1] It also gives information on how the papers are divided between the 44 disciplines. We matched each journal with their JIMs as described in Section 2 above. We assume (see Assumption A2) that the JIM of each journal measures its prestige, and the ministerial classification provides information on J_k and also on which journals are assigned to each discipline.

To calibrate the model above, further assumptions need to be made on:

1. The number of researchers and the distribution of their abilities or types, that is, a_i;
2. The unconditional and conditional probabilities of acceptance p_j and $p(a_i, p_j)$.
3. How each researcher selects the papers of interest in each category; that is, defining G_i.

We assume that for each discipline, there are $N = 1,000$ researchers, so that $i = 1, 2, ...1,000$. As our aim is not to evaluate the strengths or weaknesses of particular disciplines, but only the rationale of the reward scheme that is applied there, we assume that the distribution of the abilities or types of the researchers is non-informative, meaning they are uniformly distributed on $[0,1]$. In the robustness analysis outlined later, we also used different types of distribution for the abilities.

The unconditional probabilities of acceptance by journals are not usually known. There are some publications that give the empirical results from estimating this probability (e.g. Cherkasin et al., 2009), but the results are fragmentary, difficult to generalise and based on inside information that is not easy to obtain. Moreover, the concept of the acceptance rate, which is usually seen as the base for computing the probability of rejection, is often seen differently by different editors. Sometimes desk rejections are excluded, or papers that are withdrawn may be counted as rejected.

Acceptance rates are published for a small selection of journals, but even then it is not known whether those figures include desk rejections, or revise and resubmit requests. Moreover, the best journals in some disciplines might perversely have relatively high frequencies of acceptance if authors base their decision to submit on a realistic evaluation of their own abilities. To deal with this, we compute the probabilities using JIMs. We define the probabilities of acceptance as the inverses of the JIMs rescaled within the interval $(0.01, 0.99)$.

We simulate the model under the assumptions of the heterogeneous and the homogeneous prestige. Under heterogeneity, the probabilities are computed separately for each discipline. This means that if a multidisciplinary journal is claimed by more than one discipline, it might have different probabilities of acceptance. Under homogeneity, the probabilities are computed

[1]See https://www.bip.nauka.gov.pl/inne2/komunikat-ministra-nauki-i-szkolnictwa-wyzszego-z-dnia-18-grudnia-2019-r-w-sprawie-wykazu-czasopism-naukowych-i-recenzowanych-materialow-z-konferencji-miedzynarodowych.html

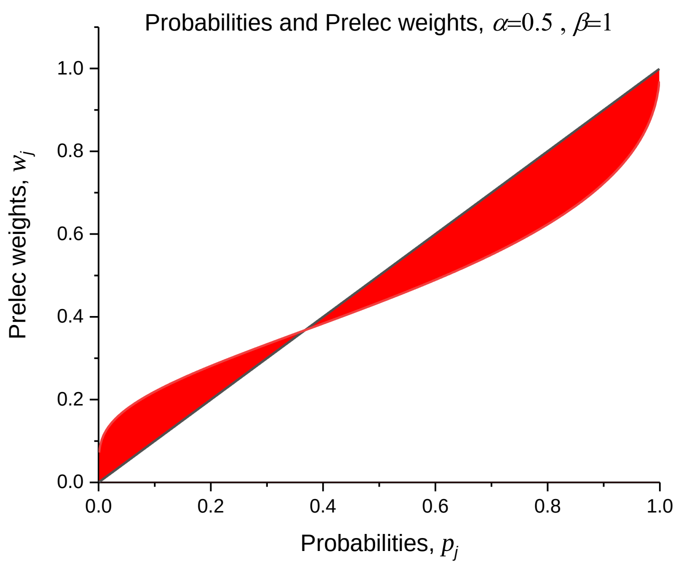
for a series of journals constructed jointly for all 44 disciplines. The Polish reward system applies the homogeneity.

One practical problem is that there is a large number of journals with zero JIMs, some of which enter highly prized RSB classes (see Table 1 and Figure 1 for the disciplines selected). Under the assumptions used, the probability of acceptance for such journals would be 0.99. As some of these journals are highly prized, we relax Assumption A7 for the results shown in this paper, and we assume instead, for the results discussed here, that the researchers do not actually observe the true probabilities of acceptance p_j, but rather their subjective probability weights w_j (see Prelec, 1998; al-Nowaihi and Dhami, 2006), defined as:

$$w_j = \exp(-\beta \ln(p_j)^{\alpha}) \quad \text{, where } \alpha > 0 \text{ , } \beta > 0 \text{ .}$$

Note that if $\alpha = \beta = 1$, $w_j = p_j$. Figure 3 shows the Prelec function for $\beta = 1$ and $\alpha = 0.5$.

Figure 3: Illustration of the Prelec probability function



In this weighting function, low probabilities are overestimated and high probabilities underestimated. We apply the Prelec function as in Figure 3, where $\beta = 1$ and $\alpha = 0.5$, by substituting w_j for p_j in (1).

The function used for modelling the conditional probabilities of acceptance $p(a_i, p_j)$, or $p(a_i, w_j)$ if the Prelec probability weights are used in (1), must fulfil the condition imposed by Assumption A7 that is, be monotonically increasing to 1 with the increase in a_i and p_j, and decreasing to 0 with a decrease in it. With this in mind, we decide to apply the *flexible sigmoid function* (FSF) of Yin et al. (2003), given by:

$$p(a_i, p_j) = p(a_i, p_j; \mu_{i,j}) = \begin{cases} \left(1 + \frac{1-a_i}{1-\mu_{i,j}}\right) a_i^{\frac{1}{1-p_j}} & \text{for } 0 \le a_i < 1 \\ 0 & \text{otherwise} \end{cases}$$

which requires the inflexion point of the sigmoid to be defined by setting the parameter $\mu_{i,j}$. We set the parameter $\mu_{i,j}$ in such a way that at the inflexion point, the ability of the i th researcher is equal to the intrinsic probability of rejection $1-p_j$. Under this assumption, the values of $\mu_{i,j}$ are obtained numerically by applying the dynamic search algorithm; see, e.g. Devroye (1986).

Finally, we have to decide on how the sets of potential target journals for each researcher, the G_i's, are defined. As we do not observe sub-disciplines, we simulate the sets of the potential target journals by drawing randomly without replacement for each i, and each k. Again, this selection has been made in various sizes as we examine the robustness of the model. In the results presented here, we assume that the number of journals drawn from each class is equal to the square root of 20% of the total number of journals in this category, rounded upwards. The number of drawings, $nrepl$, for which we present results here is 250.

A simple pseudo-code for this simulation is the following:

```
for i = 1 to N
      for repl = 1 to nrepl
            for k = 1 to K
                  draw G_{i,k}^{repl} from J_k
            endfor
            G_i^{repl} = ∪_k G_{i,k}^{repl}
            for j ∈ G_i^{repl}
                  compute p(a_i, p_j)
            endfor
            maximise (1)
      endfor
endfor
```

These drawings are the principal source of randomness in our simulations. A minor source is the rare case of multivalued argmax in (1). If this happens, we assume risk neutrality for the researchers and select the journal for submission randomly.

Table 2 summarises the main simulation settings used to obtain the results described in Section 4 and in the series of robustness experiments. The model is most sensitive to changes in the number of papers that are allowed for G_i. For example $|G_i| = 1$ implies a deterministic selection, while $|G_i| = \overline{J}$, that is, to a total number of journals for the discipline, leads to to a haphazard allocation. The model is also sensitive to the assumption of homogeneity or heterogeneity for the prestige of the journals. This is discussed further in Section 4.

Table 2: Some different settings used for the robustness analysis

Settings	**Sensitivity to a change**
With and without Prelec weighting	moderate
Different ways of computing p_j : 1. Heterogeneity and homogeneity of the prestige of journals 2. Excluding some JIMs from the set of 32 impact criteria	 Substantial (see Section 4) minor (for most disciplines)
Different settings for the number of papers in G_i	depends on the magnitude
Assuming beta distribution for abilities rather than uniform	minor
Changing the inflexion point in FSF	minor
Increasing the number of replications above 250	minor
Using logarithmically weighted sampling in place of linear sampling	minor (slows down computations)

4 Evaluating the efficiency of the Polish classification

We compare the expected prestige of each discipline obtained by allocating papers to journals under the ministerial (RSB) non-monotonical scheme with the expected prestige which would have been obtained if the original monotonical quantile allocation scheme was applied. We modify the original 0.97|0.90|0.75|0.50|0.25 quantile division slightly, as we classify all the zero-JIM journals on the ministerial list in the lowest category.

We use two aggregate measures of the allocation efficiency of the reward scheme. The first is the average prestige ratio, APR, which is defined as:

$$APR = \frac{1}{nrepl} \sum_{repl} \left(\sum_{i=1}^{N} R_i^{(RSB)} / \sum_{i=1}^{N} R_i^{(quantile)} \right) \quad ,$$

where $R_i^{(RSB)}$ is the sum of the expected prestige obtained in each replication from (1) for the allocation under the RSB allocation scheme, $R_i^{(quantile)}$ is such sum under the quantile allocation scheme and $nrepl$ is the number of replications that is, number of different drawings of G_i . It should be noted that the applied quantile allocation might not be optimal for the RSB in the sense that a different quantile division with the same number of categories might maximise the prestige. This might be particularly relevant if the distribution of abilities is not uniform. This is, however, not discussed in this paper.

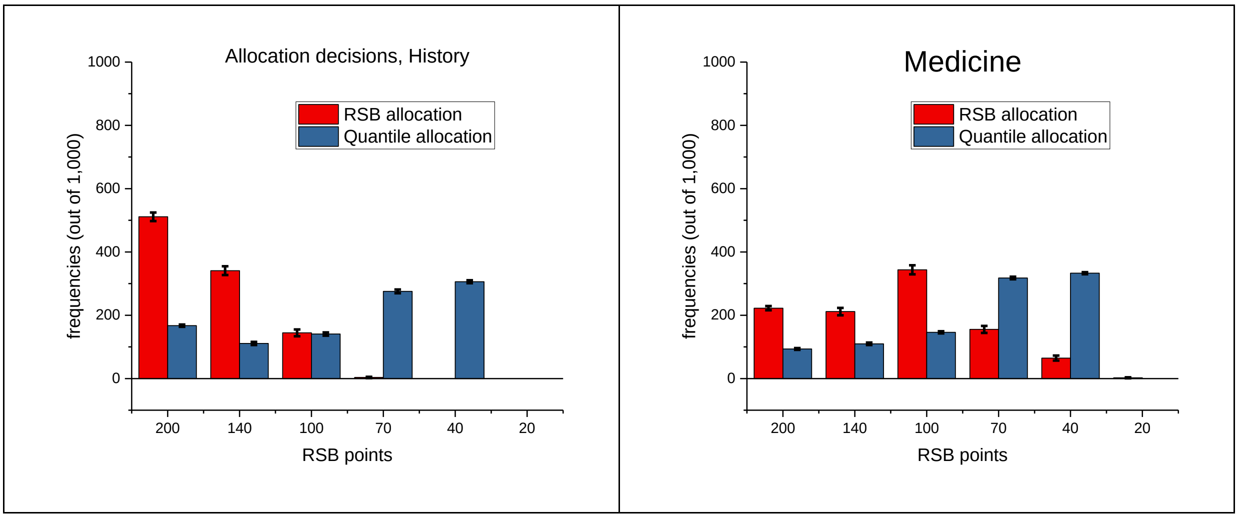
The second measure is the average misallocation coefficient, AMC, which is the average across the replications of the magnitude of misallocations in the decisions on publication taken under the ministerial rewards in comparison with that for the quantile rewards. The AMC is computed as:

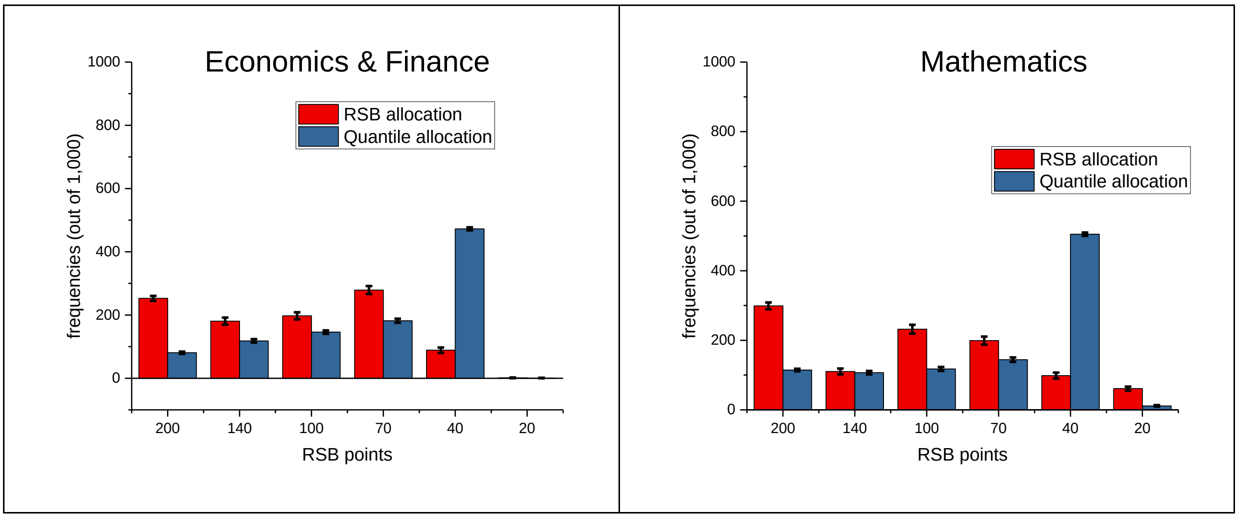
$$AMC = \frac{1}{nrepl} \sum_{nrepl} \sum_{k=1}^{K} \left(\frac{| A_k^{(RSB)} - A_k^{(quantile)} |}{2N} \right) ,$$

where $A_k^{(RSB)}$ and $A_k^{(quantile)}$ are the numbers of allocation decisions targeting journal from class k when the decisions are made under each allocation scheme. That is, separately for each class k, we count the number of cases in which $\lambda_i = j$ for some $j \in J_k$, where λ_i solves (1). Evidently, $\sum_{k=1}^{K} A_k^{(RSB)} = \sum_{k=1}^{K} A_k^{(quantile)} = N$.

Figure 4 shows the means as bars, and the standard deviations as ranges, plotted at the upper end of each bar, of the frequency of allocation decisions by 1,000 researchers in 250 simulations for the four selected disciplines of history, medicine, economics and mathematics. This means that 250,000 optimisations of (1) were made for each discipline. The red bars denote allocation following the ministerial (RSB) points, and the blue bars are for the quantile allocation. It is striking that the allocation according to the ministerial points, is much more 'single-minded' for all four disciplines, as the red RSB bars for the highly rewarded categories are overwhelmingly higher than the blue quantile bars.

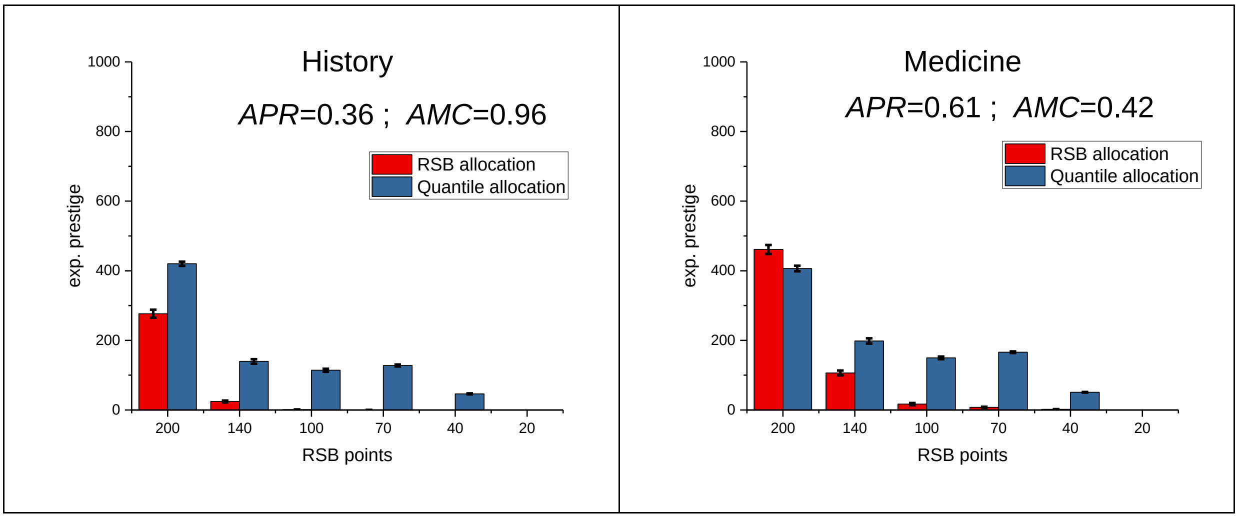
Figure 4: Means and standard deviations of the frequencies of allocation decisions

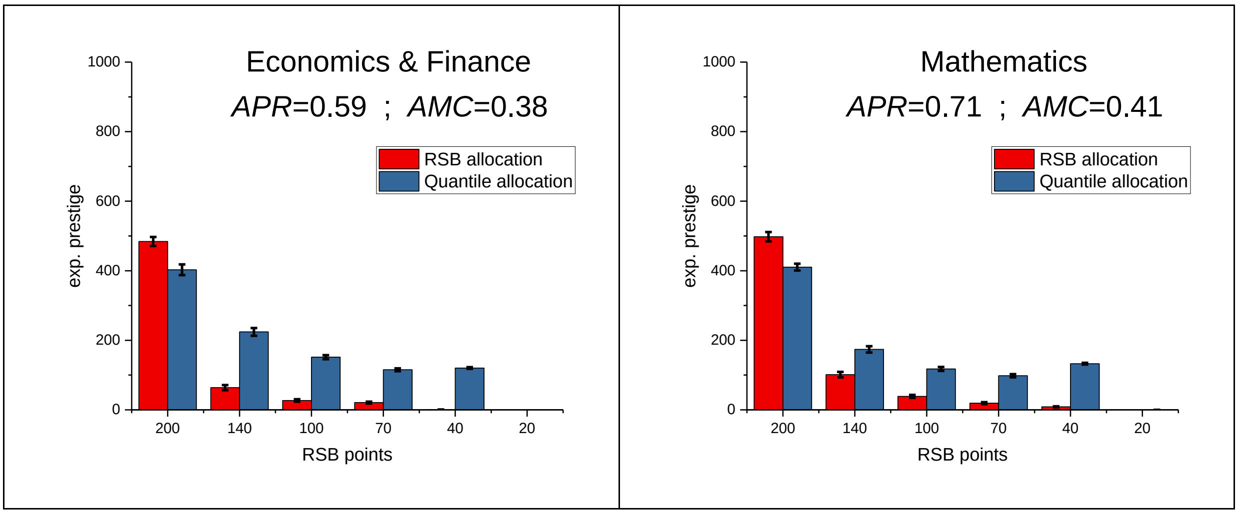




This does not, however, mean that the scheme introduced within the Polish reform is efficient. Figure 5 shows the average simulated prestige across 250 replications if the allocation is made according to the ministerial scheme shown with the red bars and the quantile allocation scheme with the blue bars. That is, each bar represents a decomposition of (2) into the RSB categories of the journals submitted. The average prestige ratios, APR, and average misallocation coefficients, AMC, are also given there. In graphical terms, the APR represents the ratio of the sum of red bars to the sum of blue bars for each panel of Figure 5. It is evident that a lot of prestige is wasted for the disciplines within the ministerial allocation scheme. Instead of submitting papers to journals worth 40-100 points, the researchers try to submit them to the highly rewarded journals with 200 or 140 points. As a lot of these highly rewarded journals have little prestige, due to breaks in the monotonicity, the overall prestige under the ministerial scheme is not as high as it is under the quantile allocation scheme.

Figure 5: Average expected prestige





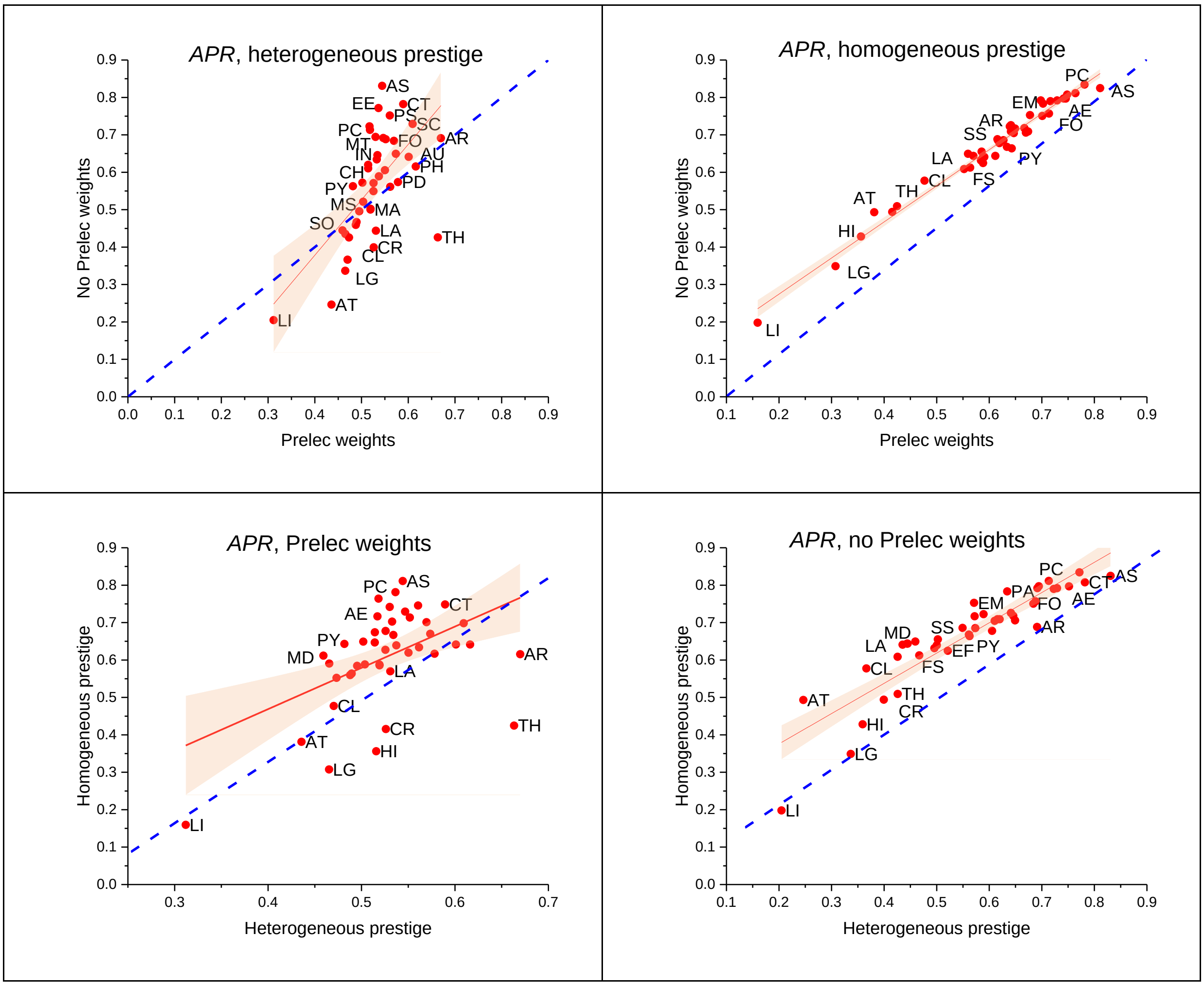
However, it should be noted that with the exception of history, the prestige for the best-rewarded category under the ministerial scheme is nevertheless higher than that given by the quantile allocation. This means that some highly rewarded journals are selected by researchers with sufficient ability, giving a realistic chance of publication success. It implies that the efficiency of allocation under the ministerial scheme might be better than is shown here if the prestige of the best journals is undervalued relative to less prestigious journals. This might indeed be the case, as some studies suggest that journals' impact measures might not offer a good representation of their quality (see, e.g., Seglen, 1997; Lüsher, 2018 and others).

The Appendix contains the results for all 44 disciplines. In Figure 6, we compare the APRs obtained in four schemes, with and without Prelec weights, so assuming that the researchers know the probabilities of acceptance by each journal, and with homogeneous and heterogeneous prestige. The schemes are compared pairwise, with linear regression lines and 95% regression confidence. The 45-degree dotted line separates cases where the APRs on the horizontal line are smaller than those on the vertical line. Two-letter abbreviations for some journals are also displayed (for a list of disciplines and abbreviations see the Appendix).

Not surprisingly, if the researchers know the true probabilities of acceptance by the journals, they are able to allocate their papers more accurately, which results in an increase in allocation efficiency. This is shown when most scatter points are above the 45-degree line, as in the upper left panel, or all of them are, as in the upper right panel. Similarly, if the prestige of the journals is assumed to be homogeneous, the allocation efficiency is markedly higher than when the prestige is heterogeneous. The advantage in allocation efficiency of the true probability case over the Prelec weights case is unambiguous for the homogeneous prestige, and, for the heterogeneous prestige, less so. There are, however, disciplines which clearly do not fit in the homogeneous scheme. These are mainly disciplines in the humanities rather than in science: literature (LI), history (HI), language studies (LG), theology (TH), cultural and religious studies

(CR), art studies (AT) and, somewhat unexpectedly, archaeology (AR); see the left bottom panel of Figure 6.

Figure 6: Pairwise comparison of average prestige ratios in all four schemes



5 Conclusions and policy recommendations

The model presented above can be used without modification to compute various different efficiency measures and benchmarks. For example, the average allocation efficiency can be compared with the desired efficiency defined by (2). Straightforward modifications might allow more complicated models to be developed that can cover repeated submissions after rejections (with some learning mechanisms), narrow or wide sets of target journals, and much more. However, these modifications would come at the expense of computational cost, which is already quite heavy if the computations are made for a large number of disciplines and several thousand journals.

In Section 2, we listed six possible reasons for the misallocation, which can cause inefficiency. The first reason is the multidisciplinary nature of some journals as some loss of allocation efficiency is inevitable for any unified across disciplines rewards system. Nevertheless, the degree of claimed multidisciplinarity is striking. It is easy to identify journals with spurious multidisciplinarity, meaning many disciplines claim them, but virtually no papers from some of those disciplines are published in them. Reducing this spurious multidisciplinarity might improve efficiency, as the rewards would better represent the prestige of the journals. Secondly, granting high rewards to some journals with low prestige might by justified by goals other than the academic objectives of the reform. It is possible that the fuzziness of these objectives might leave room for some imperfections. Thirdly, efforts to promote some newly created journals with no citations might affect efficiency further. The static nature of the model, which does not allow for expected future prestige, is evidently a shortcoming, but we feel that it might be difficult to change this without further substantial modifications and access to more detailed data. Fourthly, and interestingly, the misallocation might result from a conscious effect to improve efficiency in cases when some prior information about the distribution of abilities is available. In other words, if the quantile scale adopted for a given discipline is too ambitious for the abilities of the researchers, the scale itself might cause inefficiency. Suppose that all researchers in the discipline can publish in journals between the 0.5 and 0.75 quantiles. In the scheme given above, all the publications in this group get the same reward. This would encourage the researchers to submit to the easiest journal available, resulting in a loss in efficiency as some of the researchers might have been able to publish in the more prestigious journals in this group. In this case, then, it would make sense to adjust the scale and diversify the 0.5-0.75 interval. This argument works particularly well if the scale is too harsh, too lenient or entirely inadequate, as might be the case for the outlier disciplines identified in Figure 6. Further work is underway and consists of changing the assumption that abilities are uniformly distributed to be more informative. Fifthly, national, regional or political interests might be regarded as the priority, leading to particular journals being promoted.

That the allocation decisions might be using some subjective information about the probabilities of papers being accepted instead of using the true probabilities of acceptance creates an additional loss of efficiency, at least for some disciplines. This means that more transparency is needed to find out about the true qualities, requirements and constraints of the journals so that the researchers could evaluate these probabilities reasonably accurately.

Generally, it seems that the 'one size fits all policy' used in the Polish system, where the prestige of journals is assessed jointly for all disciplines, works to an extent. It is much easier to implement than the heterogeneous system of allocating different rewards to the same journal, depending on the discipline. However, it may be better said that 'one size fits nearly all'. There are evidently some disciplines that should be treated differently, mainly in the humanities. Equally, further inquiry into how abilities are distributed within particular disciplines is likely to conclude that diversifying the point rewards might lead to an improvement.

References

al-Nowaihi, A. and S. Dhami (2006), 'A simple derivation of Prelec's probability weighting function', *Journal of Mathematical Psychology* **50**, 521–524.

Bertoli-Barsotti, L. and T. Lando (2017), A theoretical model of the relationship between the h index and other simple citation indicators, *Scientometrics*, **111**, 1415–1448.

Bornmann, L. and K. Wohlrabe (2019), 'Normalisation of citation impact in economics', *Scientometrics* **120**, 841–884.

Callaway, E. (2016), 'Beat it, impact factor! Publishing elite turns against controversial metric', *Nature* **535**, 210–211. doi:10.1038/nature.2016.20224. PMID 27411614.

Cherkasin, I., S. Demidova, S. Imai and K. Krishna (2009), 'The inside scoop: acceptance and rejection at the Journal of International Economics', *Journal of International Economics* **77**, 120-132.

Devroye, L. (1986), *Non-uniform random variate generation*, Springer, Berlin.

Gorodnichenko, Y. T. Pham and O. Talavera, (2021), 'Conference presentations and academic publishing', *Economic Modelling* **95**, 228–254.

Lüsher, T. (2018), 'Measuring the unmeasurable: assessing the quality of science and scientists', *European Heart Journal* **39**, 1765–1769 ,doi:10.1093/eurheartj/ehy295.

Hudson, J. (2013), 'Ranking journals', *The Economic Journal* **123**, F2202-F222.

MacLeod W.B. and M. Urquiola (2021a), 'Appendix: Why does the United States have the best research universities?', https://www.aeaweb.org/content/file?id=13827 .

MacLeod W.B. and M. Urquiola (2021b), 'Why does the United States have the best research universities? Incentives, resources, and virtuous circles', *Journal of Economic Perspectives* **35**, 185–206.

McKiernan, E.C, L.A. Schimanski, C. Muñoz Nieves, L. Matthias, M.T. Niles, J.P. Alperin, (2019)' 'Use of the Journal Impact Factor in academic review, promotion, and tenure evaluations', *eLife* **8**; doi:10.7554/eLife.47338.

Prelec, D., (1998), 'The probability weighting function', *Econometrica* 66, 497-527.

Seglen, P. O. (1997). 'Why the impact factor of journals should not be used for evaluating research?', *British Medical Journal* **314** 498–502.; doi:10.1136/bmj.314.7079.497. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2126010/pdf/9056804.pdf

Stephan, P.E. (2012), 'Perverse incentives', *Nature* **484**(7392), 29-31.

Stern, D.I. (2013), 'Uncertainty measures for economics journal impact factors, *Journal of Economic Literature* **51**, 173–189

Śpiewanowski, P. and O. Talavera (2020), 'Journal rankings and publication strategy', presented at the 14th International Conference *Computational and Financial Econometrics* (CFE), King's College, London.

Yin, X., J. Goudriaan, E.A. Lantinga, J. Vos, and H.J. Spiertz (2003), 'A flexible sigmoid function of determinate growth', *Annals of Botany* **91**, 361-3003.

Appendix

Detailed results for particular disciplines

Discipline	Codes	Prelec weight				No Prelec weight			
		Heterogeneous jrnl. prestige		Homogeneous jrnl. prestige		Heterogeneous jrnl. prestige		Homogeneous jrnl. prestige	
		APR	*AMC*	*APR*	*AMC*	*APR*	*AMC*	*APR*	*AMC*
Agriculture	AG	0.51	0.50	0.67	0.35	0.62	0.49	0.71	0.45
Archaeology	AR	0.67	0.96	0.62	0.34	0.69	0.86	0.69	0.56
Architecture & Urban Studies	AU	0.60	0.93	0.64	0.50	0.64	0.93	0.73	0.62
Art	AT	0.44	0.97	0.38	0.61	0.25	0.97	0.49	0.65
Astronomy	AS	0.54	0.21	0.81	0.39	0.83	0.10	0.82	0.33
Automatics & Electronics	AE	0.52	0.48	0.72	0.43	0.72	0.43	0.79	0.39
Biology	BI	0.53	0.77	0.63	0.47	0.55	0.75	0.69	0.55
Biomedical Engineering	BM	0.47	0.84	0.55	0.51	0.43	0.83	0.61	0.56
Canonical Law	CL	0.47	0.96	0.48	0.44	0.37	0.96	0.58	0.48
Chemical Engineering	CE	0.53	0.73	0.74	0.52	0.69	0.64	0.80	0.43
Chemistry	CH	0.51	0.65	0.65	0.46	0.61	0.60	0.70	0.50
Civil Engineering & Transport	CT	0.59	0.68	0.75	0.42	0.78	0.63	0.81	0.44
Culture & Religion	CR	0.53	0.96	0.42	0.60	0.40	0.96	0.49	0.69
Ecology & Environm. Studies	EE	0.54	0.55	0.78	0.30	0.77	0.50	0.83	0.36
Economics & Finance	EF	0.50	0.77	0.59	0.38	0.52	0.77	0.62	0.46
Environm. Engineering & Mining	EM	0.53	0.73	0.68	0.42	0.57	0.72	0.75	0.42
Pharmaceutical Science	FS	0.49	0.80	0.56	0.48	0.47	0.79	0.61	0.54
Fishery & Zoology	FZ	0.57	0.65	0.67	0.50	0.65	0.58	0.71	0.45
Food Technology	FT	0.54	0.72	0.64	0.46	0.59	0.68	0.72	0.50
Forestry	FO	0.57	0.69	0.70	0.55	0.68	0.64	0.75	0.53
Geography	GE	0.55	0.74	0.73	0.33	0.69	0.73	0.79	0.44

Discipline	Codes	Prelec weight				No Prelec weight			
		Heterogeneous jrnl. prestige		Homogeneous jrnl. prestige		Heterogeneous jrnl. prestige		Homogeneous jrnl. prestige	
		APR	*AMC*	*APR*	*AMC*	*APR*	*AMC*	*APR*	*AMC*
Health Studies	HS	0.47	0.83	0.59	0.42	0.44	0.83	0.64	0.51
History	HI	0.52	0.96	0.36	0.58	0.36	0.96	0.43	0.66
Informatics	IN	0.53	0.62	0.67	0.48	0.65	0.58	0.72	0.48
Languages	LG	0.47	0.97	0.31	0.65	0.34	0.97	0.35	0.69
Law	LA	0.53	0.95	0.57	0.51	0.44	0.95	0.64	0.57
Literature	LI	0.31	0.98	0.16	0.60	0.20	0.98	0.20	0.67
Management	MA	0.52	0.85	0.59	0.47	0.50	0.84	0.64	0.51
Material Science	MS	0.50	0.73	0.58	0.45	0.50	0.70	0.63	0.45
Mathematics	MT	0.55	0.66	0.71	0.41	0.69	0.61	0.76	0.46
Mechanical Engineering	ME	0.50	0.60	0.65	0.37	0.57	0.56	0.72	0.40
Medicine	MD	0.46	0.76	0.61	0.43	0.44	0.75	0.64	0.48
Pedagogy	PD	0.58	0.89	0.62	0.51	0.57	0.84	0.69	0.50
Philosophy	PH	0.62	0.95	0.64	0.38	0.62	0.94	0.71	0.51
Physical Culture Studies	PC	0.52	0.67	0.76	0.33	0.71	0.64	0.81	0.40
Physics	PY	0.48	0.57	0.64	0.41	0.56	0.51	0.66	0.41
Politics & Administration	PA	0.53	0.81	0.70	0.41	0.63	0.81	0.78	0.44
Psychology	PS	0.56	0.73	0.75	0.34	0.75	0.72	0.80	0.46
Security Studies	SS	0.52	0.89	0.59	0.44	0.50	0.89	0.66	0.54
Social Communication & Media	SC	0.61	0.76	0.70	0.43	0.73	0.71	0.79	0.48
Sociology	SO	0.49	0.93	0.56	0.50	0.46	0.92	0.65	0.56
Technical Informatics & Telecom.	TI	0.55	0.76	0.62	0.49	0.61	0.73	0.68	0.51
Theology	TH	0.66	0.97	0.42	0.44	0.43	0.97	0.51	0.52
Veterinary	VE	0.56	0.79	0.63	0.52	0.56	0.77	0.67	0.55