ISSN 2704-7407

UNIVERSITÀ CATTOLICA DEL SACRO CUORE
Dipartimento di Economia e Finanza

Working Paper Series

Rising Stars

Erich Battistin, Marco Ovidi

Working Paper n. 106

June 2021

Rising Stars

Erich Battistin
University of Maryland

Marco Ovidi
Università Cattolica del Sacro Cuore

Working Paper n. 106
June 2021

Dipartimento di Economia e Finanza
Università Cattolica del Sacro Cuore
Largo Gemelli 1 - 20123 Milano – Italy
tel: +39.02.7234.2976 - fax: +39.02.7234.2781
e-mail: dip.economiaefinanza@unicatt.it

Rising Stars*

Erich Battistin†

University of Maryland, CEPR, FBK-IRVAPP and IZA

Marco Ovidi‡

Università Cattolica

June 2021

Abstract

We use the UK's 2014 Research Excellence Framework (REF) to study the attributes of top-scoring (four-star) publications in Economics and Econometrics. Although official documents contain aggregate scores for each institution, we show how these aggregates can be used to infer the score awarded by REF panellists to each publication. We demonstrate that this score responds to journal prestige as measured by the Thomson Reuters Article Influence Score. Several econometric analyses confirm the limited contribution of other publication attributes, such as the citation counts, to the awarded score, and publications in the top generalist and top-five economics journals are unambiguously awarded four stars. We conclude that in large-scale evaluations such as the REF, peer reviews and bibliometrics should be viewed as complementary modes of assessment: the time-consuming task of peer reviews would be more cost-effective if targeted to publications whose quality cannot be unambiguously classified using bibliometrics.

JEL Codes: H52 , H83 , I23 , I28

Keywords: Education Policy, Higher Education, Journal Ranking, Research Funding

---

*Our thanks to Graziella Bertocchi, Daniele Checchi, Francesco Fasani, Andrew Oswald, Barbara Petrongolo and James Wilsdon for helpful discussions and comments, to two referees for insighful comments, and to seminar participants at the XVI Brucchi Luchino Workshop, the 2018 Annual Conference of the Royal Economic Society and the Italian National Agency for the Evaluation of Universities and Research Institutes (ANVUR). Any responsibility for the views expressed in the article rests solely with the authors.

†Department of Agricultural and Resource Economics, 7998 Regents Drive, Symons Hall, College Park, MD 20742. Contact: ebattist@umd.edu.

‡Department of Economics and Finance, Università Cattolica, Via Necchi 5, Milan 20123, Italy. Contact: marco.ovidi@unicatt.it.

1 Introduction

Performance-based research funding has been adopted in most European countries to encourage and reward excellence in universities, which leads to a surge in the number of national assessments conducted on a regular basis (Zacharewicz et al., 2019). Research quality serves as the yardstick by which productivity and reputation are commonly assessed and compared. The extent to which assessments are based on quantitative indicators of research impact (i.e., bibliometrics) or reviews by academic experts varies across countries.

We consider the Research Excellence Framework (REF), whose next evaluation round is well under way across higher education institutions of the UK. Approximately 20% of higher education budget in the country relies on public funding (Carpentier, 2021), which is allocated using the REF outcomes. Assessments have been conducted in the UK since 1986, and the most recent results were published at the end of 2014.

One of the accountability pillars of the REF is the quality assessment of research outputs, which is the focus of our work. The research quality is assessed by peer reviews following general guidelines regarding the originality, significance and rigour. However, the contribution of each output to the awarded quality of an institution is not disclosed. For each institution, official documents only report the share of outputs by the number of awarded *stars*, which ranges from one to four. This lack of full disclosure has spurred the discussion on how to filter top-scoring (four-star) works in future submissions. Therefore, understanding how 'the quality of a publication is assessed is essential for the incentive structure faced by institutions.

We consider the latest assessment completed, which is REF2014, and use all research outputs submitted to the Economics and Econometrics sub-panel to infer the determinants of research quality awarded by experts. We complement this information with output-level data on citations, bibliometric indicators of journal impact and information about the authors as of the REF2014 submission date. These variables are the closest approximation of the information available to panellists at the time of their assessments.

Although REF statistics are aggregated by institutions, we show in Section 4 that they remain informative on the classification of single outputs. We use this result to develop an econometric model to estimate the contribution of journal impact and other publication characteristics to the classification of experts. Specifically, we show that the number of stars awarded to publications is predicted by the influence of a journal's articles, which we measure with the Article Influence Score (AIS; see Thomson Reuters, 2014). Citation counts considerably vary across publications in the same journal. However, we reveal that citations and other output characteristics such as the H-index of authors add statistically very little to the AIS in explaining how research quality is awarded by experts.[1] The limited contribution of citation counts is particularly evident for publications in high-impact journals. In particular, the results of our analyses indicate that publications in top generalist and top-five economics journals are unambiguously awarded four stars.

We mark important differences from previous research. The study by Bertocchi et al. (2015) is the closest to ours, although they consider the Italian equivalent of the REF with a methodology tailored to their case study.[2] The model in Pitt and Yan (2017) is similar in spirit to ours, although although – other than using different statistical assumptions – their analysis does not rely on output-level information. In contrast to Hole (2017), who develops an algorithm to classify journals, our analysis does not restrict all publications in one journal to contribute equally to the classification of REF outputs. In addition, Hole (2017) considers only journals with the most submissions in REF2014: instead, we use all research outputs, including those not published in academic journals. Our findings mirror the correlations between citations and REF scores in Traag and Watman (2019) across several disciplines. Their analysis is based on an indicator of excellence derived from citations. Instead, we estimate the correlation of peer reviews with journal-level metrics (AIS) and output-level characteristics (such as citations) to draw more nuanced recommendations for policy.

The predictive power of AIS documented here mirrors the high correlation in HEFCE (2015) between stars awarded in the REF and the SCImago Journal Rank index (an indicator similar to AIS). Their analyses are mostly univariate, employ anonymised data that are not publicly available and are limited to works published in 2008, which covers approximately 10% of the REF submissions for Economics and Econometrics. Our work provides a method

[1]Citation counts naturally depend on field of study and time elapsed since publication. In our analysis, we use a measure of citations that adjusts for the publication year and field of study.

[2]In the Italian ANVUR national assessment, the Economics and Econometrics panel relies on bibliometrics to assess articles in scientific journals and assesses all other outputs by peer review with a full disclosure of the final outcomes to the author of the submitted output. Checchi et al. (2019) find that the number of three- and four-star outputs in the REF can be almost perfectly predicted by the bibliometric algorithm of ANVUR. However, their work considers the institutional-level score and does not disentangle the contribution of the journal impact and output-level citations to the REF classification.

to overcome these limitations using data that can be retrieved from official documents and opens the door to replicability in fields other than economics. We select the AIS as the indicator of journal influence due to multidisciplinary research. The AIS builds on the same algorithm used by Google to rank websites (Brin and Page, 1998). The optimality of the AIS for measuring intellectual influence is demonstrated using the axiomatic approach in Palacios-Huerta and Volij (2004). Moreover, the AIS predicts expert-based evaluations of research quality (Hudson, 2013). In particular, the REF2014 data show that the AIS predicts what institutions value when deciding about submissions, as more than half of their outputs are concentrated in few journals (23 of 283 in total) with above-average AIS, as we show below.

Our results are of interest to policy makers who study the regulatory framework of future assessments. The evaluation of research outputs in REF2021 will follow the same REF2014 regulation. We do not take a normative approach on how the reviews should be conducted, which is the focus of Regibeau and Rockett (2016) among others. However, our study is informative of how peer reviews and bibliometrics can be combined in the interest of time and public spending (see also Moed, 2007). Specifically, by outlining the path to REF2021, Lord Stern's review (Stern, 2016) recommended the responsible use of bibliometrics (Wilsdon et al., 2015).[3] Our findings suggest that for Economics and Econometrics, peer reviews and bibliometrics in a large-scale evaluation such as the REF may be viewed as complementary modes of assessment to identify unambiguously top-scoring journals and review outputs outside this tier only. Our policy recommendation is that the time and resources needed for peer reviews should be devoted to finding hidden four-star gems in academic outlets with lower bibliometric indicators of impact rather than overrated outputs in top-scoring outlets.

The remainder of this paper is organised as follows. Section 2 presents the institutional background. Section 3 describes how we integrate the information in REF official documents with bibliometrics data. Section 4 documents how aggregate statistics on the REF performance by institutions are informative on the correlation between the AIS and the classification of single outputs. Our econometric model is presented in Section 5. Section 6 presents the empirical results, and Section 7 concludes the paper.

[3]The review also reported that "*bibliometric evidence could be useful to panels in determining whether there is a significant discrepancy between the grade profile for outputs [...] as determined by peer review, and citation data*" (page 21). Similar recommendations are in the Leiden Manifesto (Hicks et al., 2015) and San Francisco Declaration on Research Assessments (see http://www.ascb.org/files/SFDeclarationFINAL.pdf).

2 Background and Context

The Research Excellence Framework

The UK is considered a world leader in higher education, which contributes £73 billion yearly to the national economy and has been linked to 20% of GDP growth between 1982 and 2005 (Universities UK, 2015). The quality of research produced by institutions is assessed by the Research Excellence Framework (REF), which was commissioned by the four UK higher education funding bodies.[4] The REF provides accountability for investment in research, with implications for the allocation of public funding and reputation of institutions.

We use data from REF2014, which is the latest assessment and was completed at the end of 2014. Special panels assessed the productivity of 154 universities between 2008 and 2013 and reviewed 190,000 research outputs by 52,000 academics. In the 2020/21 fiscal year, Research England distributed approximately £1.6 bn of research funding using REF2014 outcomes (Research England, 2020). This source of funding represents 20% of universities' budgets and is their second largest source of income (Carpentier, 2021).

We consider the evaluation of research outputs, which constitute the most important component of the REF. Compared to its predecessors, REF2014 assessed the research impact in addition to research quality.[5] Research quality accounted for 65% of the profile of each institution, 20% was awarded for impact, and an additional 15% was for the research environment (e.g., infrastructure and income through research activities). The performance on research quality is what matters most in the decision to hire or expand (De Fraja et al., 2019).

The next assessment, REF2021, is under way and will be similarly implemented to REF2014. The most substantial changes concern the selection of faculty, which is not a dimension considered in this study. Specifically, the representativeness of research was not guaranteed in REF2014 because institutions could decide to submit outputs for selected faculty members.[6] Moreover, the portability of research outputs provided the incentives to hire

[4] These bodies are Research England (RE), the Scottish Funding Council (SFC), the Higher Education Funding Council for Wales (HEFCW) and the Department for the Economy, Northern Ireland (DfE). See https://www.ref.ac.uk/about/whatref/ for a description of the REF.

[5] Impact is defined as "an effect on, change or benefit to the economy, society, culture, public policy or services, health, the environment or quality of life, beyond academia" (REF, 2011).

[6] However, the amount of public funding depends on the number of academics submitted. One additional complication that affects the representativeness of research is that institutions can strategically submit staff

productive researchers from other institutions close to the REF census date. REF2021 introduces some changes to address these issues, which Stern (2016) described as sources of "*gaming*" behaviour.[7]

Assessment of research outputs

We look at the 2,600 research outputs submitted to the REF2014 Economics and Econometrics sub-panel. This unit of assessment (UoA) considered the outputs from 28 departments in the UK (Table G.1). Guidance and criteria for the evaluation process were disclosed well before the submission deadline. Institutions were invited to submit outputs authored by the employed staff and published between 2008 and 2013.

The evaluation relied entirely on peer reviews by the panellists. The Economics and Econometrics sub-panel consisted of 18 national and international academics who read all submissions to identify the overrated outputs and hidden four-star gems regardless of the publication outlet. This process implies an average workload of approximately 140 outputs per panellist. Each listed output in a submission was assessed based on its originality, significance and rigour. Since this definition allows for subjective judgement, the panellists used bibliometrics to inform assessments "*when considered appropriate*" (REF, 2012). According to official documents, bibliometrics affected the quality awarded in "*very few cases*" (REF, 2015, p. 51).[8] The submission of interdisciplinary research was encouraged, and publications falling outside the expertise of the Economics and Econometrics panel were cross-referred to and assessed by experts in panels of other disciplines. However, nearly all outputs in our data (98.8%) were assessed by the Economics and Econometrics panel.[9]

Each output was assigned to one of five mutually exclusive tiers, but the output-level classification was not disclosed. The quality depended on the number of awarded stars to distinguish among "world leading" (four-star), "internationally excellent" (three-star), "in-

---

to different units of assessment. For example, within economics departments, staff may be submitted to the Business and Management Studies sub-panel.

[7] All staff with significant responsibilities for research must be submitted to REF2021. Institutions may return the outputs of previously employed staff if publications were generated during the period of employment (REF, 2017, pp. 50-51). The weight assigned to research outputs in REF2021 remain predominant (60%) but is lower than that in REF2014. The contribution of research impact will increase to 25%.

[8] See https://www.ref.ac.uk/2014 for a description of the rules, outcomes and identity of the panellists.

[9] REF reports are accessible at ref.ac.uk/2014/panels/paneloverviewreports/. They show that cross-reference to other panels was limited to 2.7% of outputs. The outputs cross-referred to the Economics and Econometrics panel were excluded from our analysis.

ternationally recognised" (two-star) and "nationally recognised" (one-star) research. Submissions that fall short of national standards were flagged as "unclassified quality". The statistic used for accountability purposes was the share of outputs listed in a submission that were assigned to each quality level. The lack of transparency regarding the classification of single outputs has fuelled discussion on how to filter four-star work.

The research quality in Economics and Econometrics was found to be outstanding, with more than two-thirds of submitted outputs being at least "internationally excellent". However, the recognised excellence exhibited substantial variation across institutions in the proportion of four-star outputs (as demonstrated in Table G.1).

Related literature

Several studies investigated past research assessments in the UK (see Traag and Watman, 2019, for a review). Johnes et al. (1993) suggested that research ratings improved with the size and reputation of an institution. Clerides et al. (2011) concluded that departments might benefit from having members on the evaluation panel. Consistent with the latter study, De Fraja et al. (2019) suggest that institutions represented on the panel were awarded higher scores on the REF. They also show that the portability of outputs in REF2014 induced institutions to attract more-productive researchers by offering them higher salaries.

The REF generated intense debate over its regulatory framework and incentive structure. The mix between peer review and bibliometrics is often central to the discussion. Sgroi and Oswald (2013) showed that research excellence could be predicted using the journal rankings and citations. Regibeau and Rockett (2016) argued that the journal impact and citations could identify the quality of economics departments without relying on reviews from experts. Peer review leaves the door open for subjective bias that may affect assessments. Hudson (2013) showed that, conditional on various proxies for research quality, experts prefer theory journals and outlets with a strong focus on economics.

The merits and limitations of using citations to rank journals in economics and the influence on hiring and promotions in academia have been discussed at length (examples are Liebowitz and Palmer, 1984, Laband and Piette, 1994, Kalaitzidakis et al., 2003, Varin et al., 2016, and Hamermesh, 2018). A model-based approach to rank scientific outlets is shown in Bartolucci et al. (2015). In their work, the journal quality is unobserved, and indicators such as the AIS are used to proxy for this latent factor. We also address the unobserved quality awarded to each research output, which we infer from the total number of stars assigned to institutions. Palacios-Huerta and Volij (2004) adopted an axiomatic approach to rank academic journals and demonstrate the optimality of the PageRank algorithm, which is also employed by Google to rank websites (Brin and Page, 1998). We find that the AIS, which is based on the same methodology, is the strongest predictor of the research quality awarded by the REF panel.

3 Data and Bibliometrics

We accessed submissions for all institutions through the REF website. The data consist of 2,600 outputs in Economics and Econometrics, most of which are journal articles (2,388) and working papers (168). Starting from this information, we assigned to the corresponding journal all working papers that were flagged as forthcoming or published by August 2015.[10] Panel A of Table G.2 presents the breakdown by publication type that resulted from this selection. We collected citations for the journal and authors of each output to characterise the research influence and prestige. The Economics and Econometrics sub-panel had access to the citation counts for each publication made available from Elsevier's Scopus database in early 2014 and contextual data on the distribution of citations in the field and year of publication of the output. These files were deemed confidential and deleted at the end of the REF process, which makes replicating the results impossible. Therefore, we approximated the bibliometric indicators available to panellists with the most similar indicators obtained from the web.

We characterised each journal by its AIS. This choice was motivated by a study that demonstrated that the AIS predicted expert-based evaluations of research quality (Hudson, 2013).[11] We considered the AIS for 2013, which was the latest release at the time of the REF

[10]These are outputs for which we retrieved the publication status at the end of 2016. The implicit assumption here is that a working paper published by August 2015 must have been accepted for publication while the REF panel was at work. Our choice explains the small differences in submission counts by journal in Table 1 with respect to the official counts, which can be found in Hole (2017) among others.

[11]A ranking of economics journals by the AIS can be accessed at bit.ly\2Q00Eld (retrieved on 12th April 2021). The Italian ANVUR national assessments use the SCImago Journal Rank index (Gonzalez-Pereira et al., 2010), which is constructed similarly to the AIS. In our sample, the Spearman correlation between the two indicators of journal impact is 94%.

Table 1: Academic journals most frequently submitted

	(1)	(2)	(3)	(4)
			Estimated Probability:	
Journal	**Frequency**	**AIS**	**4 Stars**	**3 Stars**
4*				
Quartely Journal of Economics	30	7.05	1.000	0.000
Econometrica	70	4.48	0.936	0.064
Review of Economic Studies	63	3.44	0.904	0.096
American Economic Review	115	2.64	0.889	0.111
Probable 4*				
Journal of the European Economic Association	73	1.48	0.621	0.379
Review of Economics and Statistics	59	2.16	0.613	0.387
Possible 4*				
Economic Journal	106	1.11	0.478	0.522
Journal of Econometrics	95	0.96	0.374	0.626
3*				
Journal of Monetary Economics	42	1.14	0.299	0.701
Journal of International Economics	37	0.90	0.273	0.727
Journal of Public Economics	57	0.77	0.175	0.825
International Economic Review	30	0.76	0.166	0.834
Journal of Economic Theory	84	0.78	0.149	0.851
Journal of Development Economics	50	0.66	0.112	0.888
Econometric Theory	35	0.59	0.103	0.897
Journal of Health Economics	33	0.40	0.001	0.999
Games and Economic Behaviour	83	0.32	0.000	0.915
European Economic Review	52	0.22	0.000	0.848
Journal of Money Credit and Banking	34	0.14	0.000	0.771
Probable 3*				
Journal of Economic Behavior & Organization	42	-0.07	0.000	0.529
Possible 3*				
Journal of Economic Dynamics and Control	45	-0.13	0.000	0.389
Economic Theory	49	-0.07	0.000	0.378
2*				
Economics Letters	63	-0.37	0.000	0.114

Note. The table lists, in columns (1) and (2), journals with at least 30 submissions in Economics and Econometrics, together with their standardized Article Influence Score (AIS). Journal names are sorted by the estimated probability of scoring four stars, reported in column (3). Column (4) reports the estimated probability of scoring three stars. Journals are grouped by number of stars using the ranking methodology from Hudson (2013). See Section 6 for details.

evaluation. We standardised the AIS to have zero mean and unit variance by the Thomson's JCR field to adjust for differences in citation behaviour across disciplines. Interdisciplinary research was warmly encouraged (REF, 2011, p. 15): although 94% of submissions appeared in economics journals, and the remaining outputs spanned across fields such as psychology, mathematics and physics.[12] This finding suggests that economics has extramural influence on many other disciplines, which is consistent with the conclusions in Angrist et al. (2020).

The JCR database does not have universal coverage of journals submitted to the REF. In addition, working papers that are not forthcoming and other research outputs (e.g., books or book chapters) cannot be attributed an AIS value. The distribution of outputs for which the AIS was retrieved, which is 91% of REF submissions, is shown in Figure G.1. This distribution presents a long upper tail driven by high-impact outlets (such as the top-five journals in economics) and spikes across the entire support. Table 1 reveals the origin of these spikes and lists all journals with at least 30 submissions in our sample.

The citation count for each publication was obtained using Elsevier's Scopus, which is the same source available to the REF panellists. We measured the citations at the end of 2013 (i.e., as of the REF submission date) and retrieved information for 2,441 outputs (94% of the sample). We additionally considered Google Scholar because of its much larger array of publishing formats, although our conclusions are robust to the source of information employed.[13] This information was completed with the h-index of all authors at the time of the REF submission, which was computed from the Scopus database, and their reported affiliation in each output. Descriptive statistics for all bibliometrics are presented in Panel B of Table G.2. The citation counts and h-index in our analyses were adjusted for the publication year and field of study. Specifically, they are residuals from regressions on a full set of the year of publication and field of publication dummies.

[12] When a journal is assigned multiple fields (e.g., economics and statistics), standardisation was performed using the mean and standard deviation across all economics journals. In rare cases where all fields are outside economics, we considered the average of the field-specific standardised scores. The classification of fields available to REF panellists uses the Elsevier's ASJC categories. However, to standardise the AIS, we used Thomson's JCR categories. These alternative classifications are substantially equivalent for submissions in Economics and Econometrics.

[13] The Scopus and Google Scholar archives do not cover the same population of journals and publishers. The correlation between the two citation measurements is 86%, as computed from 2,441 outputs.

4 Graphical Analysis

Although REF statistics are aggregated by institution, we show that they are informative on the classification of single outputs. We use this insight to show that the AIS of a journal predicts the number of stars awarded to individual outputs, and publications in high-impact journals appear to have solid four stars. We also show that although citation counts considerably vary across publications in the same journal, they are less effective than the AIS at predicting the REF performance of an institution.

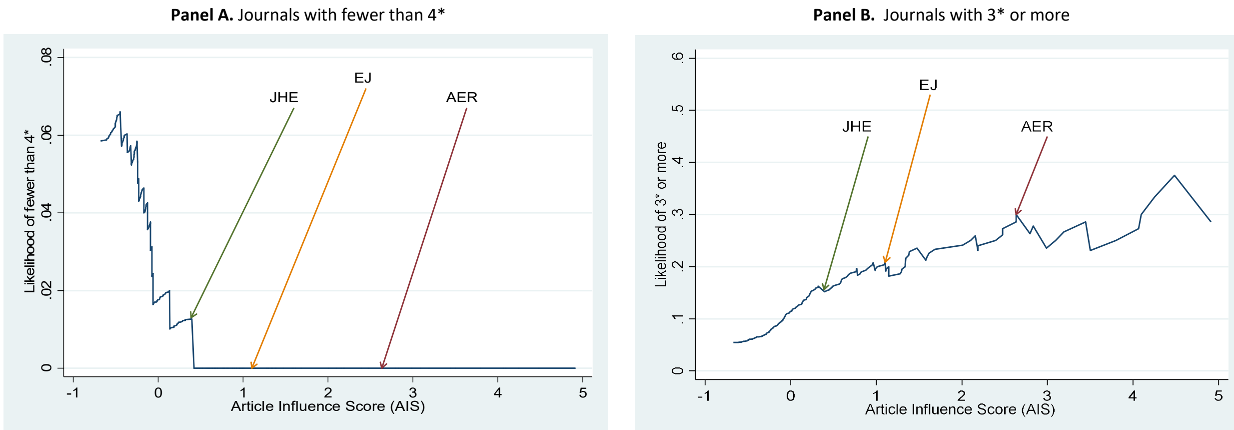
AIS predicts the classification of research outputs

To fix ideas, we consider publications in the Economic Journal (EJ). If REF submissions in this journal exceeded the number of four-star outputs in at least one institution, EJ publications must not have always been awarded four stars. Similarly, if the number of EJ submissions exceeded the number of outputs awarded one or two stars, then some EJ publications must have been awarded three stars or more. We use this idea to demonstrate that the likelihood of a top-scoring publication increases with the AIS of its journal.

Panel A of Figure 1 shows a value of 1.1 for the AIS of the EJ (standardised). Among journals with AIS at or above 1.1, the vertical axis presents the percentage of publications that are not always awarded four stars. Since this value is zero, the data do not reject the claim that outputs in journals with AIS at least as large as the EJ may be always awarded four stars. For example, these journals include the American Economic Review (AER). Panel A replicates this analysis for all AIS values and shows that a critical threshold emerges around the Journal of Health Economics (JHE). Specifically, the number of submitted JHE publications to the REF exceeds the number of four-star outputs in at least one institution, which implies that some JHE publications must have been awarded fewer than four stars. Panel A shows that the latter pattern is more likely for journals with AIS below JHE. Panel B of Figure 1 strengthens this conclusion. For example, among journals with AIS at or above 2.6 (the value for the AER), approximately 30% of the publications were awarded three stars or more . The same number is approximately 10% when the (standardised) AIS is zero.[14]

[14]The quantities in Figure 1 need not be monotone in AIS, which explains the saw-tooth pattern. As we explain in Appendix A, similar conclusions are obtained by considering combinations (e.g., pairs or triplets) of journals submitted to the REF instead of considering individual journals - see Figure A.1.

Figure 1: REF classification of outputs

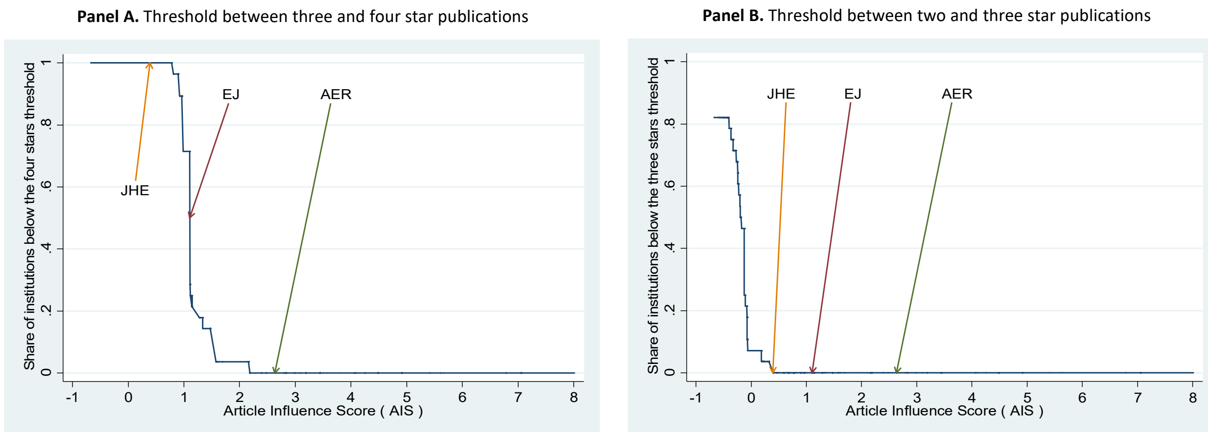


Note. This figure considers submissions by journal as explained in Section 4. Panel A shows the likelihood of being awarded fewer than four stars for publications in journals with standardized Article Influence Score (AIS) values at or above a certain value. For example, the data do not reject the hypothesis that publications in journals with AIS values at least equal to those of the American Economic Review (AER) or the Economic Journal (EJ) are awarded four stars. The Journal of Health Economics (JHE) represents a critical threshold. Panel B shows the likelihood of being awarded three stars or more for publications in journals with AIS values at or above a certain value. The support of the standardized AIS distribution is truncated at five because of the low number of journals above this value.

The relationship between output classification and AIS of the journal can be refined with assumptions on the quality of submitted outputs. For example, economists commonly assume that publications in top-five journals are always awarded four stars. These journals are highly respected and have standardised AIS values between 2.6 (AER) and 7.1 (the Quarterly Journal of Economics), which are often considered a "curse" because of their impact on visibility and career (Heckman and Moktan, 2020). If this assumption is correct, not all outputs published in top-field economics journals were awarded four stars. Figure G.2, which replicates Panel A of Figure 1, shows that submissions in the Journal of Econometrics (AIS of 0.96) or Journal of Economic Theory (0.78) cannot be consistently awarded four stars.

What would be the critical AIS values to award stars if the outputs were entirely classified based on this bibliometric indicator? Figure 2 shows that these critical values are remarkably close to those emerging in Figure 1. For example, 20.2% (or 19) of the publications submitted by Queen Mary University of London were awarded four stars (see Table G.1). We ranked all submissions in academic journals from Queen Mary University of London by AIS values and defined the critical cut-off by considering the AIS of the 20th publication, which is the first that would be awarded three stars. The remaining critical cut-offs were similarly determined. By repeating the analysis for all institutions, Panel A of Figure 2 shows the

Figure 2: Counterfactual classification based on Article Influence Score



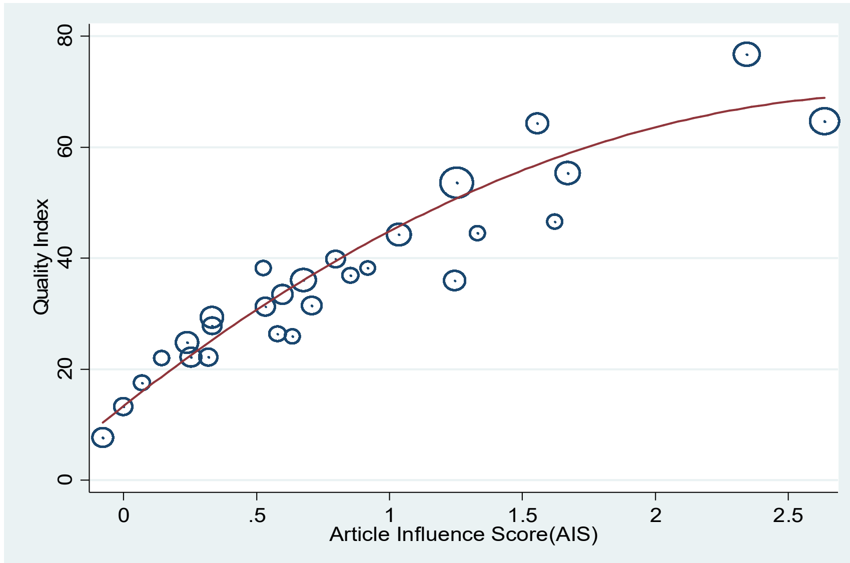
Note. All panels are derived by ordering outputs by their Article Influence Score (AIS), from the highest to the lowest. A classification of outputs based solely on AIS is maintained throughout. For each institution, outputs with the highest AIS are assigned four stars proportionately to the REF classification in column (2) of Table E.1. The remaining outputs are classified using columns (3), (4) and (5) of Table E.1 to assign three, two and one stars, respectively. Panel A in this figure shows the proportion of institutions that have passed the threshold for awarding four stars by value of AIS. For example, if classification were based solely on AIS, all publications in journals at least equal to the American Economic Review (AER) would be awarded four stars. In 43% of institutions, the Economic Journal (EJ) would represent the pass-mark between three and four stars. The Journal of Health Economics (JHE) would determine publications below four stars in all institutions. Panel B in this figure shows the proportion of institutions that have passed the threshold for awarding three stars by value of AIS. For example, if classification were based on AIS, the JHE would be the critical threshold. See Section 4 for details.

share of institutions that have reached the four-star threshold based on AIS values. We observe that publications with AIS at least as high as the AER would always score four stars. Publications with AIS lower than the JHE would always fall short of the critical cut-off to receive four stars. The critical journals that result from this analysis have AIS around the EJ, which represents the cut-off for a "world-leading" publication for 43% of institutions (12 of 28). The line in Panel B, which is analogously defined for the three-star cut-off, shows that the JHE is critical for scoring "internationally excellent".

The strong predictive power of the AIS spills over to the correlation between the average AIS and the REF score of an institution. This correlation has also been documented in other studies (e.g., Traag and Watman, 2019). Figure 3 plots the average AIS of all publications submitted by an institution against the Quality Index (QI) of the institution awarded by the REF (see the table footnote for definitions). The predictions from a regression on a quadratic polynomial in AIS are superimposed, which suggests that the AIS provides a fair approximation of the classification criteria followed by the panel. Obviously, it is important to understand whether the fit significantly improves after including additional output attributes such as citation counts. We address this empirical question in the next section.

The AIS predicts the final ranking from panellists and correlates with how institutions

Figure 3: REF Quality Index and Article Influence Score

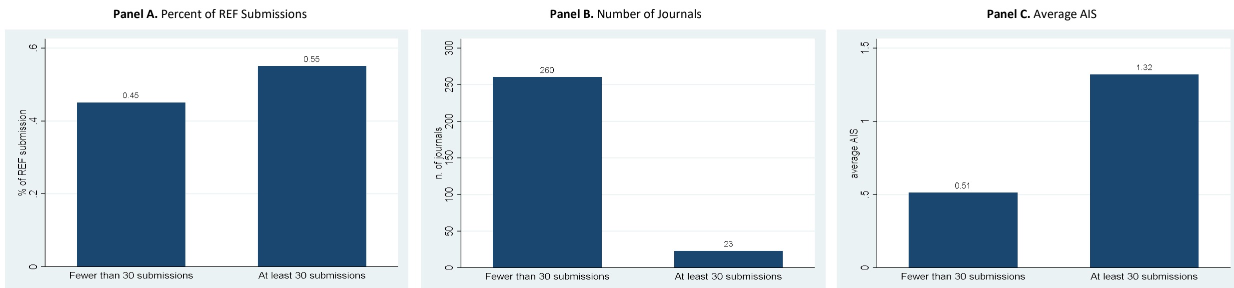


Note. The figure reports the scatterplot of an institution's Quality Index, on the vertical axis, against the average Article Influence Score (AIS) of all outputs submitted by the institution. Bubbles are proportional to the number of outputs submitted. Superimposed are predictions from a regression on linear and quadratic terms in AIS weighted by the number of outputs submitted. The Quality Index is computed using the current funding allocation formula, which depends on the incidence of top-quality outputs (80% and 20% to four- and three-star research, respectively, and no contribution of remaining outputs). See Section 4 for details.

made their strategic decisions about submissions. Specifically, we find that 55% of all submissions appeared in 23 journals with much higher AIS than the average. Thus, the departments had similar expectations about what constitutes a good output, and that perceived quality is strongly correlated with the AIS. Specifically, Panel A of Figure 4 shows that publications in journals with more than 30 entries in REF submissions account for 55% of the total (these journals are listed in Table 1). Panel B shows that these publications are concentrated in 23 journals (out of 283 in the REF), with an average standardised AIS of 1.32 (see Panel C). Remarkably, we show in Appendix B that the number of publications in 7 of these 23 journals almost perfectly predicts the number of four-star outputs awarded to institutions (with an R^2 of 98.7%).[15]

[15] These journals are (ordered from the largest AIS): Econometrica, Review of Economic Studies, American Economic Review, Journal of Monetary Economics, Economic Journal, Journal of Development Economics and Econometric Theory.

Figure 4: Concentration of REF submissions



Note. The figure reports statistics on academic journals in REF submissions. Bars depict separate results for journals with fewer than 30 outputs or with at least 30 outputs submitted to the REF. Panel A reports the share of total research outputs submitted. Panel B and Panel C report the number of academic journals and their standardised value of AIS, respectively. The average is reported on top of bars. See Section 4 for details.

Citations do not predict the classification of research outputs

The citation counts increase with the journal AIS but present significant differences within the same journal (as found in Starbuck, 2005; Anauati et al., 2020; Heckman and Moktan, 2020). This trend can be observed from column (1) of Table G.3, which shows the results from an output-level regression of Scopus citations on AIS, where we controlled for the research field and publication year. We find that a one-standard-deviation (hereafter, σ) increase in AIS is associated with a 0.164σ increase in citations, and the coefficient is highly significant. We find similar results when controlling for the average h-index of the authors (see column 3). However, with R^2 of 28.5%, the AIS alone cannot explain most of the variability in citations.[16]

Are citation counts more informative than the AIS in predicting the overall performance of institutions in the REF? We find that the answer to this question is no. Specifically, a regression (Table G.4) of the QI of an institution on the average number of citations and average h-index of submitted outputs yields an R^2 of 57%. In this regression, only the coefficient on citations is statistically significant. After including the average AIS of submitted outputs, R^2 increases to 89%. However, in the latter specification, only the coefficient on AIS remains statistically significant, which suggests that citations are uncorrelated with the REF classification once the journal impact is controlled for.

[16] Since the distribution of citations is heavily skewed, we investigate whether our conclusions are mechanically driven by the linear fit. Following Card and Della Vigna (2020), we further estimate a model for the inverse hyperbolic sine of citations in columns (2) and (4). Although R^2 substantially increases (57%), a large part of variability in citations remains unexplained. Columns (5)-(8) of Table G.3 present the estimates using Google Scholar citations, which yield a similar conclusion.

5 Empirical Specifications

We now motivate the empirical specifications in our analysis. These specifications lay out the relationship between variables in public-use data and unknown parameters in equations that describe how stars are awarded to outputs. The takeaways from these specifications are summarised at the end, and more technical details are in Appendices C and D.

Relationship between data and unobserved quantities

The REF outputs were published in 283 journals.[17] The number of submissions from institution i in journal j is X_{ij}, where $i = 1, \ldots, 28$ and $j = 1, \ldots, 283$. The terms X_{ij} for institution i are contained in the 283×1 vector $\mathbf{X}_i'$. This vector is retrieved from REF publications. We additionally collected data on the attributes of the submitted outputs (e.g., citations, field, and h-index of the authors). Let $\mathbf{Z}_{jk}'$ be the vector that contains the attributes of output k published in journal j; $\mathbf{Z}_i'$ contains the attributes of all submissions by institution i.

Variable D_{jk} denotes the number of stars awarded to output k in journal j.[18] Variable D_{jk} is *not* observed in REF data. However, we can still write:

$$Y_i^d = \sum_{j=1}^{283} \sum_{k=1}^{X_{ij}} \mathbb{1}(D_{jk} = d), \tag{1}$$

which is the number of publications awarded d stars by the REF panel for institution i, where $d = 1, \ldots, 4$. The measurements Y_i^d are contained in the 4×1 vector $\mathbf{Y}_i'$, which can be retrieved from official REF publications. Our investigation uses information on variables $(\boldsymbol{Y}_i, \mathbf{X}_i, \boldsymbol{Z}_i)$ to infer the stars awarded to single outputs.

General formulation of the problem

Equation (3) is the starting point of our investigation and follows from the REF criteria. Official documents state that the output quality was independently assessed for all outputs. Thus, the probability of awarding d stars to output k must depend on its attributes Z_{jk} and not on other outputs submitted by the institution. In econometric terms, this reasoning leads to the following equation:

[17] In this section, we work as if all outputs were articles in journals with AIS. We show in Appendix C how we account for other outputs such as books or book chapters.

[18] We set $D_{jk} = 0$ for "unclassified" outputs. The number of unclassified outputs is 16, which corresponds to 0.6% of submissions. Because of this low number, we consider only four tiers $D_{jk} \in 1, 2, 3, 4$ hereafter.

$$E\left[\mathbb{1}(D_{jk}=d)|\mathbf{X}_i,\mathbf{Z}_i\right]=\alpha_j^d+\gamma_j^d Z_{jk}. \tag{2}$$

Using (2) in (1) yields the following system of equations for $d=1,\ldots,4$:

$$E\left[Y_i^d|\mathbf{X}_i,\mathbf{Z}_i\right]=\sum_{j=1}^{283}\alpha_j^d X_{ij}+\sum_{j=1}^{283}\gamma_j^d\sum_{k=1}^{X_{ij}}Z_{jk}. \tag{3}$$

The estimation is challenged by the large number of unknown parameters. We address this problem by imposing restrictions guided by the graphical analysis in the previous section.

Adopted parametrisation

We group 283 journals into 3 mutually exclusive tiers depending on the AIS. We use the EJ as the lower limit for a top tier (Tier 1) that will include four-star outputs with high probability. Then, a middle tier is defined, i.e., Tier 2, which spans over a grey area comprising a mix of four- and three-star publications. Finally, a bottom tier is defined, i.e., Tier 3, as the complement to all included journals above. Building on the graphical analysis in Section 4, we use the JHE as the threshold to define the latter two tiers. We study the sensitivity of our conclusions to the definition of tiers in Section 6.

Because 3 tiers are now used instead of 283 journals, (3) simplifies to the system of equations (5) as follows. Specifically, journals are grouped by tier τ, where $\tau=1,2,3$. We assume that the deviation of α_j^d from the tier average $\alpha_{0\tau}^d$ depends on the journal characteristics (such as the AIS), which we denote by W_j. In addition, we impose constant effects of publication attributes within a tier. These two assumptions imply:

$$\alpha_j^d=\alpha_{0\tau}^d+\alpha_{1\tau}^d W_j, \qquad \gamma_j^d=\gamma_\tau^d, \tag{4}$$

for all j's in tier τ. By substituting (4) into (3), we have:

$$\begin{aligned} E\left[Y_i^d|\mathbf{X}_i,\mathbf{Z}_i\right] &= \sum_\tau \alpha_{0\tau}^d\left(\sum_{j\in\tau}X_{ij}\right)+\sum_\tau\alpha_{1\tau}^d\left(\sum_{j\in\tau}W_jX_{ij}\right) \\ &+ \sum_\tau\gamma_\tau^d\left(\sum_{j\in\tau}\sum_{k=1}^{X_{ij}}Z_{jk}\right). \end{aligned} \tag{5}$$

Two restrictions are imposed to further reduce the number of unknown parameters. First, we assume that the outputs in Tier 1 and Tier 2 are always awarded at least three stars. We also impose that the outputs in Tier 3 can never be awarded four stars. These restrictions are

Table 2: Classification restrictions

	Classification allowed: 4*	3*	2*	1*
Tier 1	X	X		
Tier 2	X	X		
Tier 3		X	X	X

Note. This table summarises the restrictions imposed in estimation. For example, we assume that publications in Tier 1 journals are awarded four or three stars. See section 5 for details and definitions of journal tiers.

consistent with the graphical analysis in Section 4 and allow for errors in the classification of outputs of at most one star.[19]

Summary of the estimation approach

We use all outputs to estimate the system of equations (5).[20] The four equations in this system are regressions of Y_i^d on the number of submitted outputs in each tier ($\sum_{j\in\tau} X_{ij}$), a term that involves the interactions between submitted publications and journal characteristics (e.g., AIS) in each tier ($\sum_{j\in\tau} W_j X_{ij}$), and a term that represents the attributes (e.g., citations) of the submitted publications in each tier ($\sum_{j\in\tau}\sum_{k=1}^{X_{ij}} Z_{jk}$). Estimation is performed using seemingly unrelated regressions. This approach yields the estimates of the parameters $\alpha_{0\tau}^d$, $\alpha_{1\tau}^d$ and γ_τ^d in (5), which we use to compute the probabilities of awarding stars. For example, the probability that publication k in journal j is awarded four stars is:

$$E\left[\mathbb{1}(D_{ijk}=4)|\mathbf{X}_i,\mathbf{Z}_i\right] = \alpha_{0\tau}^4 + \alpha_{1\tau}^4 W_j + \gamma_\tau^4 Z_{jk},$$

[19] This assumption is frequently made in empirical work on misclassification; e.g., Battistin and Sianesi (2011) and references therein. More generally, these restrictions on the classification probabilities mirror the results from past research on the informational content of international lists for journal rankings. By comparing various bibliometric indicators with the views of experts, Hudson (2013) concluded that some journals could be unambiguously clustered with respect to the number of awarded stars. This classification is fuzzy ("probable" and "possible" is his narrative) in other cases. Table G.5 shows that our definition of Tier 1 and Tier 2 coincides with that of unambiguously three-star or higher outputs in Hudson (2013), and most outputs in Tier 3 are expected to have a two-star classification. Similar conclusions emerge by considering other research on the REF (Hole, 2017).

[20] We assign the outputs for which AIS is not available, such as book chapters, to tiers by considering the publishing editor, as detailed in Appendix C. Our conclusions are robust to various alternative choices for assigning outputs without AIS to tiers.

which depends on the characteristics W_j (e.g., AIS) of a journal and attributes Z_{jk} (e.g., citations) of a publication.

We constrain the probabilities of four, three, two and one stars across tiers as explained in the previous section. These constraints are summarised in Table 2, where cells without crosses imply that the probabilities are set to zero. For example, these constraints imply that a "bad" publication in the AER (Tier 1) is never worth fewer than three stars. They also permit the presence of four-star gems in the grey area identified by Tier 2 (which comprises many top-field journals). Two stars is instead the expected valuation for outputs in Tier 3, although "bad" publications and hidden gems in this lowest tier may revise expectations either way by at most one star. In the econometric analysis, the constraints in Table 2 are obtained by constraining the parameters $\alpha^d_{0\tau}$, $\alpha^d_{1\tau}$ and γ^d_τ in the estimation of (5).

6 Results

The econometric analysis demonstrates that the journal tier based on the AIS strongly predicts the REF classification. Adding output characteristics such as citations and field does not change this conclusion. We additionally obtain a ranking of the most submitted journals implied by our analysis.

Baseline specifications

We start from a specification that controls for the AIS by stratifying on tiers. Specifically, we estimate (5), which imposes $\gamma^d_\tau = 0$ and $\alpha^d_{1\tau} = 0$ for all τ's and all d's. Using this specification, columns 1-3 of Table 3 show the estimated probabilities of the number of stars in each tier. Tier-2 outputs are defined as those with AIS values between JHE and EJ (the corresponding interval is $0.4 - 1.1$).

We find that Tier-1 and Tier-2 outputs are awarded four and three stars, respectively, with very high probability (89% and 93.6%). We cannot reject the hypothesis that Tier-2 publications are never awarded four stars, since the value of 6.4% in column 2 is not statistically different from zero. The classification of Tier-3 outputs is more nuanced with similar chances of scoring two and three stars (see column 3). Similar conclusions are obtained after adjusting for within-tier differences in AIS across outputs. Specifically, columns 4-6 in

Table 3: Estimation results from the baseline model

	Without adjustment			With adjustment (AIS)		
	Tier 1	Tier 2	Tier 3	Tier 1	Tier 2	Tier 3
	(1)	(2)	(3)	(4)	(5)	(6)
4* ("world leading")	0.890***	0.064		0.801***	0.161***	
	(0.043)	(0.082)		(0.037)	(0.06)	
3* ("internationally excellent")	0.110**	0.936***	0.491***	0.199***	0.839***	0.507***
	(0.043)	(0.082)	(0.033)	(0.037)	(0.06)	(0.022)
2* ("internationally recognised")			0.436***			0.422***
			(0.027)			(0.018)
1* ("nationally recognised")			0.074***			0.071***
			(0.012)			(0.012)
Number of journals in tier	48	31	205	48	31	205
Number of publications in tier	784	551	1,265	784	551	1,265
Mean of standardized AIS in tier	2.57	0.76	-0.09	2.57	0.76	-0.09

Note. Columns (1) to (3) show results from a baseline specification that allows for tier-specific intercepts. Columns (4) to (6) show results from regressions that control for within-tier heterogeneity using a quadratic polynomial in Article Influence Score (AIS). Tier is defined as in Section 5, from the Journal of Health Economics (JHE) to the Economic Journal (EJ). Columns (1) and (4) show the estimated probabilities that a publication in Tier 1 is awarded four or three stars. Columns (2) and (5) show the estimated probabilities that a publication in Tier 2 is awarded four or three stars. Columns (3) and (6) show the estimated probabilities that a publication in Tier 3 is awarded three stars, two stars or one star. The estimating equations are discussed in Section 5 and Appendix D. *** p<0.01, ** p<0.05, * p<0.1.

Panel A show the estimated probabilities from equation (5), which imposes that $\gamma_\tau^d = 0$ and enables that $\alpha_{1\tau}^d \neq 0$ when W_j includes a quadratic polynomial in AIS. The numbers in columns 4-6 are probabilities for the outputs with average AIS value in tier.[21] We conclude that Tier-1 and Tier-2 outputs are solid four and three stars, respectively.

Several checks documented in Appendix E study the robustness of this conclusion to the definition of tiers. For example, Figure E.2 shows the change in estimated probabilities of scoring four stars in Tier 1 and Tier 2 when the upper limit of Tier 2 is moved around the EJ. In this figure, any number n on the horisontal axis indicates that the upper limit of Tier 2 is defined using a journal n ranks apart from the rank of the EJ. The top line of Figure E.2 suggests that restricting Tier 1 to journals with AIS higher than EJ would reinforce our

[21]The relationship between estimated probabilities and AIS is shown in Figure G.3, which shows a more pronounced within-tier heterogeneity at the bottom end of the distribution. In addition, our specification imposes a continuity of the classification probabilities for journals at the boundaries between tiers. For example, we impose that a publication in the EJ (which marks the lower end of Tier 1) has an equal probability of being awarded four stars to a publication in the journal with the highest AIS in Tier 2. All specifications include controls for books, book chapters and other outputs as explained in Appendix ADD. The estimates that are not reported here suggest that the books published by international editors (see Table C.1) were most likely awarded four stars, and book chapters most likely receive three stars. Our analysis does not reveal any clear pattern for the field coefficients, which perhaps reflects that 94% of submissions were in economics journals.

conclusions on the probability of scoring four stars in this tier. Tier 2 would also be more likely to include four-star publications, as shown in the bottom line of the figure. Meanwhile, including journals with AIS lower than EJ in Tier 1 would result in lower probabilities of four-star publications in this tier, but the probability remains above 80%. The inclusion of these journals in Tier 1 would negatively affect the probability of four-star publications in Tier 2.

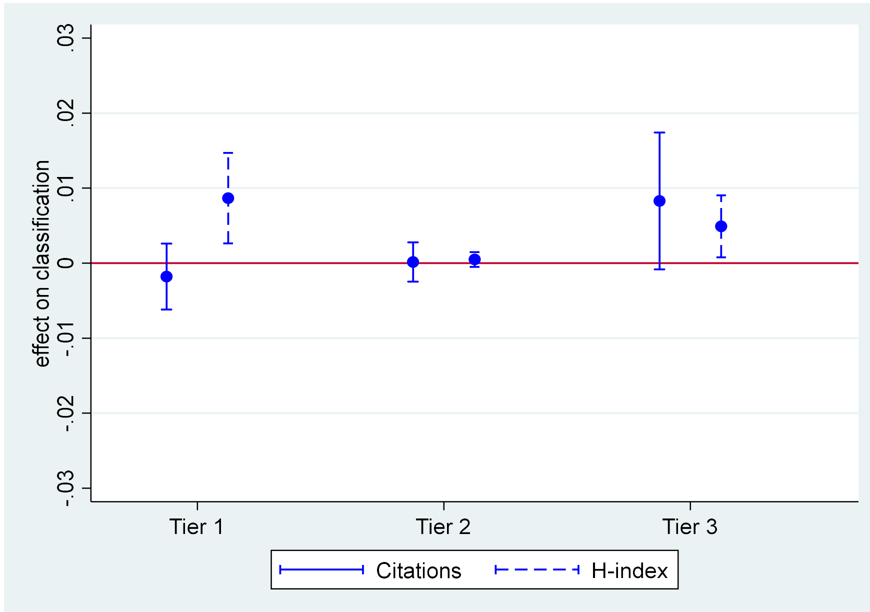
We conclude that publications in journals with higher AIS than EJ are unambiguously classified as "world leading" (four-star). On the other hand, there exists a grey area for journals with AIS lower than EJ where the classification is more ambiguous. As we shall see, this area includes a number of top-field journals, notably the Journal of Econometrics and the Journal of Economic Theory.

Effect of the output characteristics

What is the role of output characteristics such as the citations and field of publication? To answer this question, we estimate equation (5) with $\alpha^d_{1\tau} \neq 0$ as shown in columns 4-6 of Table 3. In addition, $\gamma^d_\tau \neq 0$ when Z_{jk} includes the output's citations, average h-index of authors and field. The predicted probabilities for outputs with the average value of Z_{jk} in tier are remarkably similar to the baseline estimates, as shown in Table G.6. For example, the probability that outputs in Tier 1 are awarded four stars is 80.1% in column 4 of Table 3. The corresponding probability is 77.5% after controlling for the citations and h-index of the authors in column 7 of Table G.6. A similar pattern emerges for all remaining probabilities.

We find that the output characteristics only marginally affect the probability of being awarded more stars, which is conditional on the tier membership. Specifically, Figure 5 shows the effects of a 10% increase in number of citations and h-index compared to their tier average. We show the effects and confidence intervals on the probability of being awarded the highest score: four stars in Tier 1 and Tier 2; three stars in Tier 3. For example, more citations are not associated with a better chance of scoring four stars in Tier 1; publications in this tier with authors whose h-index is 10% higher than the average have approximately 1% additional chance of being awarded four stars (from a baseline of 77.5%). The marginal association between citations and assessments is consistent with the fact that panellists used this information only in a limited number of cases (see REF, 2015, p. 51).

Figure 5: Citations and h-index effects



Note. Reported are the estimated effects on the probability of scoring four stars (in Tier 1 and Tier 2) or three stars (in Tier 3) from a change in citations (solid lines) or h-index (dashed lines). The effects correspond to an increase of 10% from the tier-specific average of citations count or h-index. The 95% confidence intervals are obtained from the specifications in columns (7)-(9) of Table F6. Citations count and h-index are measured as residuals from a publication-level regression on year and field fixed effects. H-index of a publication is the highest value among all co-authors. See Section 6 for details.

Figure 5 also shows that lower AIS values are associated with larger heterogeneity in the effects of other output attributes that contribute to the final classification by panellists. This finding suggests that peer reviews may be a cost-effective assessment for publications that are not in journals with unambiguously good bibliometrics. For example, the role of citations is relatively more important in Tier 3. However, these effects are imprecisely estimated at approximately 1% (from a baseline of 55.1% in column 7 of Table G.6).

Our conclusions are robust to further checks, as discussed at length in Appendix F. First, REF panellists assess the traits of research quality, such as significance and rigour, which are unobservable to us. We show that the results of our analyses are unlikely driven by such unobservable traits. Second, we show that the errors committed by using our model to predict the outcomes in (5), $Y_i^d - \hat{Y}_i^d$, are unrelated to the indicators of research performance in the 2008 Research Assessment Exercise (the predecessor of the REF), characteristics of higher education institutions evaluated by the REF (e.g., non-academic impact and research environment), and measures of ties between panellists and institutions. This finding implies that what our model cannot explain is not related to the dimensions that were flagged as determinants of the awarded quality in past research on the REF.

Journal ranking

Given the proliferation of rankings and their role in personnel decisions, we report in Table 1 the predicted probabilities of scoring three or four stars for the journals with the most submissions to the REF. We use all submitted outputs to estimate equation (5) and only report the probabilities for journals with at least 30 submissions. Predictions are reported for an output at the mean citation count of each journal. To ensure comparability with previous studies (Hudson, 2013 and Hole, 2017), the journals are grouped depending on values of predicted probabilities. Unambiguously four-star journals are those with at least a 65% probability of having a "world leading" classification. For probable and possible four-star journals, this probability must be larger than 50% and 35%, respectively. The same definitions are used to rank three-star and two-star journals.

Our results suggest that there is little space for "bad" outputs in top-five or generalist journals, for which our classification is unambiguously four-star. The classification at lower values of AIS becomes less clear-cut, and the EJ and Journal of Econometrics are examples of possible four-star journals. Publications in top-field outlets such as the Journal of Economic Theory and Journal of Public Economics are most likely awarded three stars.

7 Conclusion

In many countries, disciplinary panels of experts have access to bibliometrics that can be used to inform their assessments. The following practical question emerges: does the ranking of journals by informed experts mirror objective indicators of journal influence that can be more frequently attained and at much lower costs?

We have used the REF to re-consider this issue considering the resistance to bibliometric assessments by the academic community in the months preceding national evaluations (see Wilsdon et al., 2015, and UK Forum for Responsible Research Metrics, 2018). This exercise is not straightforward because the classification of outputs in the REF is not disclosed. This lack of disclosure has fuelled discussion in the national academic community regarding the determinants of top-scoring (four-star) outputs beyond these indicators of influence and citations, which are widely available.

A direct policy implication of our analysis is that a large part of Economics and Econometrics outputs can be automatically classified using bibliometric indicators of the journal impact. Specifically, we have shown that a classification based on the Thomson Reuters' Article Influence Score very closely approximates the outcome of peer reviews in Economics and Econometrics, especially at the top end of the journal impact distribution. Our definition of Tier-1 and Tier-2 journals includes $1,335$ publications, which represent 51% of REF2014 submissions. At least part of these publications could be classified using the Article Influence Score, and peer reviews could be more efficiently targeted to publications in lower-impact and interdisciplinary journals. The correlation between subjective assessments of research quality and bibliometric indicators has been documented in past research that investigates large-scale assessments in the UK (Clerides et al., 2011, Taylor, 2011, and Traag and Watman, 2019) and other countries (see, for example, Bertocchi et al., 2015). This result is often used to advocate for metric-based evaluations as opposed to peer reviews to reduce the administrative burden and risk of bias (see, for example, Laband, 2013).

Our view is that peer reviews and bibliometrics should be complementary modes of assessment. Bibliometrics could be used for accountability purposes and continuous monitoring of large-scale assessment exercises, which decreases the substantial costs in the REF assessment (a total of £246 m for REF2014; see Stern, 2016).

References

Anauati, M. V., Galiani, S., and Galvez, R. H. (2020). Differences in citation patterns across journal tiers in Economics. *Economic Inquiry*, 58(3):1217–1232.

Angrist, J., Azoulay, P., Ellison, G., Hill, R., and Lu, S. F. (2020). Inside job or deep impact? Using extramural citations to assess economic scholarship. *Journal of Economic Literature*, 58(1):3–52.

Bartolucci, F., Dardanoni, V., and Peracchi, F. (2015). Ranking scientific journals via latent class models for polytomous item response data. *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, 178:1025–1049.

Battistin, E. and Sianesi, B. (2011). Misclassified treatment status and treatment effects: An application to returns to education in the United Kingdom. *Review of Economics and Statistics*, 93(2):495–509.

Bertocchi, G., Gambardella, A., Jappelli, T., Nappi, C. A., and Peracchi, F. (2015). Bibliometric evaluation vs. informed peer review: Evidence from Italy. *Research Policy*, 44(2):451–466.

Brin, S. and Page, L. (1998). The anatomy of a large-scale hypertextual web search engine. *Computer Networks and ISDN System*, 30(1-7):107–117.

Card, D. and Della Vigna, S. (2020). What do editors maximize? Evidence from four Economics journals. *Review of Economics and Statistics*, 102(1):195–217.

Carpentier, V. (2021). Three stories of institutional differentiation: resource, mission and social inequalities in higher education. *Policy Reviews in Higher Education*.

Checchi, D., Ciolfi, A., Fraja, G. D., Mazzotta, I., and Verzillo, S. (2019). Have you read this? An empirical comparison of the British REF peer review and the italian VQR bibliometric algorithm. *CEPR Discussion Paper No.13521*.

Clerides, S., Pashardes, P., and Polycarpou, A. (2011). Peer review vs metric-based assessment: Testing for bias in the RAE ratings of UK Economics departments. *Economica*, 78:565–583.

Colussi, T. (2018). Social ties in academia: a friend is a treasure. *Review of Economics and Statistics*, 100(1):45–50.

De Fraja, G., Facchini, G., and Gathergood, J. (2019). Academic salaries and public evaluation of university research: Evidence from the UK Research Excellence Framework. *Economic Policy*, 34(99):523–583.

Gonzalez-Pereira, B., Guerrero-Bote, V. P., and Moya-Anegon, F. (2010). A new approach to the metric of journals scientific prestige: The SJR indicator. *Journal of Informetrics*.

Hamermesh, D. S. (2018). Citations in economics: Measurement, uses, and impacts. *Journal of Economic Literature*, 56(1):115–156.

Heckman, J. J. and Moktan, S. (2020). Publishing and promotion in Economics: The tyranny of the Top Five. *Journal of Economic Literature*, 58(2):419–470.

HEFCE (2015). The metric tide: Correlation analysis of REF 2014 scores and metrics (supplementary report ii to the independent review of the role of metrics in research).

Hicks, D., Wouters, P., Waltman, L., de Rijke, S., and Rafols, I. (2015). The leiden manifesto for research metrics. *Nature*, 520:429–431.

Hole, A. R. (2017). Ranking Economics journals using data from a national research evaluation exercise. *Oxford Bulletin of Economics and Statistics*, 79(5):621–636.

Hudson, J. (2013). Ranking journals. *Economic Journal*, 123:F202–F222.

Johnes, J., Taylor, J., and Francis, B. (1993). The research performance of UK universities: A statistical analysis of the results of the 1989 Research Selectivity Exercise. *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, 156:271–286.

Kalaitzidakis, P., Stengos, T., and Mamuneas, T. P. (2003). Rankings of academic journals and institutions in economics. *Journal of the European Economic Association*, 1(6):1346–1366.

Laband, D. (2013). On the use and abuse of Economics journal rankings. *Economic Journal*, 123:F223–F254.

Laband, D. N. and Piette, M. J. (1994). The relative impacts of Economics journals: 1970-1990. *Journal of Economic Literature*, 32(2):640–666.

Liebowitz, S. and Palmer, J. (1984). Assessing the relative impacts of economics journals. *Journal of Economic Literature*, 22(1):77–88.

Moed, H. F. (2007). The future of research evaluation rests with anintelligent combination of advanced metrics andtransparent peer review. *Science and Public Policy*, 34(8):575–583.

Palacios-Huerta, I. and Volij, O. (2004). The measurement of intellectual influence. *Econometrica*, 72(3):963–977.

Pitt, M. and Yan, Z. (2017). How does the REF panel perceive journals? A new approach to estimating ordinal response model with censored outcomes. Unpublished manuscript, University of Warwick.

REF (2011). Assessment framework and guidance on submissions.

REF (2012). Panel criteria and working methods.

REF (2015). Research Excellence Framework 2014: Overview report by main panel C and sub-panels 16 to 26.

REF (2017). Initial decisions on the Research Excellence Framework 2021.

Regibeau, P. and Rockett, K. E. (2016). Research assessment and recognized excellence: Simple bibliometrics for more efficient academic research evaluations. *Economic Policy*, 31(88):611–652.

Research England (2020). Research and knowledge exchange funding for 2020-21. https://re.ukri.org/finance/annual-funding-allocations/annual-funding-allocations-2020-21/.

Sgroi, D. and Oswald, A. J. (2013). How should peer-review panel behave? *Economic Journal*, 123:F255–F278.

Starbuck, W. H. (2005). How much better are the most-prestigious journals? The statistics of academic. *Organization Science*, 16(2):180–200.

Stern, N. (2016). Building on success and learning from experience, an independent review of the Research Excellence Framework.

Taylor, J. (2011). The assessment of research quality in UK universities: Peer review or metrics? *British Journal of Management*, 22:202–217.

Thomson Reuters (2014). *2013 Journal Citation Reports®*. Clarivate Analytics.

Traag, V. A. and Watman, L. (2019). Systematic analysis of agreement between metrics and peer review in the UK REF. *Palgrave Communications*, 5(29).

UK Forum for Responsible Research Metrics (2018). UK progress towards theuse of metrics responsibly: Three years on from The Metric Tide report.

Universities UK (2015). The economic role of UK universities.

Varin, C., Cattelan, M., and Firth, D. (2016). Statistical modelling of citation exchange between statistics journals. *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, 179:1–63.

Wilsdon, J., Allen, L., and Belfiore, E. (2015). The metric tide: Report of the independent review of the role of metrics in research assessment and management.

Zacharewicz, T., Lepori, B., Reale, E., and Jonkers, K. (2019). Performance-based research funding in EU Member States - A comparative assessment. *Science and Public Policy*, 46(1):105–115.

Appendix

(for on-line publication)

This Appendix is structured in different sections containing tables, figures and econometric specifications that add to those presented in the main text. The Appendix is organized by topic, with explicit reference to the relevant section in the main text.

A Derivation of graphs in Footnote 14

The following algorithm was considered to derive Panels A and B of Figure A.1, which is discussed in Section 4. To fix ideas, consider the P pairs that can be formed by considering journals j and l among the 283 that were submitted. We have a total of $P = \binom{283}{2}$ possible pairs per institution.

First, we flag pairs with a number of submissions (i.e., the sum of publications in journals j and l of the pair) exceeding the number of four-star outputs or the number of one- or two-star outputs in at least one institution. If submissions in the pair exceed the number of four-star outputs in at least one institution, it must be that some publications in the pair were not awarded four stars. Similarly, if submissions in the pair exceed the number of outputs awarded one or two stars, then some publications in the pair must have been awarded three stars or more.

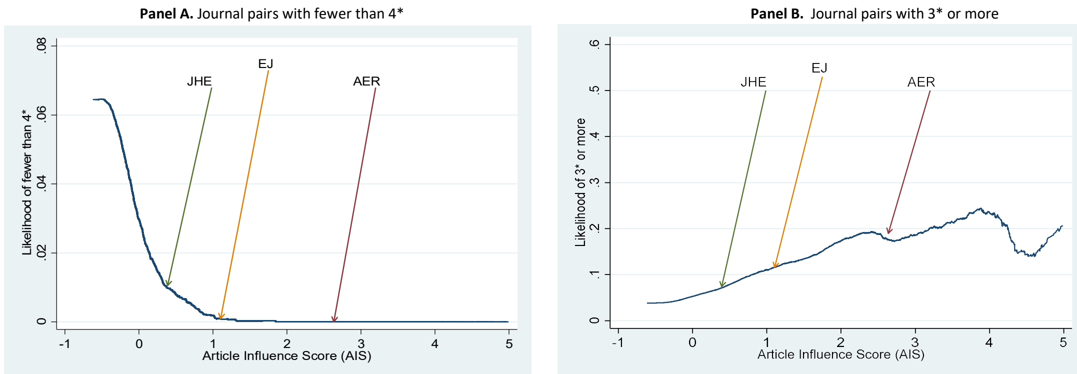
Second, we assign a value of AIS to pairs. Specifically, we do so by multiplying the standardised AIS of the two journals in the pair by the number of submissions from institution $i = 1, \ldots, 28$ in those two journals, and dividing by the total submissions in the pair. This procedure yields a total of $28 \times P$ different values of AIS.

Third, we plot the probabilities of being awarded fewer than four stars (Panel A) and more than two stars (Panel B) using these $28 \times P$ values. For example, Panel A of Figure A.1 shows that we cannot reject that pairs with values of AIS above 1 (which is approximately the AIS for the EJ) were always awarded four stars.

B LASSO estimation

We elaborate on our statement in Section 4 about the small number of journals explaining the number of stars awarded to institutions. Specifically, if only a few journals matter for determining the total number of four-star outputs, LASSO can help identify which journals

Figure A.1: REF classification of outputs (journal pairs)



Note. This figure considers submissions by journal pairs as explained in the Appendix. Panel A shows the likelihood of being awarded fewer than four stars for publications in journal pairs with AIS values at or above a certain value. For example, the data do not reject the hypothesis that publications in journals with AIS values at least equal to those of the American Economic Review (AER) or the Economic Journal (EJ) are awarded four stars. The Journal of Health Economics (JHE) represents a critical threshold. Panel B shows the likelihood of being awarded three stars or more for publications in journal pairs with AIS values at or above a certain value. The support of the standardized AIS distribution is truncated at five because of the low number of journals above this value.

have the greatest predictive power in equation (3). Building on the evidence in Figure 4, we operationalise this idea by considering the 23 journals with at least 30 submissions and which comprise 55% of the total sample. We estimate the following equation using all journals in REF submissions:

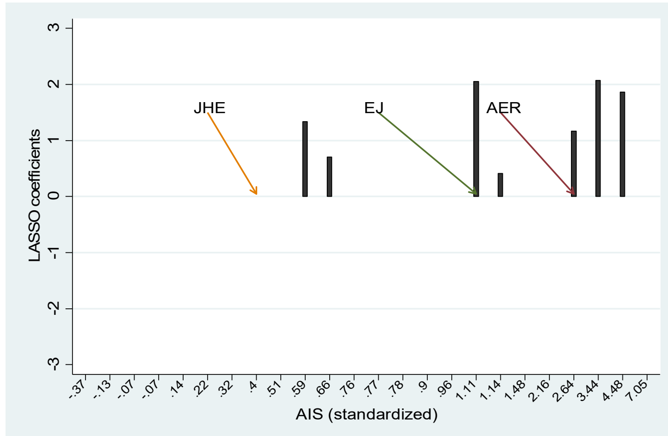
$$Y_i^d = \sum_{j=1}^{24} \alpha_j^d X_{ij} + \varepsilon_i^d, \tag{6}$$

which has 23 parameters for the journals listed in Table 1 plus one additional parameter for publications in all remaining $283 - 23 = 260$ journals. In practice, this equation allows to study the prediction accuracy of the 23 most submitted journals for the number of stars awarded to institutions. We estimate this equation by adding a penalty, determined through cross-validation, that forces "small" coefficients to be set to zero and effectively identify a smaller set of journals with the strongest prediction accuracy (this is standard with LASSO estimation).

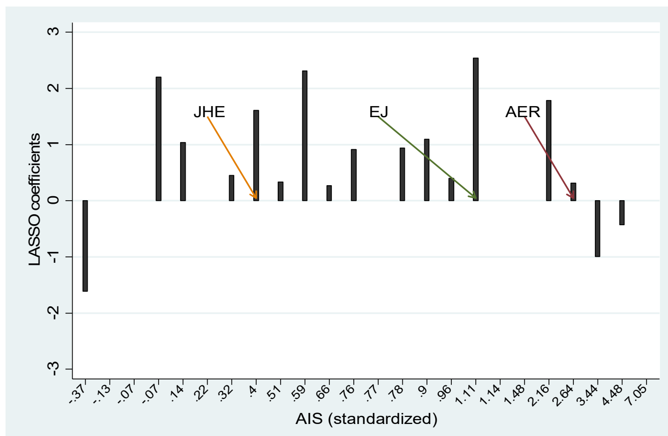
Information on as few as 7 journals predicts almost perfectly the number of four-star outputs awarded to institutions (these are listed in footnote 15 above). Panel A of Figure B.1 shows the value of LASSO coefficients obtained from (6) when Y_i^4 is considered. The horizontal axis here reports AIS of the 23 journals and, for the residual category, the average AIS across all remaining publications in the sample (which is 0.51). All journals with an AIS below the JHE have no predictive power for the number of four-star outputs at an

Figure B.1: LASSO coefficients

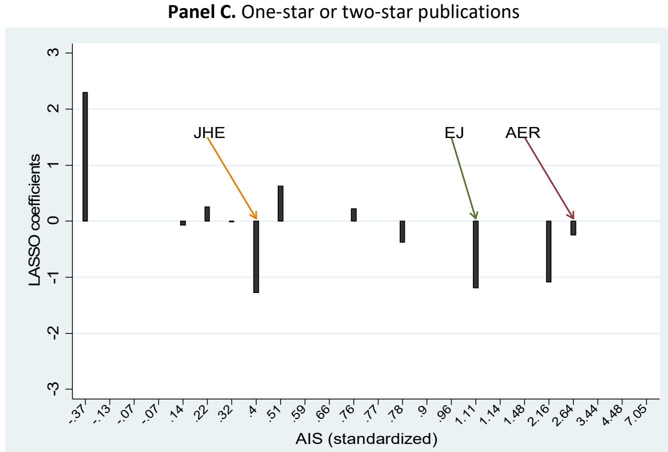
Panel A. Four-star publications



Panel B. Three-star publications



Panel C. One-star or two-star publications



Note. The figure reports LASSO estimates of equation (6) using, as outcome, the number of four-star outputs (Panel A), the number of three-star outputs (Panel B) and the number of one-star and two-star outputs (Panel C). The penalty parameters are determined through cross-validation. Coefficients for the 23 journals with at least 30 submissions listed in Table 3 are considered in equation (6), plus one additional coefficient for publications in any of the remaining 283-23=260 journals. The horizontal axis of each panel reports the journal standardised AIS. For the residual category, the average AIS of publications is reported and is equal to 0.51. Arrows in each panel mark the AIS of the American Economic Review (AER), the Economic Journal (EJ) and the Journal of Health Economics (JHE). See Appendix B for details.

Table C.1: Tier classification for books and book chapters

	baseline specification			alternative specification		
	(1)	(2)	(3)	(4)	(5)	(6)
	Tier 1	Tier 2	Tier 3	Tier 1	Tier 2	Tier 3
			Panel A. Books			
Cambridge University Press	X				X	
Harvard University Press	X				X	
John Wiley & Sons Ltd		X				X
Lambert Academic Publishing		X				X
Oxford University Press	X				X	
Princeton University Press	X				X	
Routledge		X				X
			Panel B. Book chapters			
Cambridge University Press		X				X
Canadian Tax Foundation			X			X
Elsevier			X			X
Emerald Publishing			X			X
Harvard University Press		X				X
North-Holland			X			X
Oxford University Press		X				X
Palgrave Macmillan			X			X
Princeton University Press		X				X
Springer			X			X
University of Chicago Press		X				X

Note. The table lists editors of books and book chapters included among the REF submissions. Panels A and B present the tier classification of books and book chapters, respectively. Columns (1)-(3) show the baseline specification. Columns (4)-(6) show an alternative allocation used as a sensitivity check. See Appendix C for details.

institution, a result consistent with the graphical analysis carried out in Section 4 (see, for example, Figure 1). Top-five journals in economics (the bars starting at the AER) and the EJ stand out for their prediction accuracy. An OLS regression of Y_i^4 on the set of journals selected through LASSO yields an R^2 of 98.7%. When the same set of 7 journals is used to predict the number of three-star outputs Y_i^3, the R^2 of the regression drops to 74.7%. Panel B and Panel C of Figure B.1 show coefficients from LASSO estimation of (6) using Y_i^3 and $Y_i^1 + Y_i^2$ (i.e., the number of outputs awarded two-stars at most), respectively. The non-zero coefficients in Panel B grows to 14, and journals with extreme values of AIS (above the AER) are negatively correlated with the number of three-star outputs. Here, the relationship between journals' AIS and the outcome is not as clustered as in Panel A. The results in Panel C show that only journals at the lower end of AIS spectrum are positively correlated with the outcome.

C Classification of research outputs with missing AIS

A choice has to be made on the classification of outputs for which AIS is not observed, as we explained in footnote 20 of Section 5. These outputs are approximately 8.6% of the sample, as shown in Table G.2. In our baseline specification books from international editors (e.g., Princeton University Press), the most frequent example, are assigned to Tier 1. Other books, including edited collections of chapters, are assigned to Tier 2. For chapters in edited books we follow a similar rule, downgrading by one tier the output classification: book chapters from international editors are assigned to Tier 2, and to Tier 3 otherwise. Classification in tiers of books and book chapters is shown in columns (1) to (3) of Table C.1. As a sensitivity check, we also consider a classification that assigns all book chapters to Tier 3, and books from international editors to Tier 2. This classification is described in columns (4) to (6) of the same table. Finally, the few remaining outputs (working papers not published by August 2015, datasets and reports) are assigned to Tier 3.

Our estimation results are robust to these classification criteria. For example, Table C.2 shows results obtained using classification in columns (4)-(6) of Table C.1, where the criteria described above are lowered by one tier. Results can be be compared to columns (4)-(6) of Table 3 in the main text. Results are very similar, with a slight decrease (increase) in the probability of scoring four stars for Tier 2 (Tier 3) outputs. This finding suggests that books published from international editors (see column 1 of Table C.1) are likely awarded four stars.

D Model specification and estimation

This section presents a general formulation of equation (5) that includes all outputs submitted to REF (Section 5 in the main text assumed, for simplicity, that all outputs can be assigned a value of the AIS). More precisely, outputs not in scientific journals (e.g., books) cannot be attributed a value of AIS (W_j), although other attributes (like citations, Z_{jk}) are observed. Define the dummies B_j, C_j and P_j for books, book chapters and working papers respectively.[1] Equation (4) is modified as follows:

[1] Articles in scientific journals without AIS, as well as other type of outputs, are included in the latter category.

Table C.2: Estimation results with alternative classification of outputs with missing AIS

	With adjustment (AIS)		
	Tier 1	Tier 2	Tier 3
	(1)	(2)	(3)
4* ("world leading")	0.727***	0.275***	
	(0.037)	(0.058)	
3* ("internationally excellent")	0.273**	0.725***	0.504***
	(0.037)	(0.058)	(0.021)
2* ("internationally recognised")			0.423***
			(0.017)
1* ("nationally recognised")			0.073***
			(0.011)
Number of journals in tier	48	31	205
Number of publications in tier	784	551	1,265
Mean of standardized AIS in tier	2.57	0.76	-0.09

Note. The Table shows estimation results when using the alternative classification of research outputs not in scientific journals (see columns 4-6 of Table C.1). Tier is defined as in Section 5, from the Journal of Health Economics (JHE) to the Economic Journal (EJ). Specifications and table presentation are analogue to columns (4)-(6) of Table 3. See Appendix C for details. *** p<0.01, ** p<0.05, * p<0.1.

$$\alpha_j^d = \alpha_{0\tau}^d + \alpha_{1\tau}^d W_j(1 - B_j - C_j - P_j) + \alpha_{2\tau}^d B_j + \alpha_{3\tau}^d C_j + \alpha_{4\tau}^d P_j,$$

and W_j is set to zero for outputs without AIS. Define:

$$X_{i\tau} \equiv \left(\sum_{j\in\tau} X_{ij} \right), \qquad WX_{i\tau} \equiv \left(\sum_{j\in\tau} W_j S_j X_{ij} \right), \qquad BX_{i\tau} \equiv \left(\sum_{j\in\tau} B_j X_{ij} \right),$$

$$CX_{i\tau} \equiv \left(\sum_{j\in\tau} C_j X_{ij} \right), \qquad PX_{i\tau} \equiv \left(\sum_{j\in\tau} P_j X_{ij} \right), \qquad Z_{i\tau} \equiv \left(\sum_{j\in\tau} \sum_{k=1}^{X_{ij}} Z_{jk} \right),$$

where $S_j \equiv (1 - B_j - C_j - P_j)$ is an indicator for outputs in scientific journals. Let K and H be the number of regressors in $Z_{i\tau}$ and $WX_{i\tau}$, respectively, with a slight abuse of notation to avoid the use of matrices.

The restrictions on the classification probabilities in Table 2 and Table C.1 reduce the number of unknown parameters. Since publications in tier 1 and tier 2 journals are allowed to score three stars or more, $\alpha_{i\tau}^d = \gamma_\tau^d = 0$ for $\tau = 1, 2$; $d = 1, 2$; and $i = 0, \ldots, 3$. In addition, a publication in tier 4 journals cannot be awarded four stars, implying $\alpha_{i3}^4 = \gamma_3^4 = 0$ for $i = 0, 1, 3, 4$. As a result, we estimate the following system of equations in $7 \times (K + H + 1)$ unknowns, plus additional 12 unknown parameters on output type dummies:

$$E\left[Y_i^4|\mathbf{X}_i,\mathbf{Z}_i\right]=\sum_{\tau=1}^{2}\alpha_{0\tau}^4X_{i\tau}+\sum_{\tau=1}^{2}\alpha_{1\tau}^4WX_{i\tau}+\sum_{\tau=1}^{2}\alpha_{2\tau}^4BX_{i\tau}+\alpha_{32}^4CX_{i\tau}+\sum_{\tau=1}^{2}\gamma_\tau^4Z_{i\tau},$$

$$E\left[Y_i^3|\mathbf{X}_i,\mathbf{Z}_i\right]=\sum_{\tau=1}^{3}\alpha_{0\tau}^3X_{i\tau}+\sum_{\tau=1}^{3}\alpha_{1\tau}^3WX_{i\tau}+\sum_{\tau=1}^{2}\alpha_{2\tau}^3BX_{i\tau}+$$
$$+\sum_{\tau=2}^{3}\alpha_{3\tau}^3CX_{i\tau}+\alpha_{43}^3PX_{i\tau}+\sum_{\tau=1}^{3}\gamma_\tau^3Z_{i\tau},$$

$$E\left[Y_i^2|\mathbf{X}_i,\mathbf{Z}_i\right]=\alpha_{03}^2X_{i\tau}+\alpha_{13}^2WX_{i\tau}+\alpha_{33}^2CX_{i\tau}+\alpha_{43}^2PX_{i\tau}+\gamma_3^2Z_{i\tau},$$

$$E\left[Y_i^1|\mathbf{X}_i,\mathbf{Z}_i\right]=\alpha_{03}^1X_{i\tau}+\alpha_{13}^1WX_{i\tau}+\alpha_{33}^1CX_{i\tau}+\alpha_{43}^1PX_{i\tau}+\gamma_3^1Z_{i\tau}.$$

Finally, the probabilities of being awarded each possible number of stars must add up to one. We model this constraint by restricting tier intercepts to total 1 and coefficients on output characteristic to sum up to zero. The system is estimated from seemingly unrelated regressions imposing the following $3\times(K+H+1)$ constraints, and 5 additional constraints that involve the coefficients on the output type dummies:

$$\alpha_{0\tau}^3+\alpha_{0\tau}^4=1,\qquad \tau=1,2,$$

$$\alpha_{i\tau}^3+\alpha_{i\tau}^4=0,\qquad \tau=1,2,\qquad i=1,2,3,$$

$$\gamma_\tau^3+\gamma_\tau^4=0,\qquad \tau=1,2,$$

$$\alpha_{03}^1+\alpha_{03}^2+\alpha_{03}^3=1,$$

$$\alpha_{i3}^1+\alpha_{i3}^2+\alpha_{i3}^3=0,\qquad i=1,3,4,$$

$$\gamma_3^1+\gamma_3^2+\gamma_3^3=0.$$

After imposing all constraints, tier-specific intercepts and regressors in $Z_{i\tau}$ and $WX_{i\tau}$ yield $4\times(K+H+1)$ unknowns; 7 additional unknowns arise from output type dummies. Our baseline specification estimates only tier-level intercepts, imposing $\gamma_\tau^d=0$ and $\alpha_{i\tau}^d=0$ for all τ's and d's and for $i=0,\ldots,4$. The regression adjustment includes tier-specific linear and quadratic terms in AIS. In the latter case, two additional sets of constraints are

Table D.1: Number of unknown parameters

	(1) N. of W_j	(2) N. of Z_{jk}	(3) Output type dummies	(4) N. of parameters
Baseline specification	0	0	N	4
Number of parameters after adding:				
Article Influence Score	2	0	Y	19
Citations	2	1	Y	23
H-index	2	2	Y	27
Dummy for Economics outputs	3	2	Y	31

Note. The table presents the number of parameters estimated in each specification considered. Column (1) shows the number of regressors in the journal characteristics vector, denoted H in Appendix D. Column (2) reports the number of regressors in the publication characteristics vector, denoted K in Appendix D. The latter number is 2 when adjusting for AIS since a quadratic term is included. Column (3) indicates whether dummies for output type (e.g., book, book chapter) are included. Column (4) reports the total number of unknowns, equal to 4(K+H+1), plus 7 when output type dummies are included. This follows from summing the conditions for classification probabilities and our classification restrictions, as discussed in Appendix D.

imposed to the estimation. First, continuity of classification probabilities is forced across tiers boundaries. Formally, we impose:

$$\alpha^4_{01} + \alpha^4_{11} W_1^{min} = \alpha^4_{02} + \alpha^4_{12} W_2^{max},$$

$${}^{4}_{02} + \alpha^4_{12} W_2^{min} = 0,$$

$$\alpha^3_{03} + \alpha^3_{13} W_3^{max} = 1,$$

where W_τ^{min} and W_τ^{max} are the lowest and highest value of AIS, respectively, in tier τ. The number of parameters estimated in the various specifications is shown in Table D.1. Second, we impose that the two top journals in economics (according to AIS) among ones frequently submitted (see Table 1) are deterministically awarded four stars.

E Sensitivity of results to the definition of Tier 2

This section tests the sensitivity of our conclusions to alternative definitions of journal tiers to the one employed to derive our main results in Section 6. We start by selecting the tier definition that best approximates the expected classification. We determine the optimal width of Tier 2 by means of a grid search over 60×60 possible choices of journals obtained by varying the upper and lower limits. Specifically, we begin by setting the lower limit on Tier 2 at the JHE. We then select 60 journals with an AIS in a window centred on the EJ and use them to define alternative upper limits on Tier 2. This defines a range of 60 possible intervals between the JHE and the new upper limit, which we use iteratively to estimate our model. We then select the definition of Tier 2 that yields estimates at the minimum distance from the following constraints:

$$p_2(3) = 1, \qquad p_3(3) = 0,$$

where $p_\tau(d)$ is the probability of a d-starred publication in tier τ. In words, Tier 2 is defined to maximise between-tier distance in classification probabilities while ensuring within-tier homogeneity. This procedure is replicated by replacing the JHE with 60 journals falling in a window around its AIS. Figure E.1 shows that the distance from the constraints is minimised when the upper limit on Tier 2 is the Journal of Econometrics, below the EJ, and the lower limit on Tier 2 is the Journal of Economic Dynamics and Control, below the JHE.[2]

Table E.1 presents estimates under this choice of Tier 2. As expected, columns (1) to (3) show that outputs in Tier 3 have little chance of being awarded three stars. Estimates in column (1) across the two panels suggest that publications in journals to the left of the EJ may not be consistently awarded four stars, as the probability of top-scoring outputs in Panel B drops to 80%. Columns (4) to (6) show substantially similar conclusions after adjusting for AIS heterogeneity. Overall, results convey a similar interpretation than our main estimates in Table 3.

The sensitivity of our conclusions to the choice of Tier 2 is further explored with the aid of a graphical analysis. The definition of Tier 2 is obtained by varying the upper limit while leaving the lower limit at the JHE. Figure E.2 presents the estimated probabilities of four-star outputs in Tier 1 and Tier 2. The value zero on the horizontal axis corresponds to the definition of Tier 1 in Panel A of Table 3, and the sensitivity of results to deviations below

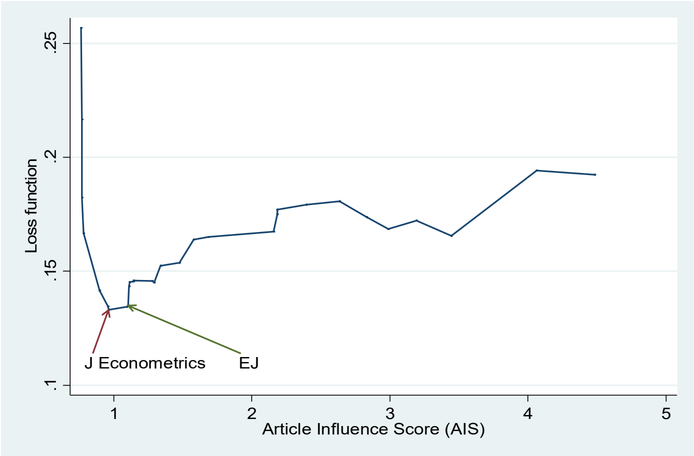
[2]Calculations available on request show that the Journal of Econometrics is always the optimal upper limit for all lower limits in our grid. Figure E.1 reports values of the following quantity:

$$\left\{[p_2(3) - 1]^2 + Var\left[p_2(3)\right]\right\} + \left\{[p_3(3)]^2 + Var\left[p_3(3)\right]\right\},$$

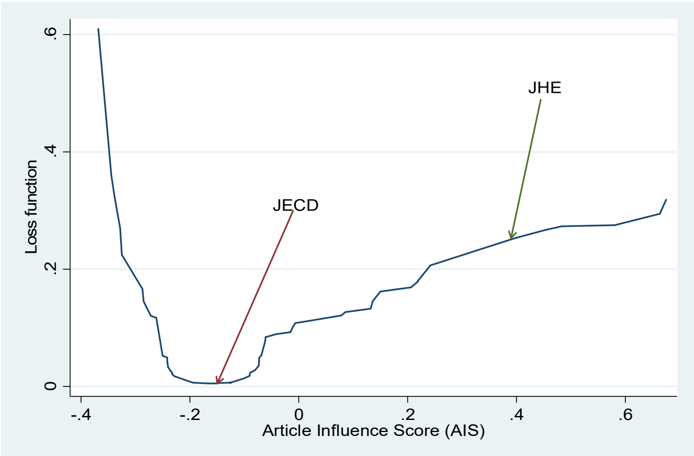
where the $p_\tau(d)$s are estimated from (5) when imposing $\gamma_\tau^d = 0$ and $\alpha_{1\tau}^d = 0$ for all τs and for all ds. Panel A shows the average value of this quantity over the 60 possible choices for the lower limit on Tier 2. Panel B reports the value of this quantity when the upper limit on Tier 2 is the Journal of Econometrics.

Figure E.1: Optimal definition of Tier 2

Panel A. Tier 2 upper limit



Panel B. Tier 2 lower limit



Note. The figure is obtained by using a grid search over 60X60 possible choices and varying the upper and lower limits of Tier 2. We start by setting the lower limit on Tier 2 at JHE. We then select 60 journals with AIS in a window centred on the EJ and use them to define alternative upper limits on Tier 2. This defines a range of 60 possible intervals between JHE and the new upper limit, which we use iteratively to estimate our classification probabilities. We then select the definition of Tier 2 to maximize between-tier distance in classification probabilities while ensuring within-tier homogeneity. Panel A reports the loss function resulting from different choices of the upper limit on Tier 2. Panel B reports the loss function for the lower limit. See Appendix E for details.

Table E.1: Estimation results with optimised Tier 2

	Without adjustment			With adjustment (AIS)		
	Tier 1	Tier 2	Tier 3	Tier 1	Tier 2	Tier 3
	(1)	(2)	(3)	(4)	(5)	(6)
4* ("world leading")	0.799***	0.037		0.714***	0.086***	
	(0.029)	(0.035)		(0.024)	(0.024)	
3* ("internationally excellent")	0.201***	0.963***	0.044	0.286***	0.914***	0.098***
	(0.029)	(0.035)	(0.041)	(0.024)	(0.024)	(0.034)
2* ("internationally recognised")			0.811***			0.770***
			(0.037)			(0.029)
1* ("nationally recognised")			0.145***			0.132***
			(0.019)			(0.021)
Number of journals in tier	53	88	142	53	88	142
Number of publications in tier	888	1,067	645	888	1,067	645
Mean of standardized AIS in tier	2.37	0.33	-0.31	2.37	0.33	-0.31

Note. Columns (1) to (3) show results from a baseline specification that allows for tier-specific intercepts. Columns (4) to (6) show results from regressions that control for within-tier heterogeneity using a quadratic polynomial in Article Influence Score (AIS). Tier 2 is defined to maximize the probability of including three-star journals (see Appendix E for details). Columns (1) and (4) show the estimated probabilities that a publication in Tier 1 is awarded four or three stars. Columns (2) and (5) show the estimated probabilities that a publication in Tier 2 is awarded four or three stars. Columns (3) and (6) show the estimated probabilities that a publication in Tier 3 is awarded three stars, two stars or one star. The estimating equations are discussed in Section 5 and Appendix D. *** p<0.01, ** p<0.05, * p<0.1.

(negative values) and above the EJ (positive values) is investigated. The value -5 corresponds to the definition of Tier 1 in Panel B of Table 3 (from the Journal of Econometrics). We find that the probability of four-star outputs in Tier 1 quickly grows to one by moving to the right of the EJ. This probability also grows in Tier 2, which implies that a positive density of top-scoring outputs exists at or above the EJ.

F Additional robustness checks

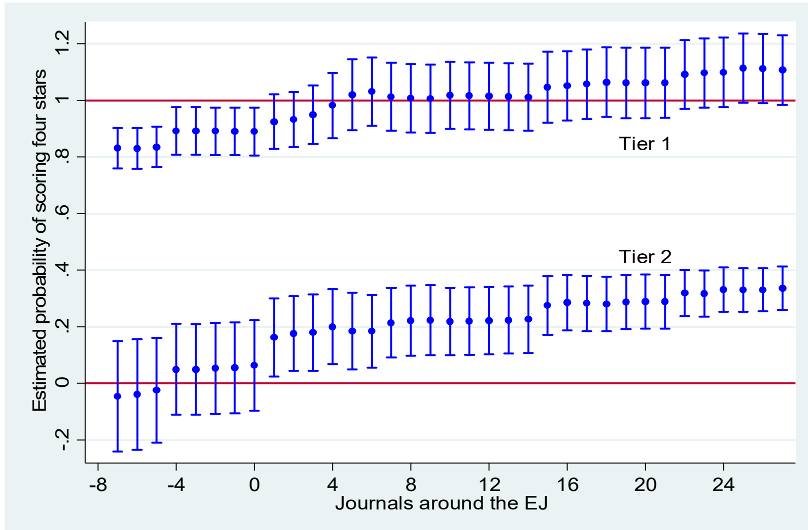
Results are not driven by unobserved traits

Consider the following version of equation (2):

$$E\left[\mathbb{1}(D_{jk} = d)|\mathbf{X}_i, \mathbf{Z}_i, \boldsymbol{U}_i\right] = \alpha_j^d + \gamma_j^d Z_{jk} + \delta_j^d U_{jk},$$

where $\mathbf{U}_i'$ represents the vector of attributes, unobservable to us, of all publications submitted by institution i. The elements of this vector are U_{jk}. The equation implies that the classification of output k depends on attributes we can observe (e.g., citations) Z_{jk} and a latent indicator of quality U_{jk} assessed by panellists. The exclusion restriction in 2 is main-

Figure E.2: Sensitivity analysis for the definition of Tier 2



Note. The figure shows estimated probabilities of scoring four stars by journal tier as a function of Tier 2 cut-offs, starting from journal tiers defined in Section 5. The 95% confidence interval is plotted for each estimate. Presented is the sensitivity of the results to shifting the Tier 2 upper bound around the Economic Journal (EJ), when the lower bound is the Journal of Health Economics (JHE). The lower bound of Tier 1 is the journal ranked x positions above the EJ in the AIS distribution. See Appendix E for details.

tained, stating that each output is classified independently of the bundle submitted by the institution. We are interested in understanding to what extent our conclusions are affected by the omitted variable U_{jk}. Specifically, we maintain the assumption that within the same journal a publication with many citations cannot be awarded a lower REF score than a publication with few citations (citation counts are standardised by year of publication and field). Under this assumption, unobservable quality detected by panellists must be non-decreasing in citations.

If the latter variable is independent of other outputs conditional on Z_{jk}, one can write:

$$E\left[\mathbb{1}(D_{jk}=d)|\mathbf{X}_i,\mathbf{Z}_i\right]=\alpha_j^d+\gamma_j^d Z_{jk}+\delta_j^d E\left[U_{jk}|Z_{jk}\right], \tag{7}$$

which clarifies the effects of omitted variables in equation (2). The important message from equation (7) is that the estimates in Table G.6 are robust to output's unobserved quality if the latter increases linearly with citations. More generally, the unobservable component U_{jk} will introduce non-linearities into the relationship between Z_{jk} and the classification probabilities. To see this, consider the following parametric relationship between unobserved quality and citations:

$$E\left[U_{jk}|Z_{jk}\right] = \rho_{0j} + \rho_{1j}Z_{jk} + \rho_{2j}Z_{jk}^2. \tag{8}$$

When this control function is linear, substituting into (7) and re-arranging terms yields new intercept and slope parameters that are a combination of α_j^d, γ_j^d, $\delta_j^d\rho_{0j}$ and $\delta_j^d\rho_{1j}$. As we are not interested in disentangling the values of these parameters, the case of unobservables linear in Z_{jk} is already embedded in the specifications considered above. It also follows that (7) will contain a quadratic term in Z_{jk} when $\rho_{2j} \neq 0$.

The classification probabilities are substantially unchanged after adding a quadratic term in citations to the estimating equations, as shown in Table F.1.[3] Results in columns (1)-(3) and (4-6) can be compared to columns (1)-(3) and (7)-(9) of Table G.6, respectively. This finding suggests that the main conclusions are unlikely to be driven by output unobservables. The effects of citations, available on request, are remarkably similar to the main results presented in Figure 5. The only slight difference is that the contribution of citations in Tier 3 is larger (1.3%) and statistically significant at the 5% level, reinforcing our conclusion that publication attributes other than journal impact only affect the classification of outputs in journals with low impact.

Model predicts institutions' performance

We compute differences between the number of outputs awarded d stars and the same number predicted by our model with linear controls (fitted values) for output's citations and h-index. These differences are uncorrelated, by construction and standard results of OLS regression theory, with all bibliometric indicators included in equation (5). Here, we study the correlation with other measures of research excellence at an institution to corroborate the validity of the assumptions underlying our analysis. We predict for the 28 institutions the number of outputs awarded d stars, $\hat{Y}_i^d$, and subtract it from the number observed, Y_i^d, for $d = 1, \ldots, 4$. We then consider the following equation defined from 28×4 observations:

$$Y_i^d - \hat{Y}_i^d = \mu_d + \xi S_i + \eta_i,$$

[3]Here the model with a linear control function in Table G.6 is augmented to include a quadratic polynomial in citations. The coefficients in Table F.1 are parametrised to yield the classification probabilities of outputs with the average citation count in a tier.

Table F.1: Estimation results adjusting for unobserved quality

	Without adjustment			With adjustment (AIS)		
	Tier 1	Tier 2	Tier 3	Tier 1	Tier 2	Tier 3
	(1)	(2)	(3)	(4)	(5)	(6)
4* ("world leading")	0.799***	0.037		0.714***	0.086***	
	(0.029)	(0.035)		(0.024)	(0.024)	
3* ("internationally excellent")	0.201***	0.963***	0.044	0.286***	0.914***	0.098***
	(0.029)	(0.035)	(0.041)	(0.024)	(0.024)	(0.034)
2* ("internationally recognised")			0.811***			0.770***
			(0.037)			(0.029)
1* ("nationally recognised")			0.145***			0.132***
			(0.019)			(0.021)
Number of journals in tier	53	88	142	53	88	142
Number of publications in tier	888	1,067	645	888	1,067	645
Mean of standardized AIS in tier	2.37	0.33	-0.31	2.37	0.33	-0.31

Note. Columns (1) to (3) show results from a baseline specification that allows for tier-specific intercepts. Columns (4) to (6) show results from regressions that control for within-tier heterogeneity using a quadratic polynomial in Article Influence Score (AIS). Tier 2 is defined to maximize the probability of including three-star journals (see Appendix E for details). Columns (1) and (4) show the estimated probabilities that a publication in Tier 1 is awarded four or three stars. Columns (2) and (5) show the estimated probabilities that a publication in Tier 2 is awarded four or three stars. Columns (3) and (6) show the estimated probabilities that a publication in Tier 3 is awarded three stars, two stars or one star. The estimating equations are discussed in Section 5 and Appendix D. *** p<0.01, ** p<0.05, * p<0.1.

where S_i is a vector of institution-level characteristics. The equation includes dummies for the number of stars awarded, the indicator of research quality available from the previous national evaluation round (RAE 2008), and the REF scores of non-academic impact and research environment (RAE stars were awarded at the department level). The value $\hat{Y}_i^d$ is computed by considering the model specification in the previous section – see columns (4) to (6) of Table F.1. Standard errors are clustered on institutions, and the results are reported in Table F.2.

We also consider the correlation with indicators of academic ties between REF panellists and institutions. Past research demonstrated that the professional network plays an important role in academic publications and promotions (see Colussi, 2018; see also the "incest factor" in Heckman and Moktan 2020). We therefore scraped the web for data on the current and past affiliations of REF panellists and their co-authors and computed two proxies for the connections between panellists and the institution. Column (2) of Table F.2 controls for a dummy equal to one if at least one panellist was ever employed at the institution, interacting this variable with dummies for the number of stars awarded. Column (3) considers the share of panellists or panellists' co-authors ever employed at the institution (see below

Table F.2: Differences between predicted quality and REF outcomes

	(1)	(2) Dummy for representation in the panel	(3) Panellists' co-authors affiliated
Dummy for one-star publications	0.002	0.020	-0.008
	(0.029)	(0.033)	(0.035)
Dummy for two-stars publications	-0.002	0.057	-0.002
	(0.020)	(0.038)	(0.020)
Dummy for three-stars publications	0.003	0.027	0.015
	(0.029)	(0.042)	(0.041)
Institution quality profile in RAE 2008	-0.000	-0.001	-0.001
	(0.001)	(0.001)	(0.001)
Institution impact profile in REF 2014	-0.000	-0.000	-0.000
	(0.000)	(0.000)	(0.000)
Institution environment profile in REF 2014	0.001	0.001	0.001
	(0.000)	(0.000)	(0.000)
Relationship of institutions with panellists x dummy for one-star publications		-0.012	-0.004
		(0.017)	(0.005)
Relationship of institutions with panellists x dummy for two-stars publications		-0.053*	-0.012
		(0.030)	(0.010)
Relationship of institutions with panellists x dummy for three-stars publications		0.028	0.007
		(0.030)	(0.010)
Relationship of institutions with panellists x dummy for four-star publications		0.038**	0.005
		(0.018)	(0.014)
Constant	-0.003	-0.012	0.008
	(0.026)	(0.024)	(0.032)
Observations	112	112	112
R-squared	0.049	0.142	0.063

Note. The table reports the results of OLS regressions using, on the left hand side, the difference between the number of outputs awarded d stars and the number predicted by our model. The latter quantity is obtained from a model that adjusts for AIS and publication characteristics – see columns (7)-(9) of Table G.6. Column (1) includes number-of-stars dummies, and indicators of the research performance of institutions in the REF and in the 2008 RAE. Columns (2) and (3) add measures of connection to experts on the panel interacted with dummies for number of stars. Column (2) considers a dummy equal to one if at least one panellist was ever employed at the institution. Column (3) considers panellists' co-authors ever employed at the institution. Standard errors are clustered at the institution level. See Appendix F for details. *** p<0.01, ** p<0.05, * p<0.1.

for the definition of these variables).[4]

The predictions of our models do not depart systematically from the actual number of outputs at any level of awarded quality. The differences between observed and predicted REF scores are uncorrelated with the indicators of research performance and institutional ties with the panel. The coefficients in the table are small and fairly precise zeros, with only one coefficient in column (3) being marginally significant.

[4] Pitt and Yan (2017) find that outputs published in the same journal contribute to the final count of four-star publications depending on the submitting institution. The variables on networks here are used to model this channel, although the findings in Pitt and Yan (2017) may be driven by output attributes unobservable to them that are controlled for in our analysis.

Academic networks

We scraped from the web curricula of the Economics and Econometrics panel members, ignoring assessors and secretariat staff who joined the panel near the submission deadline. The panel comprised 18 members, including a chair and a deputy chair. We determined the academic network by considering institutions where panel members and their co-authors were appointed throughout their professional career. We assigned to each member the institutions with which she had a professional position. For all co-authors, we then retrieved their affiliations and number of articles written with the panellists. We used this information to define various indicators of academic ties.

First, we compute the number of panellists ever employed by institution i. Define a dummy E_{ip} that takes value one if institution i has employed or is now employing panellist p. The index C_i is then computed as the sum of these dummies over the 18 panellists:

$$C_i = \sum_{p=1}^{18} E_{ip}.$$

Second, we compute the number of panellists' co-authors ever employed by institution i. Define N_{ip} as the number of co-authors of panellist p affiliated with institution i.[5] The index C_i^{coaut} is then computed as the sum over the 18 panellists:

$$C_i^{coaut} = \sum_{p=1}^{18} N_{ip}.$$

G Additional figures and tables

Table G.1 lists the departments involved in the REF and reports the evaluation results (see Section 2). Table G.2 reports descriptive statistics on research outputs submitted to the REF (see Section 3). Figure G.1 plots the distribution of standardised AIS (see Section 3). Figure G.2 replicates Panel A of Figure 1 restricting publications in top-five economic journals to be awarded four stars (see Section 4). Table G.3 reports estimates of the association between citation counts and AIS (see Section 4). Table G.4 reports estimates of the association between institution-level REF results and bibliometrics of research outputs submitted (see Section 4).

[5] Alternatively, one could consider the frequency of each co-author in panellists' scientific production or take into account the different sizes of networks across panellists. Conclusions are robust to these alternative definitions of ties.

Table G.5 reports the comparison between journal tiers selected here and the classification in Hudson (2013, see Section 5). Figure G.3 plots predicted likelihood of scoring three and four stars using our estimates in columns (4)-(6) of Table 3 in the main text (see Section 6). Table G.6 present estimation results when controlling for publication characteristics other than journal impact (see Section 6).

Table G.1: Research Excellence Framework (REF) outcomes

Institution	(1) Outputs	(2) % 4*	(3) % 3*	(4) % 2*	(5) % 1*	(6) % n.c.	(7) FTE staff	(8) % FTE submitted	(9) GPA	(10) QI	(11) PR
University College London	142	69.7	28.2	2.1	0	0	36.9	82.0	3.68	76.75	62.92
London School of Economics and Political Science	183	56.3	33.3	4.9	0.6	4.9	51.4	91.8	3.36	64.63	73.80
University of Cambridge	99	54.5	39.4	5.1	1	0	27	71.1	3.47	64.35	38.60
University of Warwick	136	42.6	50.8	6.6	0	0	41.6	80.0	3.36	55.30	51.11
University of Oxford	242	42.6	44.2	11.1	2.1	0	83.9	86.5	3.27	53.65	100.00
Royal Holloway, University of London	51	35.3	45.1	15.7	1.9	2	14.4	62.6	3.10	46.58	14.90
University of Edinburgh	55	30.9	54.6	12.7	1.8	0	17.5	62.5	3.15	44.55	17.32
University of Essex	113	29.2	60.2	10.6	0	0	33.33	83.3	3.19	44.25	32.77
University of Surrey	71	26.8	52.1	21.1	0	0	20.65	82.6	3.06	39.83	18.27
University of East Anglia	49	20.4	71.4	8.2	0	0	14	63.6	3.12	38.25	11.90
University of St Andrews	51	23.5	58.9	17.6	0	0	20.5	66.1	3.06	38.23	17.41
University of Bristol	63	22.2	58.8	19	0	0	18.6	74.4	3.03	36.90	15.25
University of Nottingham	127	19.7	65.3	14.2	0	0.8	35	76.1	3.03	36.03	28.01
Queen Mary University of London	94	20.2	62.8	15.9	1.1	0	24.45	78.9	3.02	35.90	19.50
University of Glasgow	83	18.1	61.4	18.1	2.4	0	23.75	79.2	2.95	33.45	17.65
University of Southampton	82	22	37.8	35.3	3.7	1.2	21.8	77.9	2.76	31.45	15.23
University of Leicester	80	18.8	50	28.7	0	2.5	22.4	77.2	2.83	31.30	15.58
University of York	104	14.4	59.6	24.1	1.9	0	28.07	61.0	2.87	29.30	18.27
University of Exeter	83	13.3	57.8	19.3	9.6	0	24.5	79.0	2.75	27.75	15.10
University of Sussex	54	14.8	46.3	35.2	1.8	1.9	17.4	72.5	2.70	26.38	10.20
City University London	54	16.7	37	29.6	16.7	0	13.7	52.7	2.54	25.95	7.90
University of Manchester	114	11.4	53.5	30.7	4.4	0	33.2	73.8	2.72	24.78	18.27
University of Birmingham	79	7.6	58.2	32.9	1.3	0	24.2	89.6	2.72	22.15	11.91
Birkbeck College	97	10.3	47.4	37.1	4.2	1	25.15	78.6	2.62	22.15	12.38
University of Sheffield	50	8	56	36	0	0	14.9	57.3	2.72	22.00	7.28
University of Aberdeen	63	4.8	50.8	30.1	14.3	0	19.25	80.2	2.46	17.50	7.48
University of Kent	79	2.5	43.1	37.9	16.5	0	21.9	84.2	2.32	13.28	6.46
Brunel University London	102	2	22.5	63.7	11.8	0	26.2	90.3	2.15	7.63	4.44

Note. The table reports selected results from official REF publications for the Economics and Econometrics sub-panel (see https://www.ref.ac.uk/2014/). Column (1) reports the number of research outputs submitted. Columns (2) to (6) show the quality breakdown of submissions for the 28 institutions involved (by the number of stars awarded and outputs not classified). Columns (7) and (8) present number and share of full-time equivalent (FTE) staff members submitted, respectively (source: Higher Education Statistic Agency). Columns (9) to (11) show the Grade Point Average (GPA), the Quality Index (QI) and the Power Rating (PR), respectively. The Grade Point Average (GPA), in column (9), is calculated by multiplying the percentage of research in each group by its rating, adding them all together and dividing by 100. The Quality Index (QI) in column (10) is a weighted average reflecting the current funding allocation formula, which depends on the incidence of top-quality outputs (80% and 20% for four- and three-star research, respectively, and no contribution from the remaining outputs; see HEFCE, 2016). "Research power" in column (11) by adjusting awarded quality for the fraction of full-time equivalent faculty members submitted. The QI and power rating were developed by Research Fortnight. Here, we present our own calculations based on the current funding formula and on the outputs sub-profile only. See Section 2 for details.

Table G.2: Descriptive statistics for research outputs

	(1) Mean	(2) Std. Dev.
Panel A. Characteristics of Outputs Submitted		
Publication Type:		
Journal	0.9185	0.2737
Book Chapter or Proceedings	0.0046	0.0678
Book	0.0115	0.1068
Other	0.0654	0.2473
Field:		
Economics	0.9378	0.2415
Statistics	0.0941	0.2920
Finance	0.0787	0.2694
Mathematics	0.0472	0.2122
Missing field	0.0719	0.2584
Authors:		
Number	2.2794	0.9319
H-index (at submission)	9.1165	5.9447
Panel B. Bibliometrics (December 2013)		
From Thomson Reuter's Journal Citation Reports:		
Article Influence Score (AIS)*	3.0800	2.9037
Missing AIS	0.0858	0.2801
Citations:		
From Elsevier's Scopus *	6.5834	14.9900
From Google Scholar *	29.1400	63.0600
Not in Elsevier's Scopus	0.0612	0.2397
Not in Google Scholar	0.0012	0.0340
Total number of submissions	2,600	

Note. The table presents descriptive statistics for all submissions in Economics and Econometrics. Panel A shows the breakdown by type, field and number of authors. Panel B reports the bibliometric indicators considered for the analysis: Article Influence Score, citation counts from Elsevier's Scopus, citation counts from Google Scholar and the h-index of authors. See Section 3 for details and definitions. *: conditional on non-missing data.

Table G.3: Relationship between citation count and Article Influence Score

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Elsevier`s Scopus Citations				Google Scholar Citations			
	Citations	Asinh(citations)	Citations	Asinh(citations)	Citations	Asinh(citations)	Citations	Asinh(citations)
AIS	0.164***	0.269***	0.132***	0.227***	0.207***	0.434***	0.176***	0.386***
	(0.030)	(0.047)	(0.026)	(0.045)	(0.038)	(0.076)	(0.035)	(0.075)
H-index			0.206***	0.271***			0.201***	0.308***
			(0.044)	(0.018)			(0.044)	(0.025)
Constant	0.755	3.001***	0.789	3.045***	0.045	3.932***	0.081	3.987***
	(0.735)	(0.592)	(0.806)	(0.679)	(0.290)	(0.405)	(0.358)	(0.501)
Field Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2,349	2,349	2,349	2,349	2,377	2,377	2,376	2,376
R-squared	0.249	0.536	0.285	0.570	0.174	0.435	0.207	0.468
Method	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS

Note. The table shows regressions from publication-level data that control for field and year of publication. All equations consider citations as the left-hand-side variable, using Elsevier's Scopus in columns (1)-(4) and Google Scholar in columns (5)-(8). In columns (1), (3), (5) and (7), citations are standardized to have zero mean and unit variance in the sample. In columns (2), (4), (6) and (8), the dependent variable is the inverse hyperbolic sine of citations, as in Card and Della Vigna (2017). Columns (1), (2), (5) and (6) consider specifications that control for Article Influence Score, standardized by field. Columns (3), (4), (7) and (8) add the average h-index of authors at the time of REF submission, standardized to have zero mean and unit variance in the sample. Standard errors are clustered on the journal. See Section 4 for details. *** p<0.01, ** p<0.05, * p<0.1.

Table G.4: REF Quality Index and bibliometrics

	Quality					
	Elsevier`s Scopus Citations			Google Scholar Citations		
	(1)	(2)	(3)	(4)	(5)	(6)
Citations	5.665***	5.405***	-0.422	1.045***	0.958***	-0.284
	(0.842)	(0.966)	(1.009)	(0.185)	(0.195)	(0.257)
H-index		0.596	1.787		1.394	2.025**
		(2.590)	(1.055)		(1.951)	(0.862)
AIS			21.484***			24.900***
			(2.174)			(3.746)
Constant	39.705***	34.230	2.231	39.997***	27.244	-4.807
	(2.292)	(22.517)	(10.433)	(2.092)	(17.344)	(10.059)
Observations	28	28	28	28	28	28
R-squared	0.569	0.570	0.889	0.663	0.672	0.897

Note. The table shows regressions from institution-level data that control for field and year of publication. All equations consider an institution's Quality Index as the left-hand-side variable and average citations of an institution's submitted publications as explanatory variable, using Elsevier's Scopus in columns (1)-(3) and Google Scholar in columns (4)-(6). Citations are computed as residuals from publication-level regressions of citation counts on year and field of publication fixed effects. Columns (2) and (5) add the average h-index of authors of an institution's submitted publications at the time of REF submission. Columns (3) and (6) add the Article Influence Score, standardized by field. See Section 4 for details. *** p<0.01, ** p<0.05, * p<0.1.

Table G.5: Comparison with the classification in Hudson (2013)

	(1) Tier 1	(2) Tier 2	(3) Tier 3
4*	73.33%	13.33%	13.33%
probable 4*	66.67%	33.33%	0.00%
possible 4*	50.00%	33.33%	16.67%
3*	21.95%	26.83%	51.22%
probable 3*	0.00%	14.29%	85.71%
possible 3*	0.00%	11.11%	88.89%
2*	0.00%	2.56%	94.87%
probable 2*	0.00%	0.00%	100.00%
possible 2*	0.00%	0.00%	100.00%
1*	0.00%	0.00%	100.00%

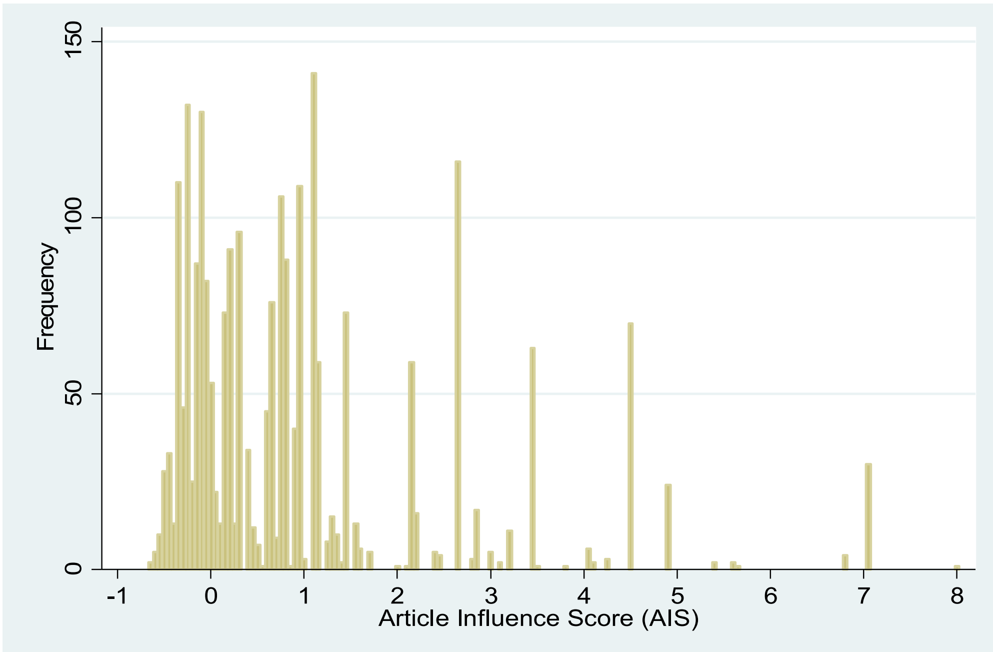
Note. The table shows the comparison between journal tiers defined in Section 5 and the classification in Hudson (2013). Cells report the share of journals in Tier 1, Tier 2 and Tier 3 for each of Hudson's (2013) categories. For example, 73.33% of journals classified as unambiguously four stars in Hudson (2013) belong to our Tier 1—see column (1); all journals classified as unambiguously one star belong to our Tier 3—see column (3). See Section 5 for details.

Table G.6: Estimation results controlling for publication characteristics

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
		Tier			Tier			Tier			Tier	
	1	2	3	1	2	3	1	2	3	1	2	3
4* ("world leading")	0.803***	0.161**		0.768***	0.231***		0.775***	0.218***		0.729***	0.257***	
	(0.045)	(0.067)		(0.036)	(0.058)		(0.041)	(0.063)		(0.036)	(0.053)	
3* ("internationally excellent")	0.197***	0.839***	0.503***	0.232***	0.769***	0.514***	0.225***	0.782***	0.511***	0.271***	0.743***	0.507***
	(0.045)	(0.067)	(0.020)	(0.036)	(0.058)	(0.020)	(0.041)	(0.063)	(0.019)	(0.036)	(0.053)	(0.018)
2* ("internationally recognised")			0.426***			0.417***			0.421***			0.423***
			(0.017)			(0.017)			(0.017)			(0.017)
1* ("nationally recognised")			0.071***			0.069***			0.068***			0.070***
			(0.011)			(0.011)			(0.010)			(0.010)
Google Scholar citations	Y	Y	Y	N	N	N	Y	Y	Y	Y	Y	Y
H-index of authors	N	N	N	Y	Y	Y	Y	Y	Y	Y	Y	Y
Economic journal indicator	N	N	N	N	N	N	N	N	N	Y	Y	Y
Number of journals in tier	48	31	205	48	31	205	48	31	205	48	31	205
Number of publications in tier	784	551	1,265	784	551	1,265	784	551	1,265	784	551	1,265
Mean of standardized AIS in tier	2.57	0.76	-0.09	2.57	0.76	-0.09	2.57	0.76	-0.09	2.57	0.76	-0.09

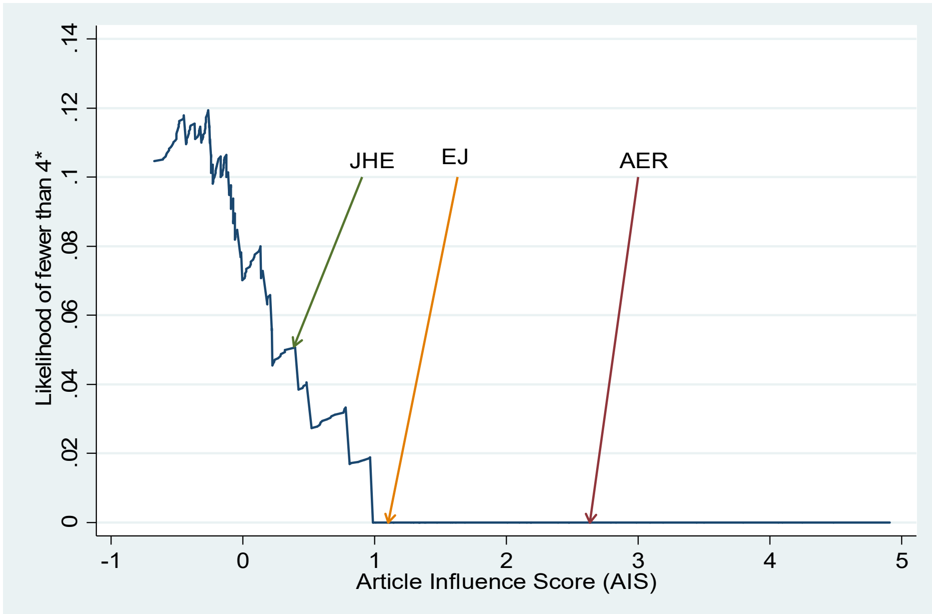
Note. The table shows estimation results when adding different publication characteristics to the baseline specification in columns (4)-(6) in Table 2. Columns (1) to (3) include standardized Google Scholar citation count. Columns (4)-(6) includes the highest h-index among authors of a research output. Columns (7)-(9) combine all variables included in the previous columns. Columns (10)-(12) add an indicator for publication in an economic journal. The estimating equations are discussed in Section 5 and in the Appendix, and include linear controls for the additional publication caracteristics considered. *** p<0.01, ** p<0.05, * p<0.1.

Figure G.1: Article Influence Score distribution



Note. The figure shows the distribution of submissions by the value of the Article Index Score (AIS) from the 2013 Thomson Reuters Journal Citation Reports database. Only outputs published in academic journals are considered. AIS is standardized to have zero mean and unit variance by research field. See Section 3 for details.

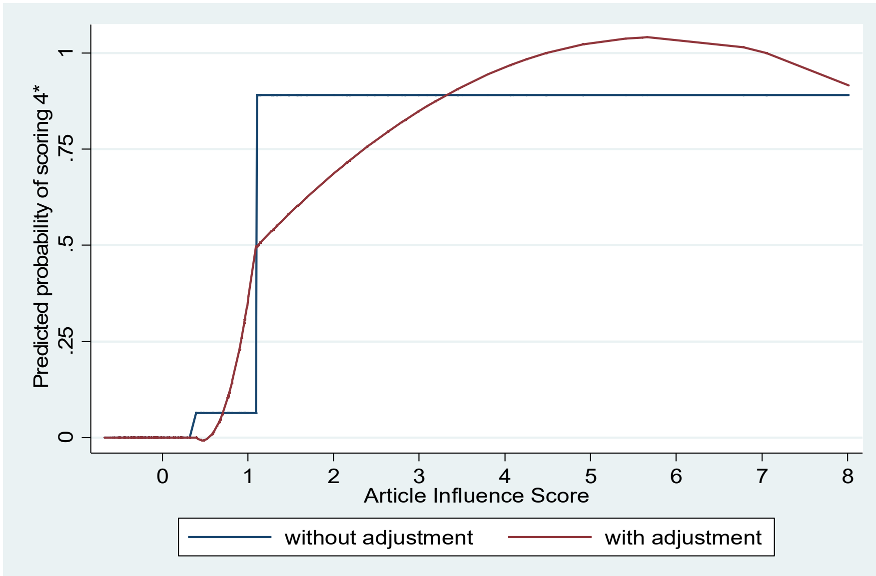
Figure G.2: Restrictions on the REF classification of outputs



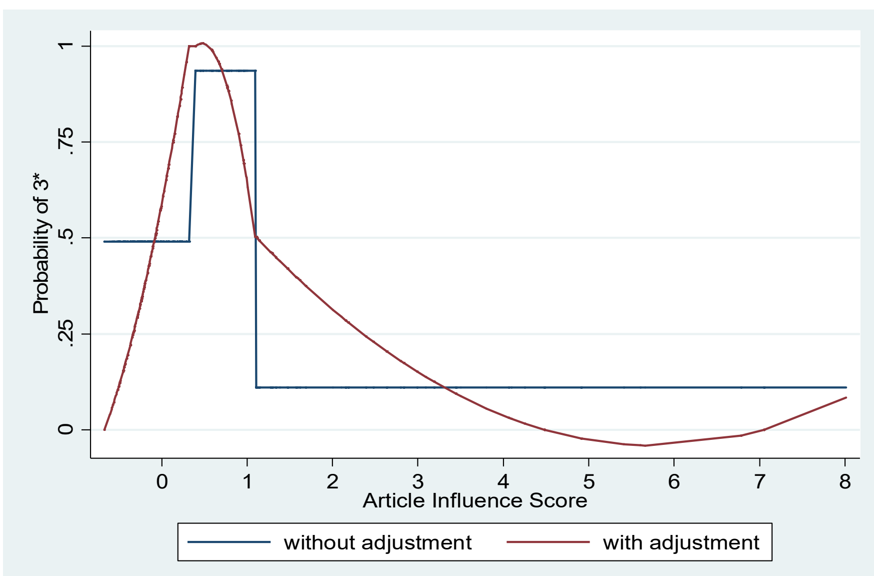
Note. The figure replicates Panel A of Figure 2 when assuming that all publications in the "top-five" Economics journals (the American Economic Review, Econometrica, the Journal of Political Economy, the Review of Economic Studies and the Quarterly Journal of Economics) are awarded four stars. It shows the likelihood of being awarded fewer than four stars for publications in journals with standardized Article Influence Score (AIS) values at or above a certain value. For example, the data do not reject the hypothesis that publications in journals with AIS values at least equal to that of the Economic Journal (EJ) are awarded four stars. The support of the standardized AIS distribution is truncated at five because of the low number of journals above this value. See Section 4 for details.

Figure G.3: Predicted classification probabilities as a function of AIS

Panel A. Likelihood of scoring four stars



Panel B. Likelihood of scoring three stars



Note. The figure plots predicted probabilities of scoring four stars (Panel A) and three stars (Panel B) as a function of AIS. Estimates are derived from results in Panel A of Table 3. The blue lines show predictions from columns (1)-(3), where only tier intercepts are included in the model. The red lines are constructed after adjusting for AIS, using the esults in columns (4)-(6). The adjustment is obtained by fitting a quadratic polynomial in AIS within each tier, and imposing continuity of classification probabilities at tier boundaries. See Section 6 for details.

Working Paper del Dipartimento di Economia e Finanza

1. L. Colombo, H. Dawid, *Strategic Location Choice under Dynamic Oligopolistic Competition and Spillovers*, novembre 2013.
2. M. Bordignon, M. Gamalerio, G. Turati, *Decentralization, Vertical Fiscal Imbalance, and Political Selection*, novembre 2013.
3. M. Guerini, *Is the Friedman Rule Stabilizing? Some Unpleasant Results in a Heterogeneous Expectations Framework*, novembre 2013.
4. E. Brenna, C. Di Novi, *Is caring for elderly parents detrimental to women's mental health? The influence of the European North-South gradient*, novembre 2013.
5. F. Sobbrio, *Citizen-Editors' Endogenous Information Acquisition and News Accuracy*, novembre 2013.
6. P. Bingley, L. Cappellari, *Correlation of Brothers Earnings and Intergenerational Transmission*, novembre 2013.
7. T. Assenza, W. A. Brock, C. H. Hommes, *Animal Spirits, Heterogeneous Expectations and the Emergence of Booms and Busts*, dicembre 2013.
8. D. Parisi, *Is There Room for 'Fear' as a Human Passion in the Work by Adam Smith?*, gennaio 2014.
9. E. Brenna, F. Spandonaro, *Does federalism induce patients' mobility across regions? Evidence from the Italian experience*, febbraio 2014.
10. A. Monticini, F. Ravazzolo, *Forecasting the intraday market price of money*, febbraio 2014.
11. Tiziana Assenza, Jakob Grazzini, Cars Hommes, Domenico Massaro, *PQ Strategies in Monopolistic Competition: Some Insights from the Lab*, marzo 2014.
12. R. Davidson, A. Monticini, *Heteroskedasticity-and-Autocorrelation-Consistent Bootstrapping*, marzo 2014.
13. C. Lucifora, S. Moriconi, *Policy Myopia and Labour Market Institutions*, giugno 2014.
14. N. Pecora, A. Spelta, *Shareholding Network in the Euro Area Banking Market*, giugno 2014.
15. G. Mazzolini, *The economic consequences of accidents at work*, giugno 2014.
16. M. Ambrosanio, P. Balduzzi, M. Bordignon, *Economic crisis and fiscal federalism in Italy*, settembre 2014.
17. P. Bingley, L. Cappellari, K. Tatsiramos, *Family, Community and Long-Term Earnings Inequality*, ottobre 2014.
18. S. Frazzoni, M. L. Mancusi, Z. Rotondi, M. Sobrero, A. Vezzulli, *Innovation and export in SMEs: the role of relationship banking*, novembre 2014.
19. H. Gnutzmann, *Price Discrimination in Asymmetric Industries: Implications for Competition and Welfare*, novembre 2014.
20. A. Baglioni, A. Boitani, M. Bordignon, *Labor mobility and fiscal policy in a currency union*, novembre 2014.
21. C. Nielsen, *Rational Overconfidence and Social Security*, dicembre 2014.
22. M. Kurz, M. Motolese, G. Piccillo, H. Wu, *Monetary Policy with Diverse Private Expectations*, febbraio 2015.
23. S. Piccolo, P. Tedeschi, G. Ursino, *How Limiting Deceptive Practices Harms Consumers*, maggio 2015.
24. A.K.S. Chand, S. Currarini, G. Ursino, *Cheap Talk with Correlated Signals*, maggio 2015.
25. S. Piccolo, P. Tedeschi, G. Ursino, *Deceptive Advertising with Rational Buyers*, giugno 2015.

26. S. Piccolo, E. Tarantino, G. Ursino, *The Value of Transparency in Multidivisional Firms*, giugno 2015.
27. G. Ursino, *Supply Chain Control: a Theory of Vertical Integration*, giugno 2015.
28. I. Aldasoro, D. Delli Gatti, E. Faia, *Bank Networks: Contagion, Systemic Risk and Prudential Policy*, luglio 2015.
29. S. Moriconi, G. Peri, *Country-Specific Preferences and Employment Rates in Europe*, settembre 2015.
30. R. Crinò, L. Ogliari, *Financial Frictions, Product Quality, and International Trade*, settembre 2015.
31. J. Grazzini, A. Spelta, *An empirical analysis of the global input-output network and its evolution*, ottobre 2015.
32. L. Cappellari, A. Di Paolo, *Bilingual Schooling and Earnings: Evidence from a Language-in-Education Reform*, novembre 2015.
33. A. Litina, S. Moriconi, S. Zanaj, *The Cultural Transmission of Environmental Preferences: Evidence from International Migration*, novembre 2015.
34. S. Moriconi, P. M. Picard, S. Zanaj, *Commodity Taxation and Regulatory Competition*, novembre 2015.
35. M. Bordignon, V. Grembi, S. Piazza, *Who do you blame in local finance? An analysis of municipal financing in Italy*, dicembre 2015.
36. A. Spelta, *A unified view of systemic risk: detecting SIFIs and forecasting the financial cycle via EWSs*, gennaio 2016.
37. N. Pecora, A. Spelta, *Discovering SIFIs in interbank communities*, febbraio 2016.
38. M. Botta, L. Colombo, *Macroeconomic and Institutional Determinants of Capital Structure Decisions*, aprile 2016.
39. A. Gamba, G. Immordino, S. Piccolo, *Organized Crime and the Bright Side of Subversion of Law*, maggio 2016.
40. L. Corno, N. Hildebrandt, A. Voena, *Weather Shocks, Age of Marriage and the Direction of Marriage Payments*, maggio 2016.
41. A. Spelta, *Stock prices prediction via tensor decomposition and links forecast*, maggio 2016.
42. T. Assenza, D. Delli Gatti, J. Grazzini, G. Ricchiuti, *Heterogeneous Firms and International Trade: The role of productivity and financial fragility*, giugno 2016.
43. S. Moriconi, *Taxation, industry integration and production efficiency*, giugno 2016.
44. L. Fiorito, C. Orsi, *Survival Value and a Robust, Practical, Joyless Individualism: Thomas Nixon Carver, Social Justice, and Eugenics*, luglio 2016.
45. E. Cottini, P. Ghinetti, *Employment insecurity and employees' health in Denmark*, settembre 2016.
46. G. Cecere, N. Corrocher, M. L. Mancusi, *Financial constraints and public funding for eco-innovation: Empirical evidence on European SMEs*, settembre 2016.
47. E. Brenna, L. Gitto, *Financing elderly care in Italy and Europe. Is there a common vision?*, settembre 2016.
48. D. G. C. Britto, *Unemployment Insurance and the Duration of Employment: Theory and Evidence from a Regression Kink Design*, settembre 2016.
49. E. Caroli, C.Lucifora, D. Vigani, *Is there a Retirement-Health Care utilization puzzle? Evidence from SHARE data in Europe*, ottobre 2016.
50. G. Femminis, *From simple growth to numerical simulations: A primer in dynamic programming*, ottobre 2016.
51. C. Lucifora, M. Tonello, *Monitoring and sanctioning cheating at school: What works? Evidence from a national evaluation program*, ottobre 2016.

52. A. Baglioni, M. Esposito, *Modigliani-Miller Doesn't Hold in a "Bailinable" World: A New Capital Structure to Reduce the Banks' Funding Cost*, novembre 2016.
53. L. Cappellari, P. Castelnovo, D. Checchi, M. Leonardi, *Skilled or educated? Educational reforms, human capital and earnings,* novembre 2016.
54. D. Britto, S. Fiorin, *Corruption and Legislature Size: Evidence from Brazil*, dicembre 2016.
55. F. Andreoli, E. Peluso, *So close yet so unequal: Reconsidering spatial inequality in U.S. cities*, febbraio 2017.
56. E. Cottini, P. Ghinetti, *Is it the way you live or the job you have? Health effects of lifestyles and working conditions,* marzo 2017.
57. A. Albanese, L. Cappellari, M. Leonardi, *The Effects of Youth Labor Market Reforms: Evidence from Italian Apprenticeships*; maggio 2017.
58. S. Perdichizzi, *Estimating Fiscal multipliers in the Eurozone. A Nonlinear Panel Data Approach*, maggio 2017.
59. S. Perdichizzi, *The impact of ECBs conventional and unconventional monetary policies on European banking indexes returns*, maggio 2017.
60. E. Brenna, *Healthcare tax credits: financial help to taxpayers or support to higher income and better educated patients? Evidence from Italy*, giugno 2017.
61. G. Gokmen, T. Nannicini, M. G. Onorato, C. Papageorgiou, *Policies in Hard Times: Assessing the Impact of Financial Crises on Structural Reforms,* settembre 2017.
62. M. Tettamanzi, *E Many Pluribus Unum: A Behavioural Macro-Economic Agent Based Model*, novembre 2017.
63. A. Boitani, C. Punzo, *Banks' leverage behaviour in a two-agent New Keynesian model*, gennaio 2018.
64. M. Bertoni, G. Brunello, L. Cappellari, *Parents, Siblings and Schoolmates. The Effects of Family-School Interactions on Educational Achievement and Long-term Labor Market Outcomes,* gennaio 2018.
65. G. P. Barbetta, G. Sorrenti, G. Turati, *Multigrading and Child Achievement*, gennaio 2018.
66. S. Gagliarducci, M. G. Onorato, F. Sobbrio, G. Tabellini, *War of the Waves: Radio and Resistance During World War II,* febbraio 2018.
67. P. Bingley, L. Cappellari, *Workers, Firms and Life-Cycle Wage Dynamics*, marzo 2018.
68. A. Boitani, S. Perdichizzi, *Public Expenditure Multipliers in recessions. Evidence from the Eurozone*, marzo 2018.
69. M. Le Moglie, G. Turati, *Electoral Cycle Bias in the Media Coverage of Corruption News*, aprile 2018.
70. R. Davidson, A. Monticini, *Improvements in Bootstrap Inference*, aprile 2018.
71. R. Crinò, G. Immordino, S. Piccolo, *Fighting Mobile Crime*, giugno 2018.
72. R. Caminal, L. Cappellari, A. Di Paolo, *Linguistic skills and the intergenerational transmission of language*, agosto 2018.
73. E. Brenna, L. Gitto, *Adult education, the use of Information and Communication Technologies and the impact on quality of life: a case study*, settembre 2018.
74. M. Bordignon, Y. Deng, J. Huang, J. Yang, *Plunging into the Sea: Ideological Change, Institutional Environments and Private Entrepreneurship in China,* settembre 2018.
75. M. Bordignon, D. Xiang, L. Zhan, *Predicting the Effects of a Sugar Sweetened Beverage Tax in a Household Production Model,* settembre 2018.
76. C. Punzo, L. Rossi, *The Redistributive Effects of a Money-Financed Fiscal Stimulus*, gennaio 2019.
77. A. Baglioni, L. Colombo, P. Rossi, *Debt restructuring with multiple bank relationships,* gennaio 2019.

78. E. Cottini, P. Ghinetti, S. Moriconi, *Higher Education Supply, Neighbourhood effects and Economic Welfare*, febbraio 2019.
79. S. Della Lena, F. Panebianco, *Cultural Transmission with Incomplete Information: Parental Perceived Efficacy and Group Misrepresentation,* marzo 2019.
80. T. Colussi, Ingo E. Isphording, Nico Pestel, *Minority Salience and Political Extremism*, marzo 2019.
81. G. P. Barbetta, P. Canino, S. Cima, *Let's tweet again? The impact of social networks on literature achievement in high school students: Evidence from a randomized controlled trial*, maggio 2019.
82. Y. Brilli, C. Lucifora, A. Russo, M. Tonello, *Vaccination take-up and health: evidence from a flu vaccination program for the elderly*, giugno 2019.
83. C. Di Novi, M. Piacenza, S. Robone, G. Turati, *Does fiscal decentralization affect regional disparities in health? Quasi-experimental evidence from Italy*, luglio 2019.
84. L. Abrardi, L. Colombo, P. Tedeschi, *The Gains of Ignoring Risk: Insurance with Better Informed Principals*, luglio 2019.
85. A. Garnero, C. Lucifora, *Turning a Blind Eye? Compliance to Minimum Wages and Employment*, gennaio 2020.
86. M. Bordignon, M. Gamalerio, E. Slerca, G. Turati, *Stop invasion! The electoral tipping point in anti-immigrant voting*, marzo 2020.
87. D. Vigani, C. Lucifora, *Losing control? Unions' Representativeness, "Pirate" Collective Agreements and Wages*, marzo 2020.
88. S. L. Comi, E. Cottini, C. Lucifora, *The effect of retirement on social relationships: new evidence from SHARE*, maggio 2020.
89. A. Boitani, S. Perdichizzi, C. Punzo, *Nonlinearities and expenditure multipliers in the Eurozone*, giugno 2020.
90. R. A. Ramos, F. Bassi, D. Lang, *Bet against the trend and cash in profits*, ottobre 2020.
91. F. Bassi, *Chronic Excess Capacity and Unemployment Hysteresis in EU Countries. A Structural Approach*, ottobre 2020.
92. M. Bordignon, T. Colussi, *Dancing with the Populist. New Parties, Electoral Rules and Italian Municipal Elections*, ottobre 2020.
93. E. Cottini, C. Lucifora, G. Turati, D. Vigani, *Children Use of Emergency Care: Differences Between Natives and Migrants in Italy*, ottobre 2020.
94. B. Fanfani, *Tastes for Discrimination in Monopsonistic Labour Markets*, ottobre 2020.
95. B. Fanfani, *The Employment Effects of Collective Bargaining*, ottobre 2020.
96. O. Giuntella, J. Lonsky, F. Mazzonna, L. Stella, *Immigration Policy and Immigrants' Sleep. Evidence from DACA*, dicembre 2020.
97. E. Cottini, P. Ghinetti, E. Iossa, P. Sacco, *Stress and Incentives at Work*, gennaio 2021.
98. L. Pieroni, M. R. Roig, L. Salmasi, *Italy: immigration and the evolution of populism*, gennaio 2021.
99. L. Corno, E. La Ferrara, A. Voena, *Female Genital Cutting and the Slave Trade*, febbraio 2021.
100. O. Giuntella, L. Rotunno, L. Stella, *Trade Shocks, Fertility, and Marital Behavior* , marzo 2021.
101. P. Bingley, L. Cappellari, K. Tatsiramos, *Parental Assortative Mating and the Intergenerational Transmission of Human Capital*, aprile 2021.
102. F. Devicienti, B. Fanfani, *Firms' Margins of Adjustment to Wage Growth. The Case of Italian Collective Bargaining*; aprile 2021.
103. C. Lucifora, A. Russo, D. Vigani, *Does prescribing appropriateness reduce health expenditures? Main e_ects and unintended outcomes*, maggio 2021.

104. T. Colussi, *The Political Effects of Threats to the Nation: Evidence from the Cuban Missile Crisis*, giugno 2021.
105. M. Bordignon, N. Gatti, M. G. Onorato, *Getting closer or falling apart? Euro countries after the Euro crisis*, giugno 2021.
106. E. Battistin, M. Ovidi, *Rising Stars*, giugno 2021.