NBER WORKING PAPER SERIES

WHAT DO EDITORS MAXIMIZE? EVIDENCE FROM FOUR LEADING ECONOMICS JOURNALS

David Card
Stefano DellaVigna

Working Paper 23282
http://www.nber.org/papers/w23282

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
March 2017

We thank Pierre Azoulay, Matthew Gentzkow, Nagore Iriberri, Lawrence Katz, Jesse Shapiro, and Fabrizio Zilibotti for their comments and suggestions. We are also grateful to the audiences at Columbia University, UC Berkeley, UC Santa Cruz, and the Annual Meeting of WEAI for useful comments. We thank Alden Cheng, Johannes Hermle, Kaushik Krishnan, Patricia Sun, and Brian Wheaton for oustanding research assistance. We are also very appreciative of the support provided by the editors and staff at the Journal of the European Economic Association, the Quarterly Journal of Economics, the Review of Economics and Statistics, and the Review of Economic Studies. The survey analyzed in this paper was approved by UC Berkeley IRB, protocol 2016-08-9029 and pre-registered as trial AEARCTR-0001669. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

What do Editors Maximize? Evidence from Four Leading Economics Journals
David Card and Stefano DellaVigna
NBER Working Paper No. 23282
March 2017
JEL No. A11,O3

ABSTRACT

We study editorial decision-making using anonymized submission data for four leading economics journals: the Journal of the European Economics Association, the Quarterly Journal of Economics, the Review of Economic Studies, and the Review of Economics and Statistics. We match papers to the publication records of authors at the time of submission and to subsequent Google Scholar citations. To guide our analysis we develop a benchmark model in which editors maximize the expected quality of accepted papers and citations are unbiased measures of quality. We then generalize the model to allow different quality thresholds for different papers, and systematic gaps between citations and quality. Empirically, we find that referee recommendations are strong predictors of citations, and that editors follow the recommendations quite closely. Holding constant the referees' evaluations, however, papers by highly-published authors get more citations, suggesting that referees impose a higher bar for these authors, or that prolific authors are over-cited. Editors only partially offset the referees' opinions, effectively discounting the citations of more prolific authors in their revise and resubmit decisions by up to 80%. To disentangle the two explanations for this discounting, we conduct a survey of specialists, asking them for their preferred relative citation counts for matched pairs of papers. The responses show no indication that prolific authors are over-cited and thus suggest that referees and editors seek to support less prolific authors.

David Card
Department of Economics
549 Evans Hall, #3880
University of California, Berkeley
Berkeley, CA 94720-3880
and NBER
card@econ.berkeley.edu

Stefano DellaVigna
University of California, Berkeley
Department of Economics
549 Evans Hall #3880
Berkeley, CA 94720-3880
and NBER
sdellavi@econ.berkeley.edu

1 Introduction

The editorial decisions at top academic journals determine the career paths of young researchers and set the direction of research in a field. Yet, remarkably little is known about how these decisions are made. How informative are the referee recommendations that underlie the peer review process? How do editors combine the referees' advice with their own reading of a paper and other prior information in deciding whether to accept it or reject it? Do editors set the same standards for established scholars as for less prolific authors?

We address these questions using anonymized data on 30,000 recent submissions at the *Quarterly Journal of Economics*, the *Review of Economic Studies*, the *Journal of the European Economics Association*, and the *Review of Economics and Statistics*. Our data include information on the field(s) of each paper, the recent publication records of the authors, summary recommendations for all papers that were refereed, and the initial editorial decision (desk reject, reject, or revise and resubmit). All submissions, regardless of the editor's decision, are matched to citations from Google Scholar and the Social Science Citation Index.

This unique dataset allows us to significantly advance our understanding of the editorial decision process at scientific journals. Most previous research has focused on published papers or aggregated submission data (e.g., Laband and Piette, 1994; Ellison, 2002a and 2002b; Hofmeister and Krapf, 2011; Card and DellaVigna, 2013; Brogaard, Engelberg and Parsons, 2014). While such studies offer many potential insights, they cannot directly shed light on the trade-offs faced by editors since they lack comprehensive information on accepted and rejected papers, including the referees' opinions. A few studies have analyzed submissions data but have focused on other issues such as the strength of agreement between referees (Welch, 2014), the effect of referee incentives (Hamermesh, 1994; Chetty, Saez, and Sandor, 2014) or the impact of blind refereeing (Blank, 1991). Two recent studies (Cherkashin et al., 2009 and Griffith, Kocherlakota, and Nevo, 2009) present broader analyses for the *Journal of International Economics* and the *Review of Economic Studies* respectively, though neither uses information on referee recommendations.

To guide our analysis we propose a simple model of the revise and resubmit (R&R) decision in which editors combine observable paper characteristics, the referee recommendations, and their own private information to select which papers to invite for revision. As a starting point we assume that editors try to maximize the expected quality of published papers and that quality is revealed by citations – i.e., citation-maximizing behavior. While highly restrictive, this stylized model is a useful benchmark for at least three reasons. First, *impact factors,* which count citations to the articles in a journal, are widely used to rank journals and are highly salient to publishers, editors and authors.[1] Second, existing studies show that citations are important determinants of career advancement (Ellison, 2012) and salaries (Hamermesh, Johnson and Weisbrod, 1982; Hilmer, Ransom and Hilmer, 2015). Finally, Google Scholar (GS) citations are available for all papers – whether accepted or rejected – including those that remain unpublished.

Nevertheless, there are at least three limitations of this simple benchmark. First, editors in

[1]There is some concern in the scientific community – e.g., Seglen (1997), Lariviere et al. (2016) – that impact factors have become too influential in making comparisons across journals.

certain journals may favor certain fields (e.g., more theoretical versus more applied fields) or certain groups of authors.[2] We show how such preferences can be easily incorporated in the model, leading to systematic differences between the way that referee recommendations and paper characteristics affect the accept/reject decision versus expected citations. Second, citations can also be directly impacted by the publication process. Although this will not necessarily have a *differential* impact on papers from different fields or different authors, we provide evidence on this issue by comparing citations for papers in the period before they are actually published, and by comparing citations for different types of papers conditional on receiving a revise and resubmit verdict. Third, citations may be systematically biased as a measure of quality by differences in citing practices across fields or a tendency to cite well-established authors (Merton, 1968). Again, this possibility can be easily incorporated in the model. Importantly, however, using only information on citations and editorial decisions we cannot distinguish between editorial preferences for certain types of papers and differential biases in the gap between citations and quality. Thus, in the final section of the paper we augment our sample of submissions with data from a survey of expert readers in which we quantify the relative bias in citations versus quality for specific types of papers.

We focus our main empirical analysis on the R&R decision for the roughly 55% of submissions that are not initially desk rejected. These papers are typically reviewed by 2 to 4 referees who provide summary recommendations on a 7-point scale, ranging from "Definitely Reject" to "Accept". We show that referee recommendations are strong predictors of citations: on average a paper unanimously classified as "Revise and Resubmit" by the referees receives many times more citations than one they unanimously agree is "Definitely Reject". We also show that the fractions of referees who rank a paper in each category provide a good summary of the information contained in the reports, with little loss relative to more flexible alternatives.

Nevertheless, the referee recommendations are *not* sufficient statistics for expected citations. In particular, submissions from authors with more (recent) publications in a set of 35 major journals receive substantially more citations, controlling for referee recommendations. For example, papers by authors with 8 recent publications have on average 2-3 times more citations than papers with similar referee rankings by authors with no recent publications. This suggests either that referees are tougher on more prolific authors (i.e., a bias arising from referee preferences) or that submissions from more prolific authors receive more citations conditional on their quality (i.e., a bias in citations as a measure of quality).

Looking at the R&R decision we find that editors are heavily influenced by the referees' recommendations. Moreover, the relative weights that editor's place on the fractions of referees in different categories are nearly proportional to their coefficients in a regression model for citations, as would be expected if editors are trying to maximize expected citations. We also find that editors have some private information over and above the summary referee opinions that are predictive of citations. Comparing papers that receive an R&R decision and those that do not, we estimate that

[2]Laband and Piette (1994), Medoff (2003) and Brogaard, Engelberg and Parsons (2014) all find that submissions to economics articles by authors who are professionally connected to the editor are more likely to be accepted, though they also find that papers by connected authors receive more citations, suggesting that the higher acceptance rate may be due to information rather than favoritism. Li (2017) similarly finds that members of NIH review committees tend to favor proposals in their own field, but are better informed about these proposals. In contrast, Fisman et al. (2016) find strong evidence of favoritism in elections to the Chinese Academies of Engineering and Science.

the editors' private signals have a correlation of about 0.2 with ultimate citations.

Among the papers that receive a positive R&R decision we can also compare citations for those that were unanimously recommended by the referees and those where the referees were split. Citations for the "close calls" are lower than for the unanimously recommended papers, further stressing the informativeness of the referee recommendations.

Moving beyond the referee recommendations, the R&R decision is also affected by other observable paper characteristics including field and author publication record, with a preference for papers from more prolific authors, conditional on the referees' evaluations. Since the referees *under-value* papers from these authors relative to expected citations, however, editors still tend to accept fewer papers from more prolific authors than would be predicted from a citation-maximizing perspective. In fact, the editors at all four journals appear to undo only a small fraction of the bias against more prolific authors exhibited by referees. This suggests either that editors agree with referees in their preference for papers from less prolific authors, or they agree with referees that the papers by more prolific authors get too many citations, conditional on their quality.

Interestingly, this pattern of underweighting of paper characteristics relative to the citation-maximizing benchmark is not unique to measures of the authors' publications. The editors put essentially no weight on the number of authors of a paper, despite the positive effect of a larger author team on future citations. They also do not consistently put more weight on fields with higher citations. Moreover, these patterns are not due to one or two journals: rather, they are shared by all four journals in our sample. They are also evident in comparisons of citations for papers that receive a positive R&R verdict. We show that under the citation maximization hypothesis, R&R'd papers with the same probability of receiving a positive R&R verdict should have the same average citations. Contrary to this prediction, however, papers by more prolific authors get many more citations than those by less prolific authors but the same propensity for an R&R.

Although our main focus is on the R&R decision, we also analyze the desk rejection decision. Desk rejections have become increasingly common, accounting for over 50% of submissions in our sample, and yet there is little evidence on how these decisions are made. We find that editors have substantial private information about paper quality at the desk rejection stage. Conditional on observable characteristics including field and author publication record, papers sent for refereeing accumulate many more citations than the papers that are desk rejected. Even papers that end up rejected after refereeing have 60 log points more citations on average than papers that are desk rejected. Since both sets of papers are ultimately rejected, this comparison is not biased by mechanical publication effects. Instead, it confirms that editors have substantial information at the desk rejection stage. As at the R&R stage, however, we find that editors appear to discount the expected citations that will accrue to papers by more prolific authors at the desk rejection stage. Indeed, desk-rejected papers by prolific authors have higher average citations than non-desk-rejected papers by authors with no previous publications.

The finding that referees *and* editors act as if they discount the citations of papers by more prolific authors is potentially surprising, since evidence from other fields suggests that more prominent authors receive preferential treatment in the publication process (e.g., Okike et al., 2016), and social scientists have long suspected that the work of more prominent scholars is viewed more positively

– the so-called "Matthew Effect" coined by Merton (1968).[3] In the case of economics, however, Blank's (1991) analysis of a randomized comparison of blind versus non-blind refereeing at the *American Economic Review* showed that blind refereeing led to *higher* relative acceptance rates for submissions from authors at top-5 schools – consistent with a bias against more prolific authors.[4] In addition, Smart and Waldfogel (1996) find *higher* citations to published articles by authors from top departments, controlling for the order of publication in the journal and page length, which they interpreted as measures of editorial treatment.[5] Similarly, Hofmeister and Krapf (2011) find higher citations to articles from authors at top-10 institutions, conditional on the editor's decision on which B.E. journal the paper is published in.

To disentangle whether the discounting of citations for the papers of more prolific authors arises because reviewers and editors think these papers get too many citations or because they are exhibiting affirmative action for less prolific scholars, we conducted a survey of faculty and PhD students in economics, asking them to compare matched pairs of papers (published in the same year in one of the top five journals) in their field of expertise. One paper in each pair was written by an author (or set of authors) with relatively many publications in the years prior to an approximate submission date, while the other paper had authors with few recent publications. We provide respondents with the actual Google Scholar citations for each paper and ask them to assess the appropriate relative number of citations based on their judgment of the quality of the papers. We then use the responses to infer the degree of discounting of citations for papers by prominent authors, using a pre-registered specification.

Our survey respondents do not think that papers by more prolific authors get too many citations. Indeed, their preferred relative citations for more prolific authors are only 2% below their actual relative citations (standard error = 5%), suggesting that the relative citations for papers by more and less prolific authors are proportional to their relative quality.

Putting together the pieces we conclude that the editorial decision process at top economics journals is close to one that maximizes the expected quality of accepted papers, as revealed by ultimate citations, with the important exception that reviewers and editors impose a higher bar for submissions from more prolific authors. This higher bar has a clear cost for the journals – a typical published paper from an author with no previous publications receives many fewer citations than one from authors with many recent publications. Nevertheless, the similarity of behavior across the editors and reviewers at the four journals, and the consistency with earlier findings, including Blank's (1991) analysis of blind versus non-blind refereeing and Smart and Waldfogel's (1996) comparisons of citations for published articles, suggest that the norm of a higher bar for more prolific authors is deeply ingrained in the field.

[3] Lee et al. (2013) present a detailed review of notions of bias in the peer review process, but conclude that evidence of bias in favor of more prominent or successful authors is actually rather limited.

[4] Blank (1991, Table 10) uses information from referees on whether they knew the names of authors even when reviewing the paper under blind conditions, and constructs IV estimates of the effect of truly blind evaluation on the probability of acceptance for different groups of authors. Her results show that the acceptance rate of papers from authors at top 5 schools rises by 20 percentage points when the reviewers do not know the author's name, though the effect is imprecisely estimated.

[5] Medoff (2006) finds that papers by authors from Harvard and University of Chicago tend to receive additional citations conditional on page length and lead article status, but that authors in other top departments do not.

2 Model

In this section we present a stylized model of the editorial decision process. Many journals, including the four in our sample, use a multi-stage process in which the editor first decides whether to desk reject the paper or not, then sends non-desk-rejected papers to referees. Based on the referee reports and his or her own reading of the paper the editor then decides whether to invite the paper for revision and resubmission. Thereafter, the authors and the editor (with or without feedback from the referees) iterate to a final decision. We first present a model of the R&R stage, which we view as the key step in the peer review process. We then discuss the earlier desk reject stage, which shares many of the same features as the R&R stage, albeit with no input from the referees. For simplicity we ignore the stages after a positive R&R verdict.

2.1 The revise and resubmit decision

The key attribute of a paper is its quality q, which is only partially observed by editors and referees. At the R&R stage the editor observes a set of characteristics of the paper and the author(s), x_1, as well as a set of referee recommendations x_R.[6] Quality is determined by a simple additive model:

$$\log q = \beta_0 + \beta_1 x_1 + \beta_R x_R + \phi \tag{1}$$

where for simplicity we treat the unobserved component of quality, ϕ, as normally distributed with mean 0 and standard deviation σ_ϕ. Notice that we allow observable paper characteristics to help forecast paper quality conditional on the referee assessments. In principle, if the referee recommendations efficiently incorporate all the information contained in any observable paper features, as well as any private information garnered by the referees from reading the paper, we would expect the coefficients on x_1 to be zero (i.e., $\beta_1 = 0$). More generally, however, the referees' recommendations may be noisy or biased predictors of quality, in which case β_1 may be different from 0.[7]

The editor observes a signal s which is the sum of ϕ and a normally distributed noise term ζ with standard deviation σ_ζ:

$$s = \phi + \zeta.$$

Conditional on s and $x \equiv (x_1, x_R)$ the editor's forecast of ϕ is:

$$E[\phi|s, x] = As \equiv v$$

where $A = \sigma_\phi^2/(\sigma_\phi^2 + \sigma_\zeta^2)$ is the signal-to-total-variance ratio of the editor's signal. This forecast is an optimally shrunk version of the editor's private signal, and is normally distributed with standard

[6]For simplicity in this paper we do not model the editor's decision over how many referees to assign to a paper, or the slippage between the number of referees assigned and the number who return reports. Bayar and Chemmaur (2013) discuss the optimal composition of the referee pool assigned to a given paper focusing on the role of specialist and generalist reviewers. We present some analysis below of the differences in the opinions of more and less prolific referees on the work of more or less prolific authors, which was investigated in the seminal study by Zuckerman and Merton (1971) of refereeing at the *Physical Review,* and is related to referee "matching" (Hamermesh, 1994).

[7]It is also plausible that the information contained in the recommendations varies across referees, or with the characteristics of the paper, in which case the coefficients β_R could vary with referee characteristics or with x_1. We have investigated the variation in the reliability of different referees, and found that this is relatively small.

deviation $\sigma_v = A^{1/2}\sigma_\phi$ and correlation $\rho = A^{1/2}$ with ϕ. The editor's expectation of the paper's quality is therefore:

$$E[\log q|s,x] = \beta_0 + \beta_1 x_1 + \beta_R x_R + v. \tag{2}$$

With this forecast in hand, the editor then decides whether to give an R&R verdict or not. Here, a natural benchmark is that the editor selects papers for which expected quality is above a threshold. Assuming v has a constant variance, he or she should therefore give a positive decision ($RR = 1$) for papers with $E[\log q|s,x] \geq \tau_0$, where τ_0 is a fixed threshold that depends on the target acceptance rate of the journal.[8] More generally, however, the editor may impose a threshold that varies with the characteristics of the paper or the authors. To allow this possibility we assume:

$$R\&R = 1 \Longleftrightarrow \beta_0 + \beta_1 x_1 + \beta_R x_R + v \geq \tau_0 + \tau_1 x_1 \tag{3}$$

where the benchmark case of $\tau_1 = 0$ corresponds to the situation where the editor cares only about expected quality. Notice that this model is similar to a random preference model (McFadden, 1973) in that the revise and resubmit decision is deterministic as far as the editor is concerned. From the point of view of outside observers, however, randomness arises because of the realization of s. Under our normality assumptions, the R&R decision conditional on x is described by a probit model:

$$\begin{aligned} P\left[RR = 1|x\right] &= \Phi\left[\frac{\beta_0 - \tau_0 + (\beta_1 - \tau_1)x_1 + \beta_R x_R}{\sigma_v}\right] \\ &= \Phi\left[\pi_0 + \pi_1 x_1 + \pi_R x_R\right], \\ &= \Phi\left[\pi x\right] \end{aligned} \tag{4}$$

where $\pi_0 = (\beta_0 - \tau_0)/\sigma_v$, $\pi_1 = (\beta_1 - \tau_1)/\sigma_v$, and $\pi_R = \beta_R/\sigma_v$.

Next, we specify how citations (which are ultimately observable) relate to quality (which is not). We assume that cumulative citations c at some future time are related to quality:

$$q = \delta c$$

where δ is a discount factor that varies by the elapsed time since submission, and can also potentially vary by field and other characteristics of the paper.[9] The simplest possible benchmark is that $\delta = \delta_0$: in this case citations form a perfect index of quality (apart from a scale factor). A more general model allows the discount rate to vary with paper characteristics:

$$\delta = \exp(\eta_0 - \eta_1 x_1) \tag{5}$$

For example, Merton (1968) argued that more prominent scholars tend to accumulate extra recogni-

[8] Assuming that the editors receive a large number of submissions and face a constraint on the total number of papers published per year, they will maximize the average quality of accepted papers by accepting a paper if and only if its expected quality exceeds some threshold. If $\log q$ is normally distributed with mean M and variance V then expected quality is $\exp(M + V/2)$, which will exceed a given threshold T if and only if $M \geq \tau_0 = \log T - V/2$.

[9] As we discuss in Section 5.2, this can be easily generalized to $q = \delta c^\theta$, which allows a convex or concave mapping between quality and citations. Allowing $\theta \neq 1$ has no substantive effect on the implications of the model so for simplicity we set $\theta = 1$.

tion for their work, simply by virtue of their prominence. In this case, citations for papers submitted by more prominent authors have to be discounted relative to those submitted by less prominent authors (i.e., the coefficient in η_1 corresponding to a prominent author will be positive). Combining this discounting with the model for quality given by equation (2) leads to a model for citations:

$$\begin{aligned}\log c &= \beta_0 - \eta_0 + (\beta_1 + \eta_1)x_1 + \beta_R x_R + \phi \\ &= \lambda_0 + \lambda_1 x_1 + \lambda_R x_R + \phi \\ &= \lambda x + \phi, \end{aligned} \tag{6}$$

where $\lambda_0 = \beta_0 - \eta_0$, $\lambda_1 = \beta_1 + \eta_1$, and $\lambda_R = \beta_R$. Clearly, when δ is constant across papers (and thus $\eta_0 = \eta_1 = 0$) we can recover β_1 and β_R from a regression of citations on paper characteristics and referee recommendations. More generally, however, the coefficient λ_1 in equation (6) will reflect both quality and any excess citation effect, so we cannot necessarily interpret differences in citations for papers with different observed characteristics as measures of relative quality.

In our empirical analysis we observe a set of characteristics of a given paper, the referee recommendations, and both the R&R decision and accumulated citations. Using these data we estimate equations (4) and (6) and identify the relative effects of x_1 and x_R on the probability of an R&R verdict and on citations. As a benchmark, consider a citation maximizing choice model with the two simplifying assumptions:

- (A1) the editor only cares about quality ($\tau_1 = 0$)
- (A2) citations are unbiased measures of quality ($\eta_1 = 0$)

Notice that we are not assuming that the referee recommendations are unbiased predictors of a paper's ultimate citations. If, for example, referees tend to give worse recommendations to certain types of authors, this will lead to a positive coefficient for the corresponding element of x_1 in equation (6). Assuming the editor is a citation maximizer, he or she will take this into account in the decision rule and weight papers with this characteristic more positively, controlling for the referees. Specifically, under assumptions A1 and A2 the editor will use weights in the R&R decision rule that are strictly proportional to the weights that the referee reports and the paper characteristics receive in the citation model, leading to the prediction:

(P1) $\pi_1 = \lambda_1/\sigma_v$, $\pi_R = \lambda_R/\sigma_v$.

Figure 1a illustrates the testable implications of this prediction. If we graph the estimated coefficients $(\widehat{\pi}_1, \widehat{\pi}_R)$ from the R&R probit against the corresponding estimated coefficients $(\widehat{\lambda}_1, \widehat{\lambda}_R)$ from an OLS model for log citations, the points should lie on a positively sloped line that passes through the origin with slope $1/\sigma_v$. As we will show, these restrictions are not fully satisfied at any of the four journals in our sample, leading us to consider the sources of the violations.

Dropping either A1 or A2 allows for systematic departures between the way x_1 and x_R affect the probability of an R&R verdict and the way they affect observed citations. Consider first the effect of dropping A1 but maintaining A2, so editors maximize something other than quality but citations are an unbiased measure of quality. In this case the combined set of referee recommendation variables will still affect citations and the R&R decision proportionally, so when we plot the estimated coefficients $\widehat{\pi}_R$ against $\widehat{\lambda}_R$ they will continue to lie on a positively sloped line with slope $1/\sigma_v$. Now, however, the coefficients of the x_1 variables may lie above or below this line, depending on whether τ_1 is positive or negative. For a paper characteristic that has a smaller positive effect on the R&R decision than would be expected given its effect on citations (i.e., one for which the coefficient pair is *above* the fitted line for the referee variables) we can infer that the editor is using a higher threshold in evaluating papers with this characteristic. Alternatively, if we maintain A1 but drop A2 (i.e., the editor maximizes quality but citations are a biased measure of quality), departures from the fitted line associated with the referee variables are caused by paper characteristics that raise or lower citations, holding constant quality. If some characteristic of a paper has a smaller positive effect on the R&R decision than would be expected given its effect on citations, we infer that citations are upward biased relative to quality for papers with this characteristic.

In the most general case when editors have different thresholds for different papers (i.e., $\tau_1 \neq 0$) and citations are potentially biased measures of quality (i.e., $\eta_1 \neq 0$), we cannot tell whether departures from the strict proportionality predictions P1 are attributable to editorial preferences or a bias in the relation between quality and citations, or both. These alternatives can be distinguished if we measure quality, which is the approach we adopt in our survey of expert readers described below.

Quantifying the editor's private signal An important feature of this model is that we can quantify the informativeness of the editor's private signal s, regardless of whether assumptions A1 and A2 are true or not. Intuitively, if the editor has a strong signal of quality then papers that receive an R&R verdict should receive more citations than papers that are rejected, controlling for x_1 and x_R. Using standard formulas it is easy to show that

$$E\left[\log c|x, RR\right] = \lambda_0 + \lambda_1 x_1 + \lambda_R x_R + \rho\sigma_\phi \times r \tag{7}$$

where r is the generalized residual from the R&R probit model :

$$\begin{aligned} r &= \frac{\left(RR - \Phi\left[\pi x\right]\right)\phi\left[\pi x\right]}{\Phi\left[\pi x\right]\left(1 - \Phi\left[\pi x\right]\right)} \\ &= \begin{cases} \frac{\phi[\pi x]}{\Phi[\pi x]} & \text{if} \quad RR = 1 \\ -\frac{\phi[\pi x]}{1-\Phi[\pi x]} & \text{if} \quad RR = 0 \end{cases} \end{aligned}$$

and $\phi(\cdot)$ and $\Phi(\cdot)$ are the standard normal density and distribution functions, respectively (e.g., Gourieroux et al., 1987). In the parlance of the program evaluation literature, equation (7) is a two-sided selection model with no exclusion restrictions. Since the correlation of the error terms in the R&R probit and citation model is $\rho = A^{1/2}$, when the editor has a relatively informative signal the coefficient on the residual from the probit model will be larger and the gap in expected citations between papers that are invited for revision and those that are rejected will be larger. From our sample of reviewed submissions we can estimate equation (7) and obtain an estimate of $\rho\sigma_\phi$: given an estimate of σ_ϕ from the residual variance of equation (6) we can then estimate ρ.[10]

[10] The estimates of λ_1 and λ_R obtained from estimating (7) are numerically identical to the estimates from (6), since the first order conditions for the probit coefficients imply that the generalized residuals are orthogonal to (x_1, x_R) –

Figure 1b illustrates the implications of equation (7). We plot on the x-axis the probability of an R&R decision, which depends on the index πx, and on the y-axis expected citations. Without conditioning on R&R status the expected relationship between observable quality and citations is upward sloping (assuming λx is positively correlated with πx, which is highly plausible). For papers with $RR = 1$ the predicted relationship lies above the unconditional line but is flatter, with a larger gap between the conditional and unconditional expected citations on the far left where a positive R&R verdict signals that the editor's signal is large and positive and therefore ϕ is likely to be positive. For rejected papers the relationship is also flattened, but among these papers the gap is largest on the far right, where a rejection decision signals that the editors signal is large and negative and therefore ϕ is likely to be negative.

Comparisons of Papers with the Same Probability of R&R We have shown that when the editor only cares about quality (i.e., $\tau_1 = 0$ in equation 3) and citations are an unbiased measure of quality (i.e., $\eta_1 = 0$ in equation 5) the coefficients of the R&R probit (other than the coefficient of the constant) are strictly proportional to the corresponding coefficients of the citation model, i.e., $\pi_1 = \lambda_1/\sigma_v$ and $\pi_R = \lambda_R/\sigma_v$. This citation-maximizing case has the important implication that among all papers receiving a favorable R&R decision, expected citations depend only on the conditional probability of an R&R decision:

$$p(x) \equiv \Phi\left[\pi x\right].$$

To see why, consider the expected citations for papers with RR=1 based on equation (7):

$$E\left[\log c|x, RR = 1\right] = \lambda_0 + \lambda_1 x_1 + \lambda_R x_R + \rho\sigma_\phi g(\pi x)$$

where $g(\pi x) = \frac{\phi[\pi x]}{\Phi[\pi x]}$ is the selectivity bias among R&R'd papers.[11] Now consider any two R&R'd papers with the same value for $p(x)$. These papers must have the same value for the covariate index $\pi_1 x_1 + \pi_R x_R$, and thus the same value for $\lambda_1 x_1 + \lambda_R x_R = \sigma_v(\pi_1 x_1 + \pi_R x_R)$. Thus they have the same expected citations. The assumption of citation maximizing behavior by editors therefore implies:

$$E\left[\log c|x, RR = 1\right] = G(p(x)) \tag{8}$$

where $G(z) = \lambda_0 + \sigma_v \Phi^{-1}\left[z\right] + \rho\sigma_\phi g(\Phi^{-1}\left[z\right])$ is a strictly increasing continuous function.

Equation (8) leads to a simple, intuitive test for the citation maximizing hypothesis: we fit a model for the probability of R&R, then classify papers into cells based on their propensity to receive an R&R verdict, and compare average citations for papers with different values of an individual covariate, such as a measure of the author's previous publication record. Under citation maximization, $p(x)$ is a sufficient statistic for expected citations among all papers with R&R=1, and there should be no difference in expected citations for papers in a given cell. If, on the other hand, editors are using a different threshold for different authors (i.e., $\tau_1 \neq 0$) or editors care about quality but

see Gourieroux et al. (1987).

[11] If the latent errors in the model are not normally distributed the functional form of the $g()$ function will be different but it will still only depend on the covariate index πx.

citations are a biased measure of quality (i.e., $\eta_1 \neq 0$) then we expect to see differences in expected citations for papers with the same R&R propensity.

2.2 The Desk Reject Decision

Having analyzed the R&R decision, we now consider the earlier desk rejection decision. At this stage the only observable information is x_1. We assume that conditional on x_1 paper quality is

$$\log q = \psi_0 + \psi_1 x_1 + \omega \tag{9}$$

where ω is a normally distributed error component with mean 0 and standard deviation σ_ω. At the desk reject stage, based on an initial reading of the paper, the editor observes a signal

$$s_0 = \omega + \varepsilon$$

where ε is normally distributed with mean 0 and standard deviation σ_ε. Conditional on this information, the editor's estimate of the expected quality of the paper is

$$\begin{aligned} E[\log q|x_1, s_0] &= \psi_0 + \psi_1 x_1 + A_0 s_0 \\ \text{where } \; A_0 &= \frac{\sigma_\omega^2}{\sigma_\omega^2 + \sigma_\varepsilon^2}. \end{aligned}$$

Define $v_0 = A_0 s_0$: this is a normally distributed random variable with mean 0 and standard deviation $\sigma_{v_0} = A_0^{1/2}\sigma_\omega$ that is observed by the editor but is unknown to outside observers. We assume the editor assigns a paper for review (i.e., does not desk reject the paper) if

$$E[\log q|x_1, s_0] = \psi_0 + \psi_1 x_1 + v_0 \geq \gamma_0 + \gamma_1 x_1$$

which has the same form as the decision rule at the R&R stage.[12] This rule leads to a simple probit model for the probability of non-desk-rejection ($NDR = 1$), conditional on the characteristics x_1:

$$P[NDR = 1|x_1] = \Phi[\frac{\psi_0 - \gamma_0 + (\psi_1 - \gamma_1)x_1}{\sigma_{v_0}}]. \tag{10}$$

Combining equation (9) with our earlier assumptions on the discounting of citations as a measure of quality leads to a model for observed citations conditional on x_1:

$$\log c = \psi_0 - \eta_0 + (\psi_1 + \eta_1)x_1 + \omega \tag{11}$$

[12] An optimal desk reject rule compares the option value of refereeing a paper to the cost of refereeing. Assume as in equation (3) that the R&R decision rule compares the conditional expectation of log quality, given x_1, x_R and a later signal s to some threshold $\tau(x_1)$. Then the optimal rule for not desk rejecting (NDR) is

$$NDR \Leftrightarrow \int\int max[0, E[q|x_1, s_0, x_R, s] - \tau(x_1)] f(x_R, s|x_1, s_0) dx_R d_S - C > 0$$

where $f(x_R, s|x_1, s_0)$ is the joint density of (x_R, s) conditional on the information observed at the desk reject stage, and C is the cost of refereeing. We assume this can be approximated by a cutoff rule of the form $E[\log q|x_1, s_0] > \gamma_0 + \gamma_1 x_1$.

which can be estimated by OLS. Under assumptions A1 and A2 above, $\gamma_1 = \eta_1 = 0$, and the coefficients of the observed characteristics x_1 in the NDR probit model (10) will be strictly proportional to the coefficients in (11). More generally, the NDR decision can reflect a combination of quality and editorial preferences, and citations can reflect a combination of quality and discounting effects.

As at the R&R stage, we can also ask how much information is obtained by the editor in the initial reading that leads to the desk reject decision. Using standard formulas we obtain

$$E\left[\log c | x_1, NDR\right] = \psi_0 - \eta_0 + (\psi_1 + \eta_1)x_1 + \rho_0 \sigma_\omega \times r_0 \tag{12}$$

where r_0 is the generalized residual from the NDR probit model and $\rho_0 = correl[\omega, v_0] = A_0^{1/2}$. Similar to equation (7) at the R&R stage, we can estimate equation (12) and thus ρ_0.

We can also compare the realized mean citations for different types of papers that receive a positive NDR decision for an alternative test of the citation maximizing hypothesis. Using the same arguments as in the derivation of equation (8), it follows that the expected number of citations for NDR papers with a given probability of desk rejection will be the same, regardless of the values of the individual covariates, when the editor seeks to maximize citations. Under the alternative hypothesis that the editor is favoring certain types of papers, however, the average number of citations received by non-desk-rejected papers that do or do not have this characteristic will differ.

3 Data

Data Assembly. We obtained permission from the four journals in our sample to assemble an anonymized data set of submissions that for each paper combines information on the year of submission, approximate field (based on JEL codes at submission), the number of co-authors and their recent publication records, the summary recommendations of each referee (if the paper was reviewed), citation information from Google Scholar and the Social Science Citation Index, and the editor's decisions regarding desk rejection and R&R status.

Our data assembly process relies on the fact that all four journals use the Editorial Express (EE) software system, which stores information about past submissions in a set of standardized files that can be accessed by a user with managing editor permissions. We wrote a program that extracted information from the EE files, queried the GS system, and merged publication histories for each author from a data base of publications in major journals (described below). The program was designed to run on a stand-alone computer under the supervision of an editorial assistant and create an anonymized output file that is stripped of all identifying information, including paper titles, author names, referee names, and exact submission dates. For additional protection the citation counts and publication records of authors are also top-coded.[13] We constructed our data bases for the *Review of Economics and Statistics* (REStat) and the *Quarterly Journal of Economics* (QJE) in April 2015, and the data base for the *Review of Economic Studies* (REStud) in September 2015. The data base for the *Journal of the European Economic Association* (JEEA) was constructed over

[13]The top-code limit is lower for REStud than the other journals. We adjust for this using an imputation procedure based on the mean of citations at the other journals for papers that are above the REStud topcode.

several months up to and including September 2015.

Summary Statistics. We have information on all new submissions (i.e., excluding revisions) to each of the four journals from their date of adoption of the EE system until the end of 2013, allowing at least 18 months for the editor to reach an initial R&R decision. As shown in Table 1 we have data beginning in 2005 for the QJE (N=10,824) and REStud (N=8,335), beginning in 2006 for REStat (N=5,767), and beginning in 2003 for JEEA (N=4,942).

Table 1 and Figure 2a present information on the editorial decisions for the papers in our sample. Desk rejections are more common at the QJE and REStat (60% and 54% of initial submissions respectively) than at REStud or JEEA (20% and 24%, respectively). The R&R rate is lowest at the QJE (4%) and highest at REStat (12%). We do not keep track of the revision stages that occur after an initial R&R decision though we do have information on final publication status for REStud and JEEA. Among papers submitted up to 2011 the final publication rate for papers that received a positive R&R verdict was approximately 90% at JEEA and 75% at REStud.

Figure 2b and Columns 6-10 of Table 1 provide information on a key input to the editorial process: the referee recommendations for papers that are not desk-rejected. The EE system allows referees to enter one of 8 summary recommendations ranging from definitely reject to accept.[14] The modal recommendation is "reject" at all four journals; a majority of recommendations (ranging from 54% at REStat to 73% at QJE) are definitely reject or reject.[15]

We use the JEL codes provided by the author(s) to determine whether the paper belongs to one of 15 field categories listed in Table 1. To account for multiple field codes we set the indicator for a field equal to 1/J where J is the total number of fields to which the paper is assigned. The most common fields are labor, macro, and micro. The field distributions vary somewhat across journal, with a higher share of theory submissions at REStud and a higher share of labor economics at QJE.

An important variable is the publication record of the author(s) at the time of submission. To code this variable, we extracted all articles published in 35 high-quality journals between 1991 and 2014. The set of journals (shown in Appendix Table 1) includes the leading general interest journals as well as top field journals in a majority of fields. We construct the total number of papers published by a given author in these journals in a 5-year window ending in each year from 1995 to 2013, as well as the total number in the top 5 economics journals. We then assign these counts to each co-author and take the highest publication record of all co-authors, setting the count to 0 if we can find no previous publications. For example, a paper written by a team in which the most prolific coauthor published 4 papers in the 35 journals in the 5 years up to and including the year of submission is coded as having 4 papers. We also keep track of the number of coauthors, since this is a positive predictor of citations among published papers (Card and DellaVigna, 2013).

As shown in Table 1 and Figure 2c, 46% of submissions in our overall sample were submitted by authors with no previous publications (or whose names could not be matched to our publication database), while 17% were submitted by authors with 4 or more publications. Submissions at the QJE tend to come from the most prolific authors, followed by REStud, then REStat and JEEA.

[14] The top two categories are conditionally accept and accept. Since these recommendations are rare, we pool both under the accept category.

[15] Welch (2014, Table 3) shows the distributions of referee recommendations at 6 economics journals (including the QJE and 5 others) and 2 finance journals. These distributions are quite similar to the ones in our data.

A final key piece of information is the number of citations received by a paper. We recorded citations as of April 2015 for QJE and REStat and as of August 2015 for REStud and JEEA. For our main measure we use Google Scholar, which provides information regardless of whether a manuscript is published or not. This is particularly important in our context because we are measuring citations for some of the papers in our sample only 2-3 years after the paper was submitted, and we want to minimize any mechanical bias arising because papers that are rejected take some time to be published in other outlets, or may never be published. As a robustness check, we also use counts of citations from the Social Science Citation Index (SSCI), which are reported in Google Scholar but are only available for published papers (and only count citations in other published works).

We merge citation counts to papers using the following procedure. First we extract a paper's title from EE and query GS using the *allintitle* function, which requires all words in the EE title to be contained in the GS title. We capture the top 10 entries found under the allintitle search, and verify that a given GS entry has at least one author surname in common with the names of authors in EE. Then the GS and SSCI citation counts for all entries with a matching name are summed to determine total citations. Papers with no match in Google Scholar are coded as having zero citations. In Online Appendix Table 8, however, we show that our main results are robust to an alternative choice in which papers with no match in GS are dropped from the analysis.

Working with citations raises two issues. First, citation counts are highly skewed: about 30% of submitted papers have no citations, with an even higher rate among recent submissions. Second, citations to a given paper rise with the passage of time. We use two complementary approaches to address these issues. For our main specifications we use the inverse hyperbolic sine (asinh) of the citation count and include journal-year fixed effects. The asinh function closely parallels the natural logarithm function when there are 2+ citations, but is well defined at 0.[16] Appendix Figure 1a shows the distribution of $asinh(citations)$ in our sample, with a spike at 0 (corresponding to 30% of papers with 0 cites) and another mode at around 3 (corresponding to around 10 cites). Under this specification, we can interpret the coefficients of our models as proportional effects relative to submissions from the same journal-year cohort (i.e., as measuring log point effects). As an alternative we assign each paper its citation percentile within the pool of papers submitted to the same journal in the same year. To eliminate heaping we randomly perturb the number of citations received by each paper, smoothing out the 30% of papers with 0 citations (see Appendix Figure 1b).

4 Empirical Results

4.1 Models for Citations and The R&R Decision

Summarizing Referee Opinions

How informative are referee recommendations about future citations? We consider the 15,177 papers that were not desk-rejected and were assigned to at least two referees. This choice reflects the fact that in many cases assignment to a single referee is equivalent to desk rejection. In particular, papers

[16] Asinh(z)=ln($z + \sqrt{1+z^2}$). For z$\geq$ 2, asinh(z)$\approx$ ln(z) + ln(2), but asinh(0)=0.

at REStud assigned to only one referee have a 99% rejection rate. We therefore exclude the 2,271 papers assigned to one referee, though the estimated coefficients in our main models are very similar regardless of whether we include or exclude these papers at all journals or only at REStud.

Figures 3a and 3b show how citations are related to referee recommendations. To construct these figures we take each paper/referee combination as an observation and calculate mean citations by the referee's summary recommendation, weighting observations by the inverse of the number of referee recommendations for the associated paper. Figure 3a uses asinh of the number of citations, while Figure 3b uses the citation percentile within the same journal×year submission cohort.

Both figures show a clear association between referee recommendations and citations, though the effect is somewhat nonlinear, with a relatively large jump between *Definitely Reject* and *Reject,* and a negligible change between *Strong Revise and Resubmit* and *Accept*. The slope of the relationship is quite similar across journals, suggesting a similar informativeness of referees across journals. The *levels* of the citation measures differ, however, with the highest citation levels at the QJE and the lowest at JEEA. The differences in the citation percentile measures are driven by differences in the degree of selectivity of the papers that are reviewed relative to the overall submission pool at each journal. This selection process is strongest at the QJE, where only 40% of papers are reviewed and the average citation citation percentile for all reviewed papers is 65, and weakest at JEEA, where about 65% of papers are reviewed and the average citation percentile of these papers is 53.[17]

Figures 3a and 3b relate mean citations to the opinions of individual referees. How do citations vary with the collective opinions of the entire team of referees? Figure 3c presents a heat map of mean citations for papers with 2 reports, showing the data for each of the 7×7=49 possible cells for the two referee's recommendations.[18] The figure reveals that average citations depend on the average opinions of the referees. For example, papers receiving two *Reject* recommendations have a mean $asinh(citations)$ of 2.5, while papers with two *Strong R&R* recommendations have a mean of 4.1. Papers with one *Reject* and one *Strong R&R* fall in the middle with a mean of 3.2. In Online Appendix Figure 2c we present parallel evidence for papers with 3 reports, creating a heat map using all possible pairs of recommendations. These data support the same conclusions.

In light of this evidence, we summarize the referee recommendations using the fractions of recommendations for a given paper in each of the 7 categories. For example, if a paper with 3 reports has two referees recommending *Reject* and one referee recommending *Weak R&R* then the fractions are 2/3 for *Reject,* 1/3 for *Weak R&R* and 0 for all other categories. This simple procedure has the benefit that can be used irrespective of the number of reports.

Column 1 in Table 2 reports the estimates of an OLS regression model for $asinh(citations)$ that includes journal × year fixed effects and measures of the fractions of referee reports in each category. As in the figures, the estimates suggest a strong positive effect of referee enthusiasm on mean citations. The increases in the estimated coefficients between categories are substantially larger than the slopes in Figure 3b, reflecting the fact that the model yields partial regression coefficients

[17] With 40% of papers reviewed, the expected citation percentile if the desk rejection process perfectly eliminated the bottom tail is 80, while with a 65% review rate the expected citation percentile under perfect selection is 67.5. Using these as benchmarks, the efficiency of the desk rejection is 65/80=0.81 at QJE and 53/67.5=0.79 at JEEA.

[18] The referees' recommendations are modestly positively correlated, with rank order correlations of around 0.25 for 2-referee papers. Welch (2014) shows similar correlations for referee recommendations at a broader sample of economics and finance journals.

whereas the figure is essentially showing univariate coefficients for each category of report.

To document the validity of our averaging specification we return to the subsample of papers with two reports, and display in Figure 3d the predicted citations from the model in column 1 of Table 2 in each of the 49 cells. Comparing these predictions with the actual citations in Figure 3c shows that the model does a very good job of summarizing the two recommendations. The model also does well for papers with 3 reports, as shown by comparing Online Appendix Figures 2c and 2d. Moreover, as shown in Online Appendix Table 4 (columns 1-3) when we compare the coefficients of the referee category variables for papers with 2, 3, and 4 referees, the coefficients are remarkable similar.

Other Determinants of Citations

Next we consider other possible determinants of citations, including the recent publication record of the authors, the number of authors and the field of the paper. Without controlling for referee recommendations, these variables are strong predictors of citations (column 2 of Table 2). An increase in the number of author publications from 0 to 4 or 5, for example, raises citations by about 100 log points, a large (and highly statistically significant) effect. The effect of the number of authors is not as large, though still sizable (and highly significant). Relative to a single-authored paper (the base category) a paper with 3 co-authors has 24 log points more citations (roughly 27% more). There are also systematic differences in citations for different fields (see Online Appendix Table 2): papers in theory and econometrics have the lowest citations, while papers in international and experimental economics have the highest citations. These differences are broadly consistent with patterns in the existing literature based on published papers (e.g., Card and DellaVigna, 2013).

To what extent do these effects persist after controlling for referee recommendations? As noted in Section 2, if the referee reports are sufficient statistics for quality, and citations are unbiased measures of quality, then the other covariates should have no effect on citations after controlling for the referees' recommendations. Within the framework of our model, variables that remain significant predictors of citations indicate that the referees either believe that citations should be discounted for certain groups to properly measure quality, or that certain types of papers should be more highly rated holding constant their quality.

Column 3 in Table 2 presents a full specification with both referee recommendations (x_R in our notation) and the other controls (x_1). The referee variables remain highly significant predictors, with coefficients attenuated by about 15 percent relative to the specification with no controls in column 1. Interestingly, the other controls also remain significant in the joint model. For example, papers by authors with 4-5 recent publications have about 85 log points higher citations than those with 0 recent publications. Interpreted through the lens of our model, this implies that referees evaluate papers as if they were substantially discounting the citations received by more prolific authors. There is a similar effect for papers with more co-authors and papers in more-cited fields.

The Revise and Resubmit Decision

Having examined the predictors of citations, we turn to the predictors of the R&R decision. As discussed in Section 2, under the joint assumptions that editors only care about the expected quality of papers and that citations are an unbiased measure of quality, the coefficients in a probit model for the R&R decision should be proportional to the coefficients in an OLS model for citations that includes the same variables. Under more general assumptions, however, this proportionality prediction will break down.

We first present some graphical evidence. Figure 4a (which is constructed like Figure 3a using paper×referee observations) shows that the probability of an R&R is strongly increasing in the recommendation of any one referee. To examine how editors aggregate multiple recommendations, we show a heat map in Figure 4b of the probability of an R&R verdict for all 49 possible combinations of the referee recommendations when there are 2 referees. This probability is essentially zero with two negative recommendation, rises to 25 percent with two *Weak R&R* recommendations, and to 80 percent or higher with two *R&R* recommendations.[19] Similar patterns are present looking at all possible pairs of recommendations for papers with three referees (Online Appendix Figure 3a).

Columns 4-6 of Table 2 present the estimated coefficients for probit models that parallel the citation models in columns 1-3, using only the referee recommendations (column 4), only the other control variables (column 5), and finally both sets of variables (column 6). As might be expected given the patterns in Figure 4a, the model with only the referee recommendations and journal×submission year controls is remarkably successful, with a pseudo R^2 of 0.48.[20] The quality of fit is apparent in the comparison between Figure 4c which graphs predicted probabilities for each of the possible referee combinations for 2-referee papers, and Figure 4b, which shows the actual probabilities. The relatively close fit of the model across the cells is also true when we look at pairs of reports for papers with 3 referees (see Online Appendix Figures 3a-b).

Column 5 presents a model with only the x_1 (paper characteristic) variables. The R&R rate is increasing with the number of previous publications of the author team, but does not appear to be systematically affected by the number of coauthors, despite the effect of these variables on citations. The same is true of the field variables. Specifically, a comparison of the field effects in the R&R model and the citation model (reported in columns 1 and 3 of Online Appendix Table 2) shows little relation between the relative citations received by papers in a field and the relative likelihood the paper receives an R&R decision.

Column 6 presents the full specification of (4) with both the referee variables and the other covariates. The addition of the latter variables raises the pseudo-R^2 of the probit very slightly (from 0.48 to 0.49), with most of the extra explanatory power coming from the author publication variables, which continue to exert a positive effect on the R&R rate, even controlling for the referee's recommendations. As in column 5, the number of authors has no systematic effect.

Coefficient Plots. With these results in hand, we turn to an examination of the relative

[19] Welch (2014) compares referee recommendations and editorial decisions for an anonymous journal and shows that editorial decisions are highly related to the referees' opinions.

[20] The journal-year fixed effects contribute very little to the fit, with a pseudo-R^2 of 0.03 when they are the only controls.

magnitude of the coefficients of the various paper characteristics in the citation model and the R&R decision model. Our focus is on evaluating predictions P1, which state that if the editor is maximizing citations, the coefficients in these two models will be strictly proportional.

Figures 5a-b plot the coefficients from the R&R probit model (Column 6 of Table 2) against the corresponding coefficients from the citation model (column 3 of Table 2). For visual clarity, Figure 5a displays only the coefficients on the referee recommendation variables and on the author prominence variables, while Figure 5b shows all the coefficients. For interpretive purposes, the figures also show the best-fitting lines through the origin for various subgroups of coefficients. Under the null hypothesis of the model, these lines should all approximately have the same slope.

The referee recommendation coefficients in Figure 5a are remarkably aligned: referee categories that are associated with higher citations are also associated with a higher probability of an R&R decision. For example, the large jump in citations in moving from *Weak Revise and Resubmit* to *Revise and Resubmit* is mirrored by a large rise in the probability of R&R, while the negligible impact of moving from *Strong R&R* to an *Accept* recommendation on citations is also reflected by negligible effect on the probability of R&R. From this pattern one might conclude that the decision-making of editors is closely aligned to the views of the referees, and both are focused on higher citations.

When it comes to the other paper characteristics, however, the parallelism between citations and the R&R decision breaks down. As shown in Figure 5a, for example, measures of author publications exert a much smaller effect than would be expected given their impacts on citations. The red line displays the degree of proportionality between the author publication variables in the R&R model and the citation model. The slope is only about one fifth the slope of the black line which shows the degree of proportionality between the referee recommendation coefficients in the models.

This conclusion is confirmed by a close examination of the coefficients in columns 3 and 6 of Table 2. The coefficients of the referee variables are about twice as big in the R&R model as they are in the citation model, implying in the context of our model that the standard deviation of the latent error in the editor's decision model (σ_v) is about 0.5 (since $\pi_R = \beta_R/\sigma_v$). In contrast, the coefficients of the author publication variables are only about 40% as large in the R&R model as the citation model. The two ratios differ by a factor of about 5, as is visible in Figures 5a and 5b.

Figure 5b also display the coefficients of two other groups of variables: those associated with the number of authors and those associated with field. Both sets of variables have a significant effect on citations, yet editors put essentially no weight on the number of authors, nor do the coefficients on the field indicators appear to line up with their effects on citations (compare the coefficients in columns 2 and 4 of Online Appendix Table 2). Evidently, editors are putting much higher relative weight on the referee recommendations than on other variables that also predict citations.

Do these patterns differ by journal? Figures 6a-d show that several key patterns are common across all 4 journals. (The underlying coefficients are reported in Online Appendix Table 3). First, within each group of variables, the coefficients line up nicely on a line. Second, the line for referee recommendations is systematically steeper than for other variables, implying that editors give more weight to the referee recommendations than to any of the other variables in forming their R&R decisions. Third, at all journals the measures of author publications have a particularly large and systematic impact on citations, but a much smaller relative impact on the R&R decision. This gap is particularly notable at REStud and REStat, where the editors appear to assign *no weight* to any variable other than the referee recommendations. At these journals, the R&R models seem to suggest that the editors simply follow the referees, with no attempt to undo any biases that referees exhibit in evaluating papers from different types of authors or from different fields.

Other Citation Measures. A potential concern with the findings in Table 2 and Figures 5 and 6 is that the results are affected by our use of the inverse hyperbolic sine transformation in modeling citations. To address this concern in Table 3 we reestimate the citation model using alternative transformations. Column 1 shows our base specification, reproduced from column 3 of Table 2. Column 2 uses our percentile citation measure, which controls for differences in citations across journal-year cohorts very flexibly by computing citation percentiles within cohorts. Column 3 uses a transformation motivated by the hypothesis that editors focus exclusively on the probability that a paper becomes a "major hit" and is among the most highly cited papers in their submission pool. Specifically, we define a paper to be *top cited* if it is in the top p percent of citations in a journal-year cohort, where p is set to the R&R rate for that journal and year. We then estimate a probit model to predict the probability of being top cited. We also consider a specification in column 4 using $\log(1 + citations)$ as an alternative to the asinh specification. Finally, in column 6 we re-estimate our citation model using SSCI citations. Since SSCI citations only accrue to published papers, we restrict the sample to submissions in the years from 2006 to 2010 to ensure enough time for publication. To check the robustness of our main specification to the choice of sample, column 5 shows a model for $asinh(GS\ citations)$ fit to the 2006-2010 sample, which looks very similar to the baseline model in column 1.[21]

The results are very consistent across the alternative citation measures, with coefficients that are nearly proportional across specifications. For example, the coefficients in column 3 have a correlation of 0.998 with the coefficients in column 1, implying that the same index of observed paper characteristics predicts both mean asinh of citations and the probability of being in the upper tail of citations. In all cases referee recommendations are strong predictors of the measure of citations, with coefficients that are roughly proportional to the coefficients in the R&R probit, but of different scales depending on the citation measure. All the models also indicate significant positive effects of the author publication variables on the measure of citations, with a relative magnitude about 50% as large as the effects of the referee variables. Since the author publication variables enter the R&R probit model with coefficients only about 10% as large as the referee variables, we conclude that editors under-weight author publications by a factor of about 5 in their R&R decision, regardless of whether editors are maximizing expected asinh GS citations (column 1), the expected percentile of GS citations (column 2), the probability of being in the top tail of GS citations (column 3), or the expected asinh or percentile of SSCI citations (columns 6-7). The only notable difference across specifications is that the number of authors has a small and insignificant effect on SSCI citations, compared to a significant positive effect on GS citations.

Additional Measures of Author Publications. In our baseline specification we measure author productivity by the number of articles published in 35 high-impact journals over the 5-year

[21] As shown in Online Appendix Table 5, when we re-estimate our baseline R&R probit model using data from 2006-2010 the estimates are very similar to those from the whole sample period.

period prior to submission. To probe the robustness of our under-weighting conclusion we checked three additional measures of productivity. The first is the count of publications in the previous 5 years in top-5 economics journals (REStud, QJE, the *American Economic Review*, *Econometrica*, and the *Journal of Political Economy,* excluding the Papers and Proceedings of the AER). The second is the count of publications in our 35-journal sample in the 6 to 10 years prior to submission. The third is an indicator for the prominence of the authors' home institutions, which may proxy for the quality of their past work or their promise as scholars (in the case of young researchers).

Table 4 presents citation models and R&R probit models in which we augment our baseline models from Table 2 (reproduced in columns 1 and 4) with these additional measures. The specifications in columns 2 and 5 include indicators for the number of author publications in top 5 journals. Since top-5 publications are relatively infrequent, we censor our measure at 4 publications in the past 5 years. As is evident from the estimates in column 2, measures of previous top-5 publications are important predictors of citations: a paper from an author team with 2 recent top-5 publications is associated with an extra 43 log points of citations, even conditional on all the other variables. They also strongly affect the R&R decision. Nevertheless, their effect on the R&R decision relative to the effect of the referee recommendation variables is much smaller than in the citation model, suggesting a significant under-weighting of top-5 publications by editors relative to a citation-maximizing benchmark (see Online Appendix Figure 4f).

We also report the estimated effects of publications in the 35 high-impact journals in the period 6-10 years before submission. Although papers from authors with more publications in this earlier time frame do not receive significantly more citations (controlling for their recent publications), earlier publications do have a small positive effect on the R&R decision. Moreover, controlling for earlier publications and recent top-5 publications, the effects of recent publications in the broader 35 journal sample are all small and insignificantly different from 0.

Finally, in columns 3 and 6 we report the impacts of a measure of institutional prominence for the author team at the time of submission, distinguishing between US institutions (coded into 3 groups), European institutions (coded into 2 groups) and institutions in the rest of the world (coded into 2 groups). We use the rankings in Ellison (2013) to classify US institutions, while for non-US institutions we use the 2014 QS World University Rankings for Economics.[22] Since we only collected institutional prominence variables for REStud and JEEA, the models in columns 3 and 6 are fit to the subsample of submissions at these two journals.[23]

The results in column 3 show that institutional prominence is an important predictor of citations, even conditional on a broad set of measures of the authors' publication record. For example, having at least one coauthor at a top-10 US economics department at the time of submission increases citations by 52 log points, while having a coauthor at an 11-20 ranked US institution increases citations by 44 log points. Institutional affiliations also affect the R&R decision (column 6), but as with other characteristics included in x_1 their relative impact on the R&R decision is much smaller

[22]The institutional prominence dummies for each paper are defined within region, so that the dummies for each region sum to at most one, and the sum of the institutional dummies ranges from 0 to 3. Similar to our measure of author publications, we take the top-ranked U.S. institution among coauthors when defining the U.S. institution dummies, and the top-ranked European institution when defining the European dummies.

[23]Estimates of the models in columns 2 and 5 for these two journals are very similar to the ones for the full sample.

than the relative impact of the referee variables (see Online Appendix Figure 4g).

A particularly interesting set of findings concern the effects of institutional affiliation in Europe. Conditional on the referee recommendations, having a co-author at a top-10 department in Europe increases citations by 36 log points, a large and highly significant effect. Yet this affiliation has no significant effect on the R&R decision. Since *REStud* and *JEEA* are based in Europe, and many of the editors are drawn from top-10 European departments, this extreme downweighting cannot be explained by a lack of information about the relative standing of different schools. It appears that these two journals are "leaving citations on the table" by implicitly raising the threshold for an R&R decision when the author is from a top European department.

Value Added of the Editor

Our simple model suggests that the difference in citations between papers that receive an R&R decision and those that are rejected provides a measure of the editor's "value added" or, more formally, of the correlation between the editor's private signal and unobserved paper quality. Figure 7a presents the empirical analogue of the simulation in Figure 1b. For each paper we predict the probability of a revise-and-resubmit decision using the specification in Column 6 of Table 2. We then sort papers into deciles by this predicted probability, splitting the top decile into two top groups, and plot mean citations for papers with a positive and negative decision. We also show the number of papers in each probability range with each decision.

As shown along the $x-axis$ of the Figure, for papers in the bottom 5 deciles of predicted citations the probability of an R&R is near zero, reaching just 1% in the fifth decile. The probability is still only 18% in the 8^{th} decile, but increases sharply to 37% in the 9^{th} decile and equals 90% for papers in the top 5 percent of submissions. The vertical gap between the mean citations for R&R's and rejected papers is relatively large – on the order of 60-80 log points. A simple calibration based on the vertical gap for papers with around a 50% probability of R&R implies that the correlation of the editor's signal with true quality is around 0.22.[24] The vertical gap is also wider to the left, as predicted by our model: the editor has to receive a very positive signal for papers with relatively low observable quality in order to reach a positive R&R decision.[25] Online Appendix Figure 7a displays the same data as in Figure 7a along with the predicted fit from our model, showing that the model does a relatively good job of capturing the patterns in Figure 7a.

Another salient feature of Figure 7a is that even among papers that receive a positive R&R recommendation, expected citations are increasing in the strength of the observable predictors. For example, mean $asinh(citations)$ for R&Rs in the top group in the figure (the top 5% of predicted citations) is about 4.1, while the mean for those in the 7th group (the top 60-70% of predicted citations) is about 3.6 – a gap of 50 log points. Thus, the close calls where the editor appears to have made a positive decision despite only lukewarm enthusiasm from the referees (and no offsetting $x_1's$) yield lower average citations than cases where the referees are very positive (and the editor agrees).

[24] As shown by equation (7), the gap in expected citations between R&R and rejected papers when the probability of R&R equals 0.5 is $\phi(0)/\left(\Phi(0)(1-\Phi(0)\right)\rho\sigma_\phi \approx 1.6\rho\sigma_\phi$.

[25] The visual interpretation of Figure 7a is affected by the scaling of the $x-axis$, which has highly unequal differences in the probability of R&R, unlike Figure 1b, which has evenly spaced probabilities.

This is consistent with the model and stresses the informativeness of referee recommendations.

Publication Bias in Value Added? A concern with our measure of editors' value added is that some of the gap in citations between papers with RR=1 and those with RR=0 is mechanical. Papers that receive an R&R are likely to be published relatively quickly, while those that are rejected have to be submitted and evaluated at other outlets, and may never get published. To the extent that published papers attract more attention and garner more citations than unpublished papers, this will bias upward our estimate of the information content in the editor's private signal. To assess this issue we compare the value added plots for recent submissions (2012-13) versus older submissions (2006-10). Given the lags in publication, submissions from 2012 and 2013 are unlikely to have been published by mid-2015, when citations were measured).[26] Thus, the mechanical bias is unlikely to have emerged among these papers.

Figure 7b shows that the difference between citations—our measure of value added—is indeed smaller for more recent submissions. Even for these submissions, however, the difference in citations between rejected papers and R&Rs is still 30 log points and statistically significant (Column 3 in Online Appendix Table 6), with an implied correlation of 0.13 between the editor's private signal and the citations received by the paper. This suggests that there may indeed be some mechanical bias in the correlation between R&R status and citations, though this explains only about half the average correlation. In any case, this bias does not explain the under-weighting of paper characteristics – including the author publication record – in the editor's R&R decision, relative to the weights of these characteristics in predicted citations.

4.2 Desk Rejections

While our main focus is the R&R decision, in this section we present a brief discussion of the desk rejection decision, building on the framework suggested by our simple model. An empirical analysis of this stage is useful given that more than half of the submissions to many journals are desk rejected, and that the previous empirical literature has largely ignored desk rejections.[27]

Using the full sample of 29,868 submitted papers, we compare predictors of citations with predictors of the decision to not desk reject (NDR) the paper. Author publications and the size of the author team are important predictors of citations (column 1 of Online Appendix Table 7). As would be expected if the NDR process selects papers based in part on the editor's private information about potential citations, the impacts of these variables are *larger* than when we estimate the same specification using only the subset of papers assigned to referees. For example, the coefficients of the publication measures in column 1 of Online Appendix Table 7 are approximately 1.3 times larger than the coefficients in the model in column 2 of Table 2, while the coefficients of the team size variables are about 1.1 times larger. These two sets of variables, plus field dummies and journal×year

[26] In this time period, for example, the *Journal of the European Economic Association* had a publication queue of about 1.5 years.

[27] On the theoretical side, Vranceanu et al. (2011) present a model in which papers with a poor match to the editorial mission of the journal are desk-rejected, but quality per se is irrelevant. Bayar and Chemmanur (2013) present a model in which the editor sees a signal of quality, desk rejects the lowest-signal papers, desk accepts the highest-signal papers, and sends the intermediate cases to referees. Schulte and Felgenhauer (2015) present a model in which an editor can acquire a signal before consulting the referees or not.

fixed effects have a combined R-squared of about 0.23 in predicting GS citations. Thus, there is considerable information in observed paper characteristics that can be used to predict citations.

A probit model for NDR, reported in column 2 of Online Appendix Table 7, show that editors use the prior publication record of authors in making their initial NDR decision, but put little systematic weight on the number of co-authors or the field of the papers. A plot of the coefficients from the NDR probit against those of the citation model therefore shows systematic deviations from null hypothesis of citation maximization (Online Appendix Figure 6), with editors downweighting information in the number of coauthors and field relative to the information in prior publications.

Value Added of the Editor at the Desk Reject Stage

How much information does the editor have at the desk-rejection stage? This is a potentially important question because the desk rejection process is sometimes characterized as arbitrary or uninformed. Figure 8a plots mean citations for four groups of papers in various quantiles of the predicted probability of NDR. We show mean $asinh(citations)$ for papers that are desk rejected (the red line at the bottom) and those that are not desk rejected (the blue line) as well as separate lines for NDR paper that are ultimately rejected at the R&R stage (the green line) and those that receive a positive R&R decision (the orange line at the top of the figure).

The figure reveals large gaps in mean citations between desk-rejected and NDR papers, and between papers that are not desk rejected and then receive a positive or negative R&R.[28] On average, NDR papers receive about 75 log points more citations than those that are desk rejected, implying that the editor obtains substantial information from scrutinizing a paper before making the desk reject decision. In the context of our model this gap implies that the correlation between the editor's initial signal s_0 and future citations is about 0.32, and that s_0 reveals about 10% of the unexplained variance of citations given the observed characteristics at the desk reject stage.[29]

The gap between NDR papers that are ultimately given an R&R and those that are rejected is also large – around 125-175 log points. This gap reflects the discriminatory power of the entire second stage of the review process, including the inputs of the referees and the editor's private signal at the R&R stage. For example, comparing papers that are reviewed by the referees and had an 80% probability of NDR based on x_1, those that are ultimately R&R have mean $asinh(citations)$ of 4.0 while those that are ultimately rejected have a mean of 2.25 - implying about 5.7 times more citations for the R&R group.

Finally, the gap in average citations between desk rejected papers and those that are NDR but ultimately rejected is about 60 log points. This gap is interesting because both sets of papers are rejected - thus, there is no mechanical publication effect biasing the comparison. This gap can be decomposed as the sum of 75 log point gap in citations attributable to the NDR decision, minus a 15 log point gap attributable to the "bad news" of a rejection in the second stage. Viewed this way, the editor's signal at the desk reject stage is relatively informative.

[28] The gap between papers that are R&R'd and those that are rejected after review is larger than the corresponding gap in Figure 7a (for the same set of papers) because of the different ways of grouping papers along the x-axis – by probability of NDR in Figure 8a (based only on x_1) and by probability of R&R in Figure 7a (based on x_1 and x_R).

[29] Recall that according to our model the signal to total variance ratio is $A_1 = \rho_1^2$, where $\rho_1 = 0.31$ is the implied correlation of the editor's signal and the citation residual.

So far, we have seen that author publications are highly predictive of the desk rejection decision. Since we do not have referee recommendations to benchmark the relative effect of the publication record, however, it is not clear whether editors over-weight or under-weight authors' publications in reaching their decision. Building on the test proposed by equation (8), we evaluate the hypothesis that desk rejection decisions are consistent with citation maximization by comparing citations for NDR papers with similar probabilities of desk rejection from more and less prolific authors.

We present this comparison in Figure 8b, focusing on authors (or author teams) with 4 or more recent publications versus those with 0 or 1 publications. Mean citations are about 100 log points higher for papers by more prolific authors, conditioning on NDR status *and* the quantile of the predicted probability of NDR. Indeed in most quantile bins the mean citations of desk rejected papers by more prolific authors have higher mean citations than the non-desk-rejected papers by less prolific authors. This pattern parallels our results at the R&R stage, where editors significantly under-weight the citations of more highly published authors, effectively imposing a higher bar (in terms of expected citations) for these authors. At both stages there appears to be a higher bar for authors with a stronger track record.

5 Interpretation and Survey

To summarize our findings so far: at all four journals in our sample, referees and editors appear to impose a higher bar for papers by more prolific authors (or groups of authors). Figure 9a revisits the evidence for referees. We display mean $asinh(citations)$ for papers by more and less prolific authors with a given referee recommendation (using the same classification of prolific as in Figure 8b). If the referees were evaluating papers based on expected citations the two lines would be similar. Instead, mean citations for prolific authors are 100 log points higher. In other words, referees evaluate papers as if the citations received by more prolific authors should be discounted by roughly e^1.

Columns 1 and 2 of Table 5 quantify this discounting effect for the full set of publication dummies in our models. We begin in column 1 with a model for $asinh(citations)$, fit to the subsample of papers that were assigned to at least two referees, that includes only the author publication variables and journal×year dummies. Relative to the omitted group of authors with no recent publications, papers from authors with 8+ publications have 139 log points higher citations. When we add in the referee recommendations and controls for field and number of coauthors (column 2), this gap falls by about 35 log points, but is still highly significant.

In the R&R decision model in column 6 of Table 2 we saw that editors put positive weight on author's publications (given the referee opinions), effectively "undoing" some of the bias against more prolific authors. How much does the R&R selection process reduce the effects of the publication variables? The answer is shown by the models in columns 3 and 4 of Table 5, which are fit to the subsample of papers that receive a positive R&R decision. Editors undo only about 20 percent of the implicit discounting imposed by referees, reducing the expected citation premium for papers from authors with 8+ previous publications, for example, from 105 to 82 log points. Interestingly, the estimated citation premiums for more prolific authors in the subsample of R&R papers are not very sensitive to whether we include the referee recommendations (as in column 4) or not (in column 3).

The estimated publication coefficients from the model in column 4 are very similar to the coefficients obtained when we implement the test described in Section 2.1, based on comparisons of citations for papers with the same probability of obtaining an R&R verdict. Specifically, we estimated a model for $asinh(citations)$ with dummies for papers in each decile of the predicted probability of an R&R verdict as well as an additional dummy for papers in top vingtile, and indicators for authors' previous publications. The estimated coefficient for 8 or more publications in this specification is 0.87 – quite close to the corresponding coefficient the model in column 4. (The other publication coefficients are also quite close). This confirms that we can clearly reject the hypothesis of citation-maximizing decision-making by editors.

To what extent does the citation advantage for papers from more prolific authors change when we condition on final publication status? While we do not know the publication status for all the R&R'd papers in our sample, we assume that the vast majority were ultimately published. We therefore used EconLit to construct a sample of all papers published in the 4 journals in our sample between 2008 and 2015. Assuming an average 2 year delay between first submission and publication, these papers should correspond to the R&R papers in our sample from 2006 to 2013 (minus the papers that were rejected after an initial positive R&R verdict). We then constructed the x_1 variables for these published papers, coding author publications at an assumed submission date 2 years before the publication date, and using the JEL codes in EconLit (which may differ from the codes at initial submission used in our main analysis). The estimated model for GS citations for these papers, shown in column 5 of Table 6, reveals a set of estimated publication coefficients that are slightly smaller than the ones in column 3 for R&R papers, but still relatively large.[30]

Finally, for completeness we constructed a third sample of papers published in the top 5 economics journals between 1997 and 2010, coding the x_1 variables for these papers by assuming a 2 year lag between submission and publication, and using Google Scholar citations as of late 2016. Since these papers have all been published for at least 6 years, any concerns about publication-status-related biases are eliminated. Moreover, the model includes journal×year effects which control for differences in citations accruing to papers in more or less prestigious journals. The citation model for this sample (reported in column 6) yields estimated author publication effects that are attenuated by about 20-30% relative to the effects in our R&R sample (column 3), but are still highly significant.

5.1 Interpretations

Our model suggests two main interpretations for the key finding in Tables 2-5 that referees and editors significantly under-weight the expected citations that will be received by papers of more prolific authors. The first is that citations are inflated measures of paper quality for prolific authors, leading referees and editors to discount citations accordingly. There are several potential explanations for why the number of citations to a given paper could be higher for more prolific authors, controlling for quality. For one, there are often multiple papers, written around the same time, that contain

[30]We measure Google Scholar citations in late 2016 for these papers using the same search protocols as for our main sample. We find 1,534 published papers in EconLit at the four journals, compared to 2,209 R&R recommendations. We believe the relative size of the published sample is reasonable, given that given that not all of the R&R papers are published and that the EconLit sample probably excludes most papers submitted to JEEA in 2003-05.

similar ideas. More prolific scholars may have broader networks of colleagues, students, etc., who know their work and cite it rather than a similar paper by some less prolific scholar. A closely related idea – Merton's (1968) "Matthew effect" – is that people tend to cite the best known author when there are several possible alternatives. A second possibility is that more prolific authors have access to working paper series and other channels that publicize their work. This is particularly likely to inflate their relative citations in the first few years after papers are written. To the extent that current citations beget future citations, however, any such initial advantage can lead to a permanent gap in citation rates for more and less prolific authors.

An alternative interpretation is that citations may be an appropriate measure of quality, but referees and editors impose a higher bar for more prolific authors. Such a process may be due to a desire to keep the door open to less established scholars (i.e., affirmative action) or a desire to explicitly prevent established authors from publishing marginally acceptable papers (i.e., animus).[31] We note that there are at least two pieces of evidence in the existing literature that support this interpretation. The most direct evidence is Blank's (1991) analysis of blind versus non-blind refereeing at the *American Economic Review,* which showed that blind refereeing increased the relative acceptance rate of papers from authors at top-5 schools. A second finding, closely related to the analysis in Table 9, is that published papers written by authors who were professionally connected to the editor at the time of submission tend to have more rather than less citations (Laband and Piette, 1994; Medoff, 2003; Brogaard, Engelberg and Parsons, 2014).

Before we turn to some survey-based evidence designed to distinguish between these two interpretations, however, we briefly discuss a third possibility that is sometimes raised in the editorial context: elite favoritism.[32] According to this hypothesis, more accomplished authors are *favored* in the publication process by other prolific authors who review their work positively, and by editors who are typically in the same professional networks. If one takes citations as unbiased measures of quality, we clearly find substantial evidence against the elite favoritism hypothesis. It is possible, however, that the citations received by more prolific authors are highly inflated, and that after appropriate discounting (e.g., a discount of >100 log points) we would see that more prolific authors actually face a lower bar in the editorial process.

A plausible test of the elite favoritism hypothesis is to examine whether papers by prolific authors are evaluated more positively by other prolific scholars. Though all the editors in our sample have strong publication records, placing them squarely in the prolific category, the prior publication records of the referees vary widely. It is therefore interesting to test whether the citation gap in Figure 9a differs when the referee has a strong publication record (and is therefore a potential member of the elite) or not.[33] The comparison, shown in Figure 9b, gives no evidence of elite favoritism: the gap in citations between papers of prominent and non-prominent authors is about the same whether the recommendation comes from a prolific referee or a non-prolific referee. Interestingly, a similar conclusion was reached in the seminal study by Zuckerman and Merton (1971), which showed similar

[31]A related possibility is that editors may believe that less prolific authors are more likely to deliver a responsive revision if invited to provide one.

[32]This hypothesis is often raised informally by commentators who are skeptical of the integrity of the peer review process. See Campanario (1998a, 1998b) and Lee et al. (2013) for some context.

[33]Berk, Harvey and Hirshleifer (2017) argue on the basis of interviews with former editors that relatively junior scholars are often harsher in all their reviews.

assessments of papers by more and less prominent authors by more and less prominent referees.

5.2 Survey Evidence on Quality vs. Citations

Using information on citations and the R&R decision we cannot distinguish between the two main explanations for the down-weighting of citations for papers written by more prolific authors. The two alternatives can be distinguished by data that allow us to measure quality independently of citations. In this section we present evidence from a survey designed with this purpose in mind.

The survey aims to replicate the quality assessment of referees and editors, using pairs of published papers. Specifically, the survey respondents compare two published papers that differ by the publication record of the author(s) at the time of submission, but are otherwise matched in terms of journal quality, publication year, and field. This contrast is designed to mirror the R&R decision faced by a journal editor in selecting among the submissions. It also mirrors our empirical specifications which include fixed effects for broad fields of the paper, as well as for each year-journal cohort. The comparison of papers *within* a field also makes the evaluation easier for the survey respondents, and resembles the evaluation of referees who typically assess submissions in their field.

To identify pairs of papers, we consider articles published from one of the traditional top-5 journals in economics between 1999 and 2012, excluding AER Papers and Proceedings articles, notes, and comments. We code articles in the subfields of (i) unemployment; (ii) taxation; (iii) crime; (iv) education; (v) family economics; and (vi) behavioral economics.[34] We also code the articles as (mainly) theoretical or empirical.

Following the same procedure as in our main analysis sample, we measure the publications of authors in the same set of 35 high-impact journals in the 5 years prior to submission. Given the delays between submission and publication, we assume that papers were submitted 2 years prior to the year of publication. We then take, as in our main analysis, the maximum across all coauthors. We classify an author or author team as prolific if there is at least one coauthor with 4 or more publications in the 5 years prior to the assumed submission date. Likewise, we classify the author or team as non-prolific if all co-authors have no more than 1 publication in this period. Notice that some of the authors coded as non-prolific at the assumed submission year may be coded as prolific in later years. This is as intended and reflects the procedure we used in our main analysis and the information available to the referees and editors at the time of submission.

We then identify balanced pairs of papers – one written by a prolific author, one by a non-prolific authors – published in one of the top-5 journals[35] in the same year, in the same field, and with the same theory or empirical component. To simplify our design we exclude papers by authors with intermediate publication records. We exclude pairs that have too large a differential in citations (a ratio of citations outside the interval from 0.2 to 5.0). We also exclude a small number of pairs that included a paper written by one of us, or that we viewed as too far apart in content. The final

[34]The coding of the fields uses a combination of keywords. We search for the keywords in either the title of the paper, or in the description for one of the JEL codes associated with the paper.

[35]In constructing potential pairs we focused on papers from the *American Economics Review*, the *Quarterly Journal of Economics*, and the *Journal of Political Economy*, which tend to publish articles that are similar in the level of mathematical formality. For behavioral economics, given the smaller sample of articles, we include one article from *Econometrica*.

sample includes 60 pairs of papers, with 8 pairs on the topic of unemployment, 12 pairs on taxation, 6 pairs on crime, 12 pairs on education, 10 pairs on family economics, and 12 pairs on behavioral economics. The number of distinct papers is 101, since some papers appear as part of two pairs.

Survey Wording. The survey was administered on the Qualtrics platform, with all the questions displayed on one page (see Online Appendix Figure 8). The respondents were asked two main questions about each pair of papers they are asked to consider. The first asks their "opinion in comparing various features of the two papers," focusing on four specific criteria: (i) Rigor (theoretical structure and/or research design); (ii) Importance of Contribution; (iii) Novelty; (iv) Exposition (organization, clarity, detail, writing). For each criterion the respondent is asked to indicate whether Paper A is better, Paper A is slightly better, the two papers are about the same, Paper B is slightly better, or Paper B is better. We randomize the order in which the four criteria are asked, as well as whether Paper A or Paper B is the paper written by a prolific author.

Second, the survey informs the respondent of the Google Scholar citations as of August 2016 for the two papers and asks: *In light of the ____ citations accrued by Paper A and your assessment above, please indicate whether you think that the number of citations for Paper B is (i) about right, (ii) too high, (iii) too low.* We then elicit a quantitative measure of the appropriateness of citations:

> *In light of the ___ citations accrued by Paper A and your assessment above, what do you think the appropriate number of citations for Paper B should be?*

Let c_A and c_B denote the actual citations of papers A and B, and let $\hat{c}_B$ denote the elicited *appropriate number of citations* for paper B. When paper B is the one written by a prolific author, the ratio $\hat{c}_B/c_B$ represents the respondent's desired discount factor for the citations of the more prolific author. A value for this ratio that is less than 1 means that the respondent thinks the paper is "over-cited" relative to paper A, whereas a value greater than 1 means that he or she believes paper B is "under-cited". In the alternative case when paper A is the one written by a more prolific author, the desired discounting factor for citations to the paper by the more prolific author is $c_B/\hat{c}_B$.

The second half of the survey presents the same questions for a second pair of papers, and ends with an opportunity for the respondents to provide feedback.

Survey Respondents. The survey population includes faculty and PhD students who specialize in the fields covered by the papers in the survey. The survey was administered at in late September and October 2016. Our analysis follows a pre-registered analysis plan, AEARCTR-0001669.

Out of 93 emails sent to 73 faculty and 20 PhD students, 74 surveys were completed, 55 by faculty and 19 by PhD students, for an overall response rate of 80 percent. Each respondent compared 2 pairs of papers in their field, yielding $74 \times 2 = 148$ comparisons covering 58 distinct pairs.

Estimating the Mean Discount for Citations of More Prolific Authors

For paper pair j, let R_j represent the ratio of the number of citations for the paper written by the prolific author to the number of citations for the paper written by the non-prolific author. Using the respondent's answer to the question about the appropriate number of citations to paper B, we construct the respondent's quality-adjusted citation ratio as:

$$\widehat{R}_j = \widehat{c}_B / c_A \text{ if paper B is by the prolific author}$$
$$= c_A / \widehat{c}_B \text{ if paper A is by the prolific author.}$$

We interpret $\widehat{R}_j$ as the respondent's assessment of the ratio of the quality of the paper by the prolific author (q_P) to the the quality of the paper by the non-prolific author (q_N), i.e.,

$$\widehat{R}_j = q_P / q_N.$$

Our model asserts that the relation between quality and citations is s $q = c\delta$. Normalizing $\delta = 1$ for papers by non-prolific authors, it follows that

$$\widehat{R}_j = \delta R_j. \tag{13}$$

Thus, we fit the simple regression model:

$$\log \widehat{R}_j = d_0 + d_1 \log R_j + \varepsilon. \tag{14}$$

With a constant discount factor for citations to papers by more prolific authors, as specified by equation (13), we should estimate $d_0 = \log(\delta)$ and $d_1 = 1$.

A slightly more general model of the relationship between citations and quality is $q = c^{\theta}\delta$, which allows a potentially concave or convex mapping from quality to citations (holding constant the discount factor). It is straightforward to show that all the implications of the model developed in Section 2 remain unchanged when $\theta \neq 1$.[36] In this case, however, equation (13) becomes:

$$\widehat{R}_j = \delta R_j^{\theta}$$

and the predicted value for the coefficient d_1 in equation (14) is $d_1 = \theta$.

Figure 10a illustrates two possible patterns of results using simulated data. We bin papers into deciles by the citation variable ($\log R_j$) and plot the average of the y variable ($\log \widehat{R}_j$) within each bin. The dotted red line illustrates a case with, to a first approximation, no quality discounting: the regression line runs through the origin. The continuous blue line shows simulated data, assuming quality discounting for a value of $\delta = 0.75$, implying an intercept for the regression of $\alpha = -0.28$.

Figure 10b shows a bin-scatter of our actual data. Following our pre-analysis plan we winsorize the dependent variable at the 2^{nd} and 98^{th} percentiles. The average quality-adjusted citation ratios are clearly correlated with the actual citation ratios, with a slope close to 0.7 and an estimated intercept close to 0. Panel A of Table 6 provides a series of estimates of the model specified by equation (14), with a simple OLS regression in Column 1 and a specification in Column 2 in which

[36]The only change is that the coefficients in the citation model, equation (6), take on the values $\lambda_0 = (\beta_0 - \eta_0)/\theta$, $\lambda_1 = (\beta_1 + \eta_1)/\theta$, and $\lambda_R = \beta_R/\theta$ and the residual in the citation model becomes $\theta^{-1}\phi$. Under citation maximizing behavior the coefficients of the R&R probit are still proportional to the coefficients in the citation model, but the factor of proportionality is θ/σ_v.

we weight the responses for a given paper pair by the inverse of the number of respondents who evaluated the pair, thus giving equal weight to pairs evaluated by different numbers of respondents. In Column 3 we limit the sample to pairs with more comparable citations ($-0.5 \leq \log R_j \leq 0.5$). Under either specification, we obtain a precise estimate of no quality discounting: the intercept $\hat{d}_0 = -.02$ (s.e. 0.05) implies $\hat{\delta} = 0.98$.

In columns 4 and 5 we fit separate models for respondents who are either graduate students and younger faculty with relatively few publications (column 4) or faculty who would be classified as prolific (i.e., have published 4 or more papers in the past 5 years in one of the 35 journals in Online Appendix Table 1). The results show that any discounting of the relative citations of more prolific authors comes, if anything, from the prolific faculty. This provides no evidence of elite favoritism and suggests that any discounting may stem in part from competitiveness among prolific authors.

Qualitative Ratings. For paper pair j, the survey respondents also assess the relative strength of the two papers on a five-point scale, which we re-scale from -2 to +2 so positive values correspond to a higher rating for the paper by the prolific author. As shown in Figure 10c, there is at best a weak relationship between the respondents' assessments of the relative strengths of the papers and their relative citations $\log R_j$, with the the strongest relationship for relative importance (plotted with red dots). None of the scatters suggest a negative intercept, as would be expected if citations for more prolific authors are upward biased relative to quality.

Panel B of Table 6 present a series of regression models in which we relate the relative strength of the paper by the prolific author in a given pair to the relative citation measure and a constant. Consistent with Figure 10c, only the model for "Importance" (column 2) has an R-squared coefficient above 0.05. Again, the key coefficient for our purposes is the constant, which we interpret as the discount factor applied to the relative citations of the paper by the prolific author in the particular domain. None of the estimated constants are large or even marginally significant, confirming the main result in Panel A. Overall, our survey results provide no evidence that papers by prolific authors receive more citations than those by non-prolific authors, controlling for their relative quality.

6 Conclusion

Editors' decisions over which papers to publish have major impacts on the direction of research in a field and on the careers of researchers. Yet little is known about how editors combine the information from peer reviews and their own prior information to decide which papers are published.

In this paper, we aim to provide systematic evidence using data on all submissions over an 8-year period for 4 high-impact journals in economics. We analyze recommendations by referees and the decisions by editors, benchmarking them against a simple model in which editors maximize the expected quality of the papers they publish, and citations are an ex-post measure of quality.

We show that this simple model model is consistent with some of the key features of the editorial decision process, including the systematic relationship between referee assessments, future citations, and the probability of an R&R decision, and the fact that R&R papers receive higher citations that those that are rejected, conditional on the referees' recommendations.

Nevertheless, there are important deviations from the citation-maximizing benchmark. On the referee side, certain paper characteristics are strongly correlated with future citations, controlling for the referee assessments of a paper. This suggests that referees impose higher standards on certain types of papers, or that that they are effectively discounting the future citations that will be received by these papers. At best the editors only partially offset these tendencies – indeed at two of the four journals in our sample it appears that the editors essentially follow the referees.

We focus our attention on one of the most important determinants of citations - the publication record of the authors at the time of submission. Referees appear to substantially discount the future citations that will be received by more prolific authors and the editors offset the referees only slightly. Thus, among the papers that receive a revise-and-resubmit decision, those written by more prolific authors receive many times more citations, on average, than those written by less prolific authors, controlling for the referee assessment.

We consider two main interpretations. Citations may be an inflated measure of paper quality for prolific authors, leading referees and editors to discount citations accordingly. Alternatively, citations may be an appropriate measure of quality, but referees and editors may be using affirmative action to support less prolific authors. While our main analysis cannot separate the two interpretations, we present evidence from a survey of economists asked to evaluate the quality of pairs of papers. The survey results are most consistent with the affirmative action interpretation.

We view this just as a step in the direction of understanding the functioning of scientific journals, with many questions remaining. For example, are there similar patterns of citation discounting in other disciplines? Okike, Hug, and Kocher (2016) provide some evidence from a medical journal of favoritism towards prolific authors, a finding different from ours. Also, can a simple model explain the decision of editors to wait for another report or decide with what is at hand? We hope that future research will get at these and other questions.

References

Bayar, Onur and Thomas J. Chemmanur. 2013. "A Model of the Editorial Process in Scientific Journals." Working Paper.

Berk, Jonathan B., Campbell R. Harvey and David Hirshleifer. 2017. "How to Write an Effective Referee Report and Improve the Scientific Review Process." *Journal of Economic Perspectives*, 31(1): 231-244.

Blank, Rebecca M. 1991. "The Effects of Double-Blind versus Single-Blind Reviewing: Experimental Evidence from The American Economic Review." *American Economic Review*, 81(5): 1041-1067.

Brogaard, Jonathan, Joseph Engelberg and Christopher Parsons. 2014. "Networks and Productivity: Causal Evidence from Editor Rotations." *Journal of Financial Economics*, 111(1): 251-270.

Campanario, Juan Miguel. 1998a. "Peer Review for Journals as It Stands Today – Part 1." *Science Communications*, 19(3): 181-211.

Campanario, Juan Miguel. 1998b. "Peer Review for Journals as It Stands Today – Part 2." *Science Communications*, 19(4): 277-306.

Card, David and Stefano DellaVigna. 2013. "Nine Facts about Top Journals in Economics." *Journal of Economic Literature*, 51(1): 144-161.

Cherkashin, Ivan, Svetlana Demidova, Susumu Imai and Kala Krishna. 2009. "The inside scoop: Acceptance and rejection at the journal of international economics." *Journal of International Economics*, 77(1): 120-132.

Chetty, Raj, Emmanuel Saez, and Laszlo Sandor. 2014. "What Policies Motivate Pro-Social Behavior? An Experiment with Referees at the Journal of Public Economics." *Journal of Economic Perspectives*, 28(3), 169-188.

Ellison, Glenn. 2002. "The Slowdown of the Economics Publishing Process." *Journal of Political Economy*, 110(5): 947-993.

Ellison, Glenn. 2002b. "Evolving Standards for Academic Publishing: A q-r Theory." *Journal of Political Economy*, 110(5): 994-1034.

Ellison, Glenn. 2012. "Assessing Computer Scientists Using Citation Data." Working Paper.

Ellison, Glenn. 2013. "How Does the Market Use Citation Data? The Hirsch Index in Economics." *American Economic Journal: Applied Economics*, 5(3): 63-90.

Fisman, Raymond, Jing Shi, Yongxiang Wang, and Rong Xu. Forthcoming. "Social Ties and Favoritism in Chinese Science", *Journal of Political Economy.*

Gourieroux, Christian, Alain Monfort, Eric Renault and Alain Trognon. 1987. "Generalised Residuals." *Journal of Econometrics*, 34(1-2): 5-32.

Griffith, Rachel, Narayana Kocherlakota, and Aviv Nevo. 2009. "Review of the Review: A Comparison of the Review of Economic Studies with its Peers." Unpublished Working Paper.

Hamermesh, Daniel S. 1994. "Facts and Myths about Refereeing." *Journal of Economic Perspectives*, 8(1): 153-163.

Hamermesh, Daniel S., George E. Johnson and Burton A. Weisbrod. 1982. "Scholarship Citations and Salaries: Economic Rewards in Economics." *Southern Economic Journal*, 49(2): 472-481.

Hilmer, Michael J., Michael R. Ransom, and Christiana E. Hilmer. 2015. "Fame and the fortune of academic economists: How the market rewards influential research in economics." *Southern Economic Journal*, 82(2): 430-452.

Hofmeister, Robert and Matthias Krapf. 2011. "How Do Editors Select Papers, and How Good are They at Doing It?" *The B.E. Journal of Economic Analysis & Policy*, 11(1): article 64.

Laband, David N. and Michael J. Piette. 1994. "Favoritism versus Search for Good Papers: Empirical Evidence Regarding the Behavior of Journal Editors." *Journal of Political Economy*, 102(1): 194-203.

Larivière, Vincent, Véronique Kiermer, Catriona J. MacCallum, Marcia McNutt, Mark Patterson, Bernd Pulverer, Sowmya Swaminathan, Stuart Taylor, Stephen Curry. 2016. "A simple proposal for the publications of journal citation distributions." *bioRxiv* preprint.

Lee, Carole J., Cassidy R. Sugimoto, Guo Zhang and Blaise Cronin. 2013. "Bias in Peer Review." *Journal of the American Society for Information Science and Technology*, 64(1): 2-17.

Li, Danielle. Forthcoming. "Expertise vs. Bias in Evaluation: Evidence from the NIH" *American Economic Journal: Applied Economics.*

McFadden, Daniel. 1973. "Conditional Logit Analysis of Qualitative Choice Behavior". In *Frontiers in Econometrics*, edited by Paul Zarembka, 105-142. New York: Academic Press.

Medoff, Marshall H. 2003. "Editorial Favoritism in Economics?" *Southern Economic Journal*, 70(2): 425-434.

Medoff, Marshall H. 2006. "Evidence of a Harvard and Chicago Matthew Effect." *Journal of Economic Methodology*, 13(4): 485-506.

Merton, Robert K. 1968. "The Matthew Effect in Science." *Science*, 159(3810): 56-63.

Okike, Kanu, Kevin T. Hug, Mininder S. Kocher and Seth S. Leopold. 2016. "Single-blind vs Double-blind Peer Review in the Setting of Author Prestige." *JAMA: The Journal of the American Medical Association*, 316(12): 1315-1316.

Schulte, Elisabeth and Mike Felgenhauer. 2015. "Preselection and Expert Advice." Macie Paper Series Working Paper.

Seglen, Per O. 1997. "Why the impact factor of journals should not be used for evaluating research." *BMJ: British Medical Journal*, 314(7079): 498-502.

Vranceanu, Radu, Damien Besancenot and Kim Huynh. 2011. "Desk rejection in an academic publication market model with matching frictions." ESSEC Working Paper.

Welch, Ivo. 2014. "Referee recommendations." *Review of Financial Studies*, 27(9): 2773-2804.

Zuckerman, Harriet and Robert K. Merton. 1971. "Patterns of Evaluation in Science: Institutionalisation, Structure and Functions of the Referee System." *Minerva*, 9(1): 66-100.

Figure 1a. Model Prediction I: Predictors of Citation versus Predictors of Editor Decision

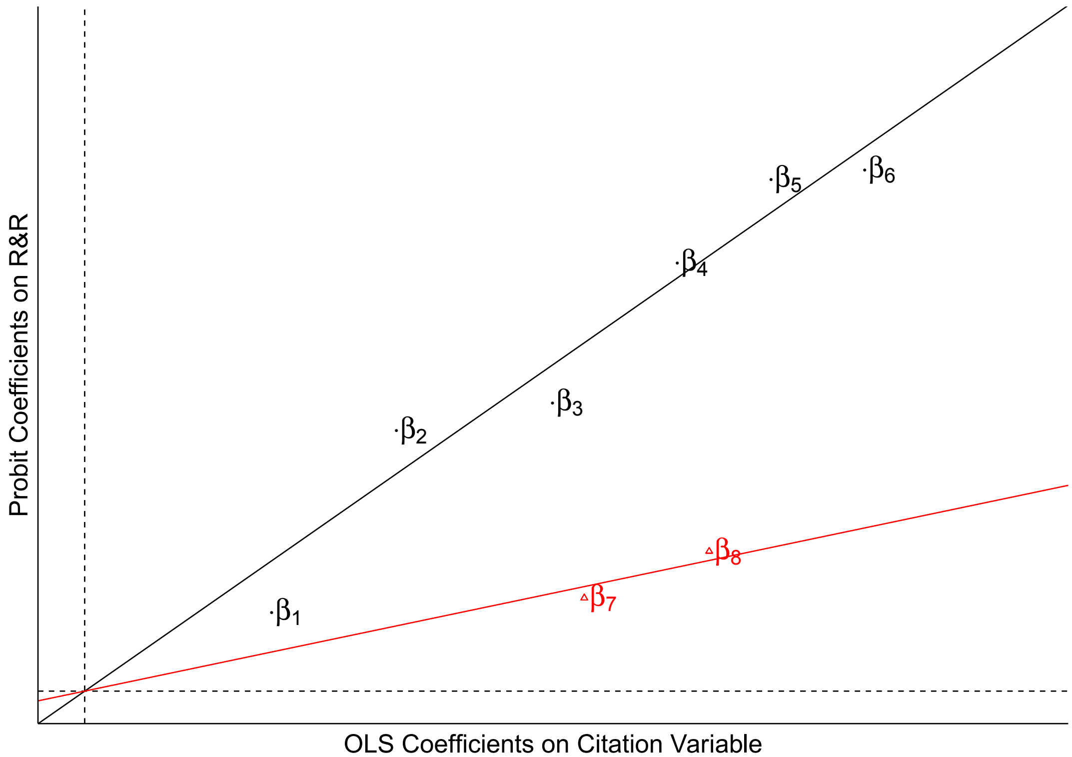
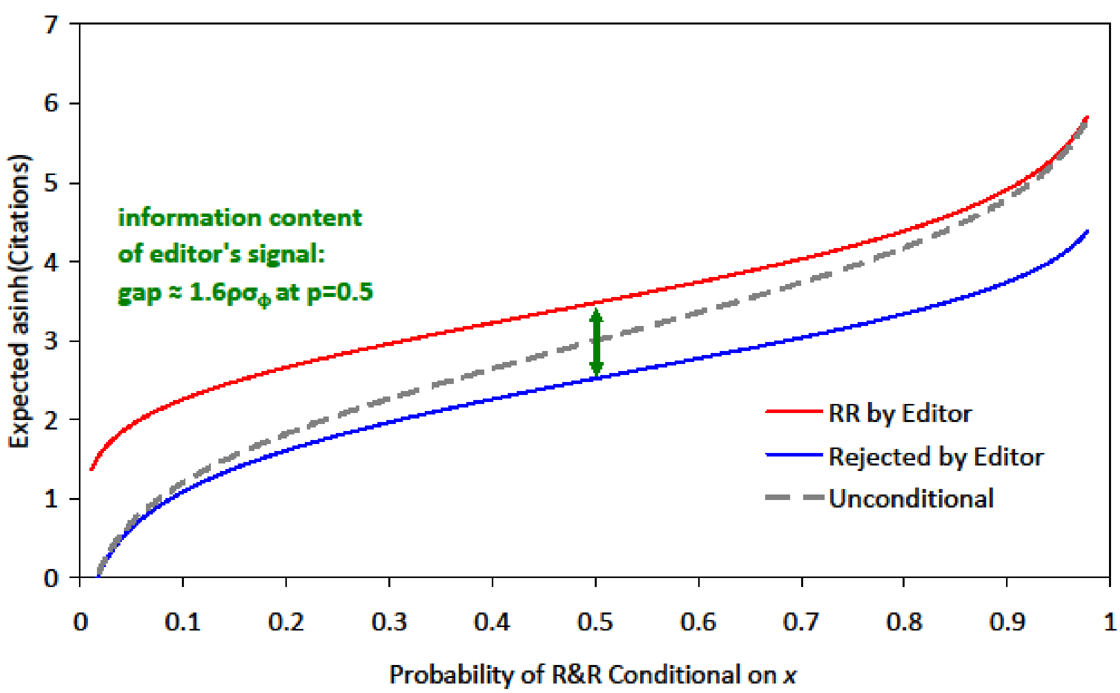


Figure 1b. Model Prediction II: Value Added of the Editor



Notes: Figure 1a plots, for simulated values, the coefficients for a citation regression (x axis) and an R&R probit (y axis). If the coefficients all line up on one line, the evidence is consistent with editors maximizing citations; if the coefficients are on multiple lines, the evidence implies a deviation from this model. Figure 1b reports simulated data from the model for papers that (hypothetically) receive and R&R versus those which do not, for a given value of the probability of R&R. The vertical distance between the red and blue line provides a measure of the editor private information about the paper quality.

Figure 2. Summary Statistics

Figure 2a. Distribution of Editorial Decisions

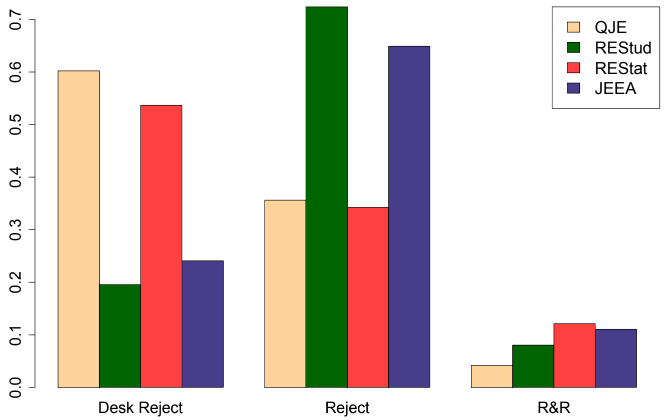


Figure 2b. Distribution of Referee Recommendations

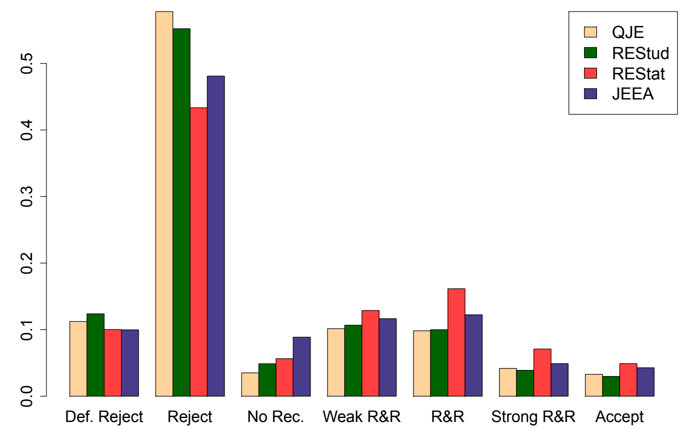
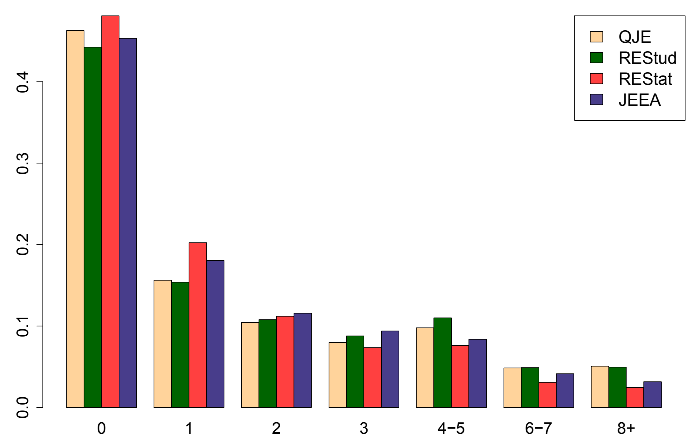


Figure 2c. Distribution of Author Prominence



Notes: Figure 2 displays a few key summary statistics by journal. Figure 2a plots the distribution of the editor's decision and Figure 2b shows the distribution of referee recommendations. Figure 2c plots the distribution of author publications in 35 high-impact journals in the 5 years leading up to submission, for the papers in our dataset. The unit of observation is a paper, and for papers with multiple coauthors, we take the maximum publications among coauthors.

Figure 3. Referee Recommendations and Citations

Figure 3a. Impact on Asinh of Citations

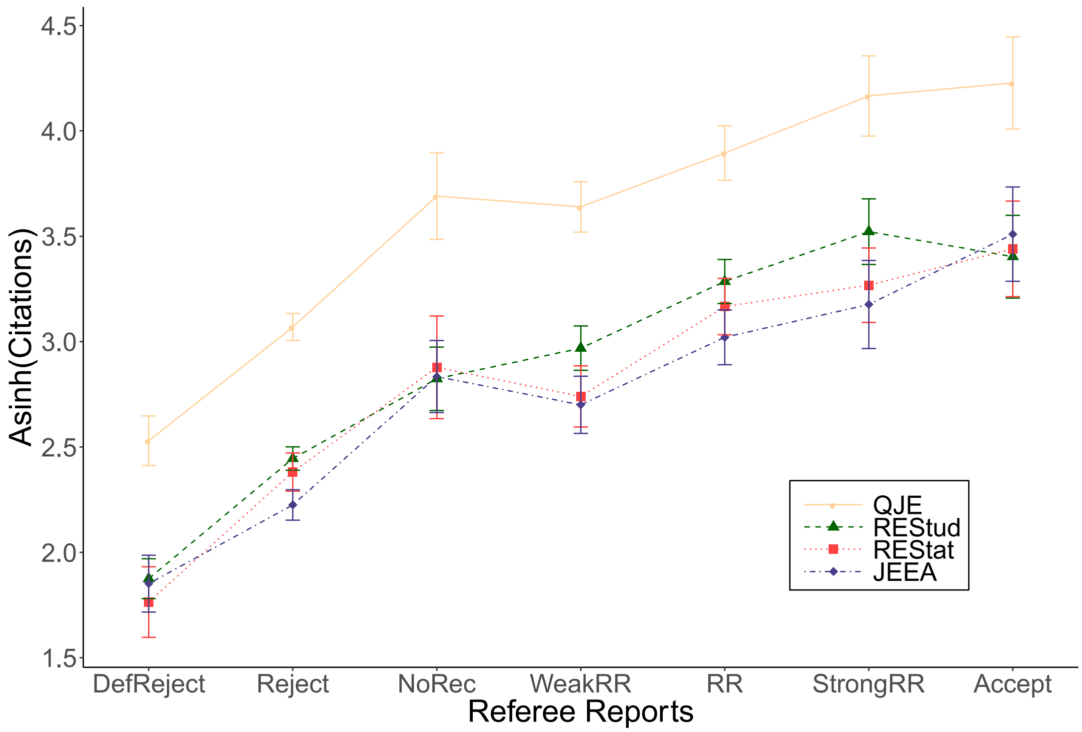


Figure 3b. Impact on Citation Percentile

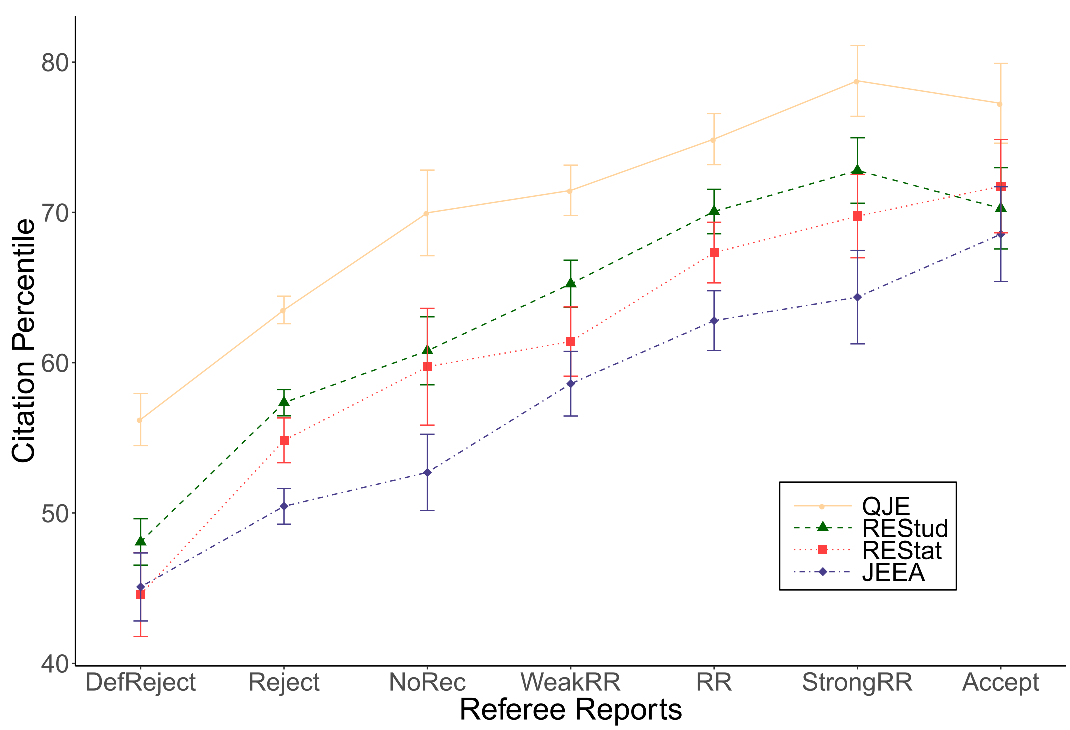


Figure 3c. Citation and Combination of Reports, Papers with 2 Reports, Data

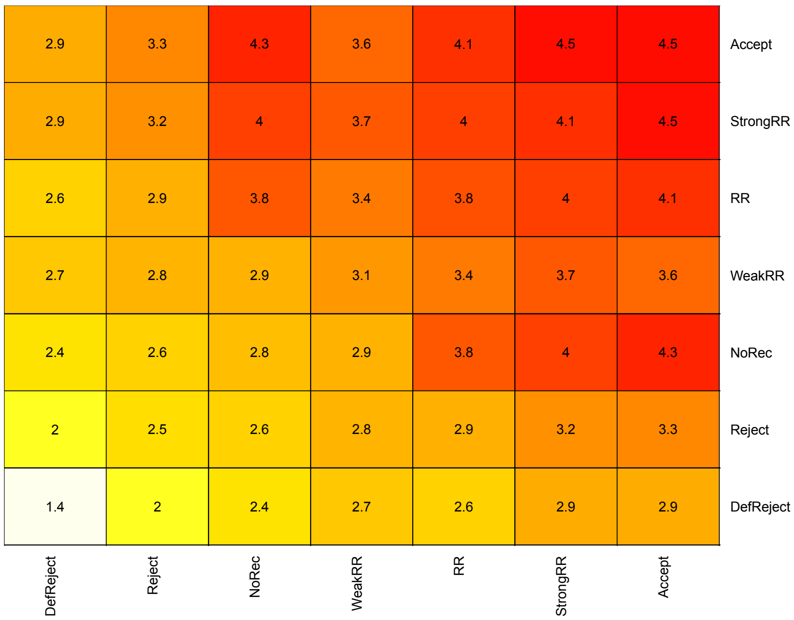
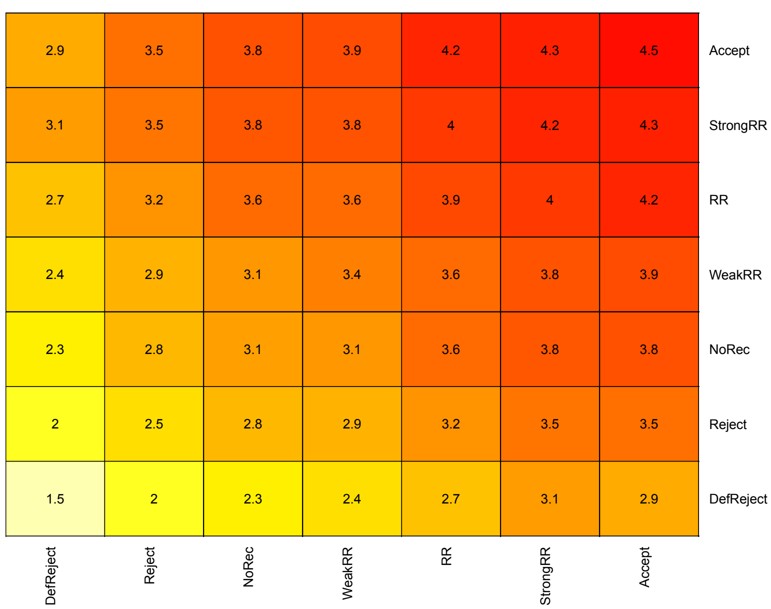


Figure 3d. Citation and Combination of Reports, Papers with 2 Reports, Model Prediction



Notes: Figures 3a and 3b show the weighted average citation measure for a paper receiving a given recommendation. The unit of observation is a referee report, and observations are weighted by the number of referee reports for the paper to ensure that each paper receives equal weight. Standard errors are clustered at the paper level. Figure 3a uses citation percentile as the citation measure, whereas figure 3b uses the Asinh of citations. The higher level of the line for QJE in figure 3a reflects in part the higher desk-rejection rate at the QJE, while in figure 3b it also shows that reviewed papers at the QJE tend to receive higher citations. Figures 3c and 3d display evidence at the paper level, focusing on papers with exactly 2 referee reports. Figure 3c shows a heat map of actual citations for all combinations of 2 reports whereas figure 3d does the same using predicted citations from a regression using only fraction of referee recommendations and year-journal fixed effects. Darker colors in the heat map correspond to higher values of citation.

Figure 4. Referee Recommendations and the Probability of Revise and Resubmit
Figure 4a. Referee Report and R&R Rate, By Journal

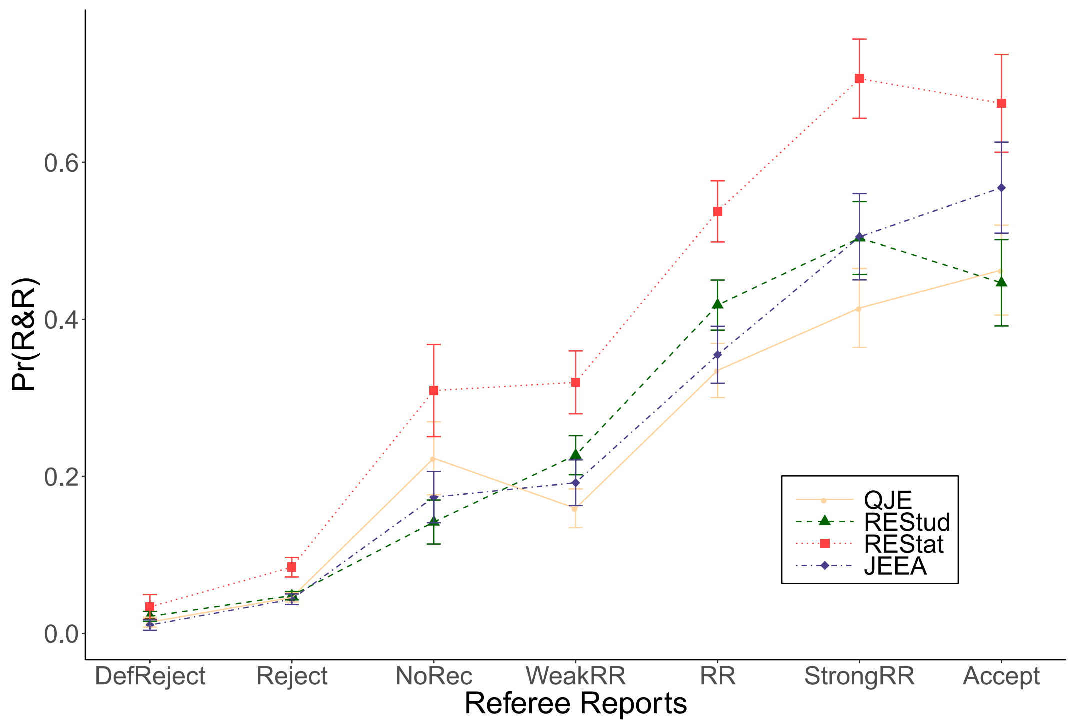


Figure 4b. Combinations of Referee Recommendations and R&R, 2-Report Papers

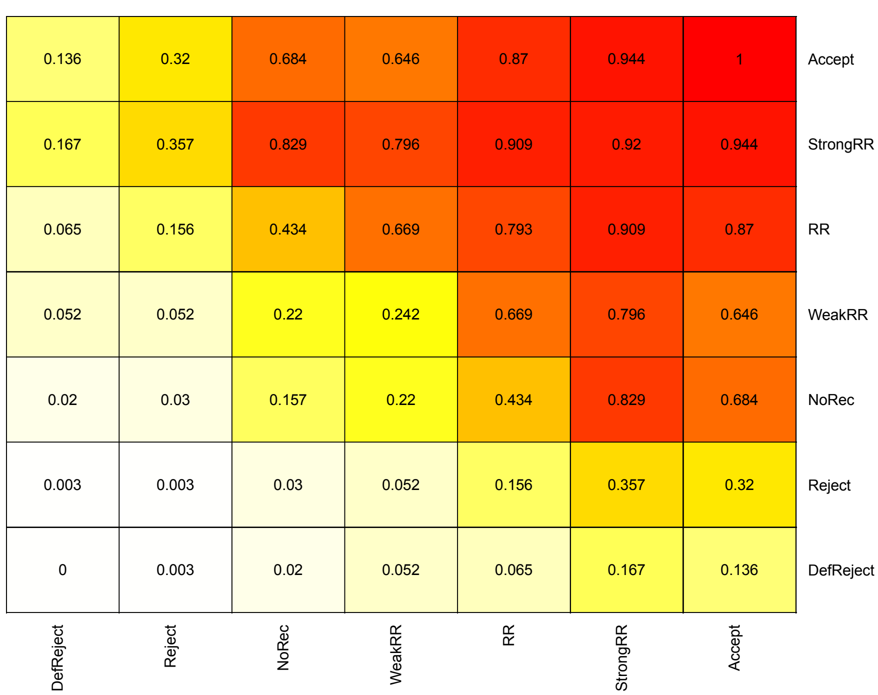


Figure 4c. Combinations of Referee Recommendations and R&R, 2-Report Papers, Model

	DefReject	Reject	NoRec	WeakRR	RR	StrongRR	Accept
Accept	0.195	0.327	0.663	0.757	0.923	0.966	0.963
StrongRR	0.226	0.375	0.737	0.788	0.937	0.975	0.966
RR	0.108	0.211	0.557	0.633	0.863	0.937	0.923
WeakRR	0.024	0.06	0.253	0.324	0.633	0.788	0.757
NoRec	0.014	0.039	0.188	0.253	0.557	0.737	0.663
Reject	0.001	0.004	0.039	0.06	0.211	0.375	0.327
DefReject	0	0.001	0.014	0.024	0.108	0.226	0.195

Notes: Figure 4 displays visual evidence of the correlation between referee reports and the editor's review-and-resubmit (R&R) decision. Figure 4a shows the weighted R&R rate for a paper receiving a given recommendation. The unit of observation is a referee report, so for example the value of the Accept category should be interpreted as the R&R rate for papers with (at least 1) referee recommending Accept, taking into account that the other referee(s) recommendations likely differ. Observations are weighted by the number of referee reports for the paper to ensure that each paper receives equal weight, and standard errors are clustered at the paper level. Figures 4b and 4c display evidence at the paper level, focusing on papers with exactly 2 referee reports. Figure 4b shows a heat map of actual R&R rates for all combinations of 2 reports whereas figure 4c does the same using predicted R&R probabilities from a probit regression using only fraction of referee recommendations and year-journal fixed effects. Darker colors in the heat map correspond to probabilities of R&R.

Figure 5. The Relative Effect of Referee Recommendations and Paper Characteristics on Citations and the Probability of Revise and Resubmit

Figure 5a. Coefficients for Referee Recommendations and Author Publications

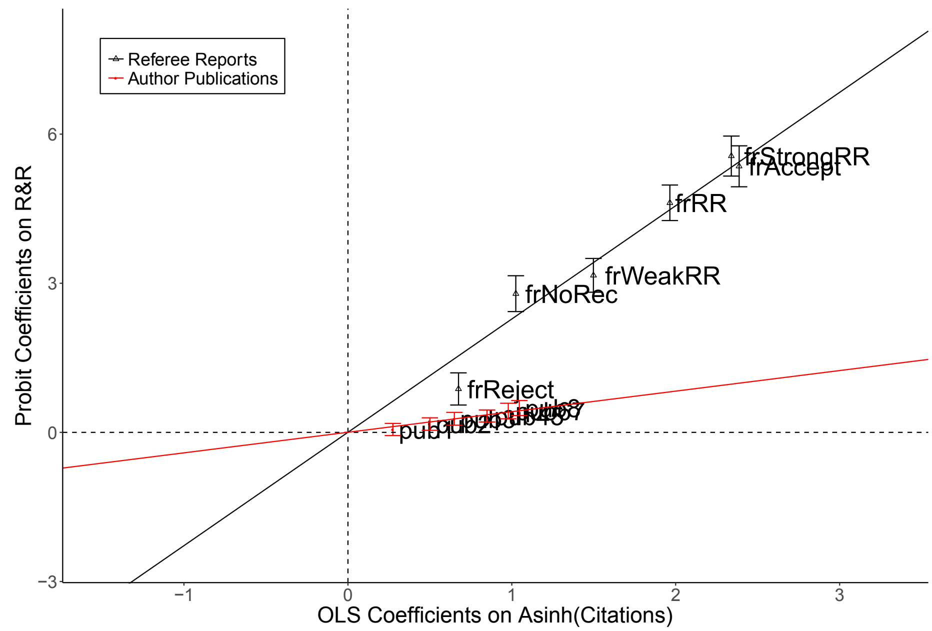
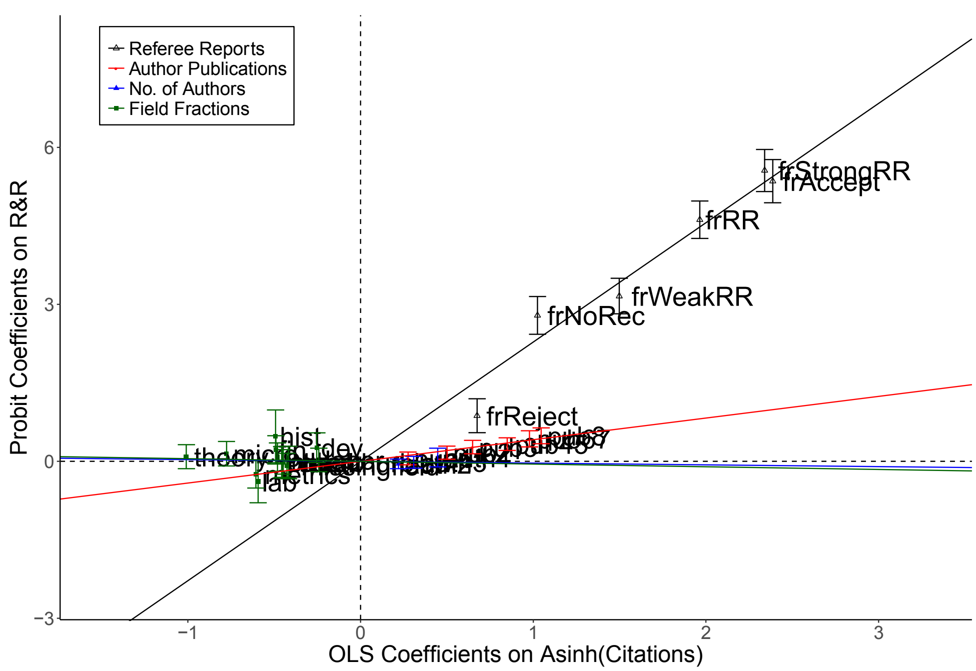
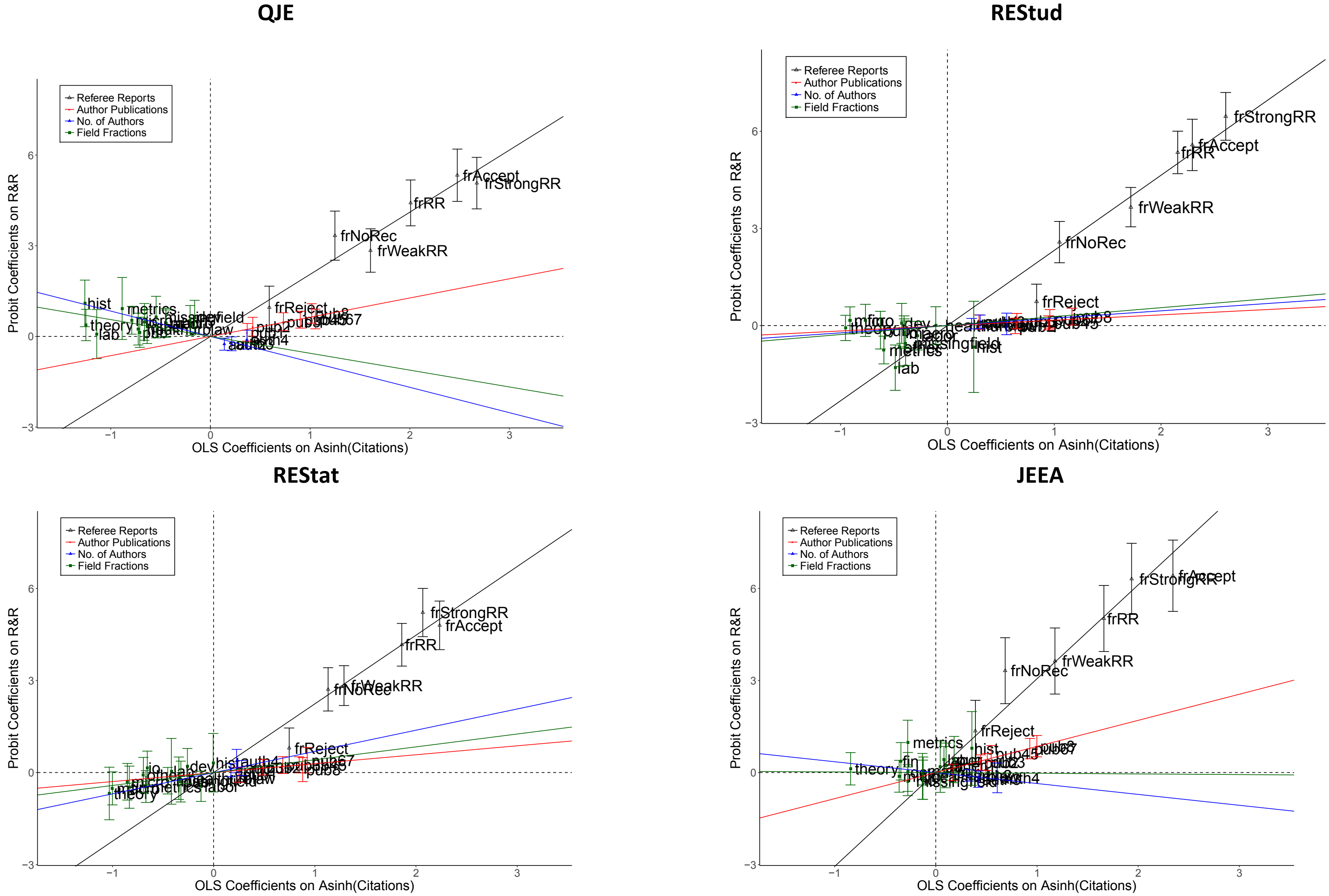


Figure 5b. Coefficients for Referee Recommendations, Author Publications, Number of Authors and Fields



Notes: Figure 5 plots the coefficients from the main specifications of the citation and R&R regressions (Columns 3 and 6 in Table 2). Best fit lines through each group of coefficients are also shown (weighted by the inverse variance of the probit coefficient from the R&R regression). Figure 5a shows that while both referee reports and author publications are both predictive of ex-post citations, the editors' R&R decision is influenced much more by referee reports than by author publications, relative to the extent to which these variables predict citations (as evidenced by the steeper slope of the line for referee reports). Figure 5b shows the coefficients for number of authors and fields in addition.

Figure 6. The Relative Effect of Referee Recommendations and Paper Characteristics on Citations and the Probability of Revise and Resubmit – By Journal



Notes: Figure 6 plots the coefficients from the main specifications of the citation and R&R regressions as in Figure 5b, separately by journal. Best fit lines through each group of coefficients are also shown (weighted by the inverse variance of the probit coefficient from the R&R regression).

Figure 7. The Relationship Between the Editor's Revise and Resubmit Decision and Realized Citations
Figure 7a. Plot by Quantiles in probability of revise-and-resubmit

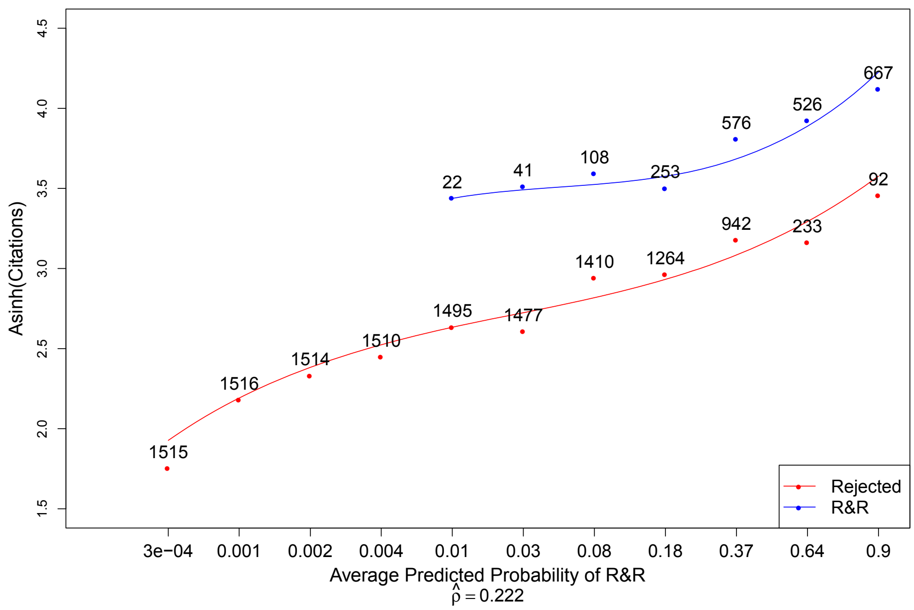
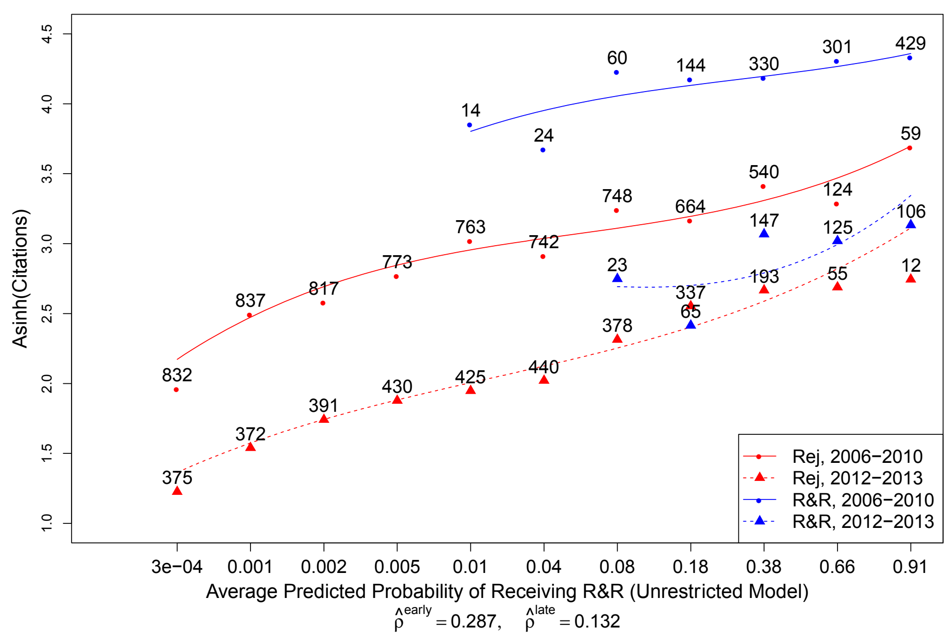


Figure 7b. Separating Bias due to Publication



Notes: Figure 7 shows the average Asinh(citations) by deciles of predicted probability of R&R where the top decile is further split into two ventiles. Figure 7a considers separately papers that were rejected and those that the editor granted a revise-and-resubmit (using papers from the entire sample period), and figure 7b breaks these two groups of papers down further into whether they were submitted between 2006 and 2010, or between 2012 and 2013 (leaving out papers submitted in other years). The smoothing lines are obtained via cubic fits to all data points. The estimate of the editor signal described in the text is shown at the bottom of figure 7a, whereas the editor signal is estimated separately for the early and late period within a single regression by interacting the editor signal term with dummies indicating whether the paper was submitted in the earlier or later period.

Figure 8. The Relationship Between the Editor's Desk Rejection Decision and Realized Citations

Figure 8a. Plot by Quantiles in probability of avoiding desk-rejection

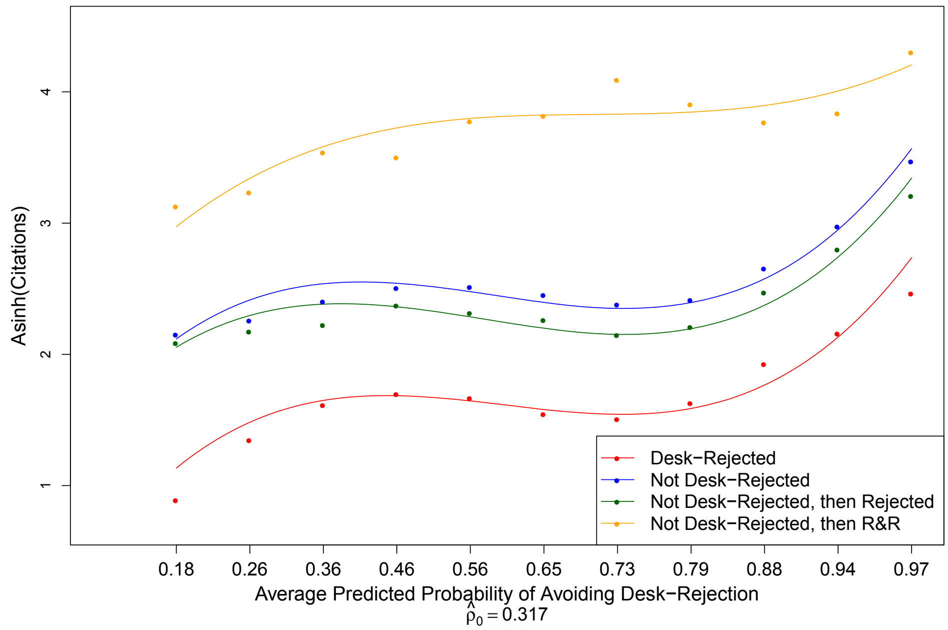
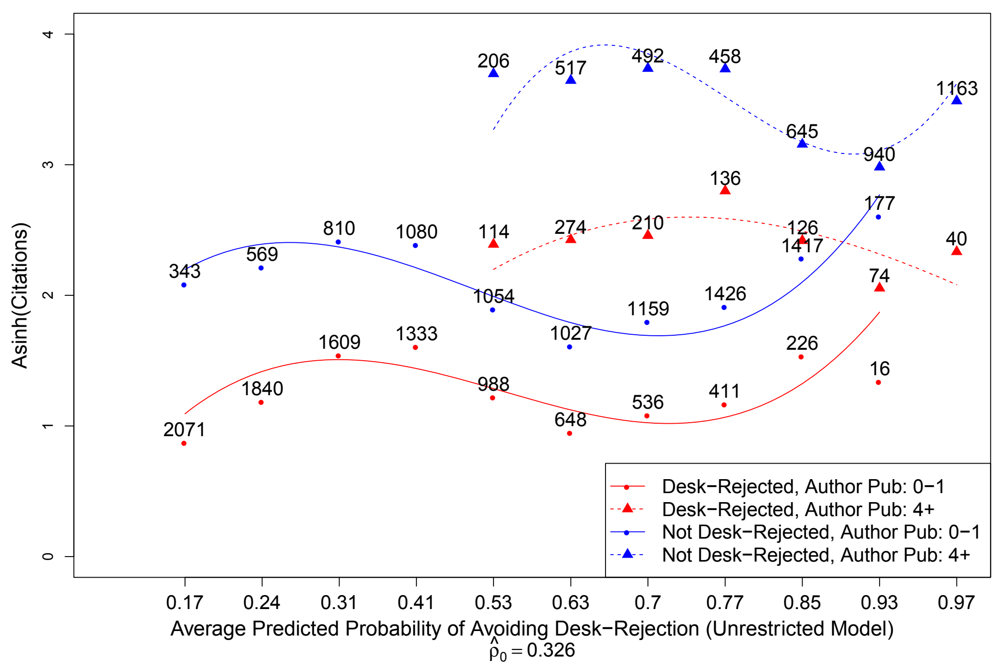


Figure 8b. Plot by Quantiles in probability of avoiding desk-rejection, split by author publications



Notes: Figure 8 shows the average Asinh(citations) by deciles of predicted probability of non-desk-rejection, where the top decile is further split into two ventiles. Figure 8a considers separately papers that were desk-rejected, those that were not but were rejected later on, and those that ultimately received an R&R (using all papers in our data). Figure 8b breaks the desk-rejected and non-desk-rejected papers down further into whether the authors' recent publications were in the 0-1 or 4+ range (leaving out papers submitted by authors with 2-3 recent publications). The smoothing lines are obtained via cubic fits to all data points. The estimate of the editor signal described in the text is shown at the bottom of both figures.

Figure 9. Discounting of Citations of Prolific Authors, Referees

Figure 9a. All Referees

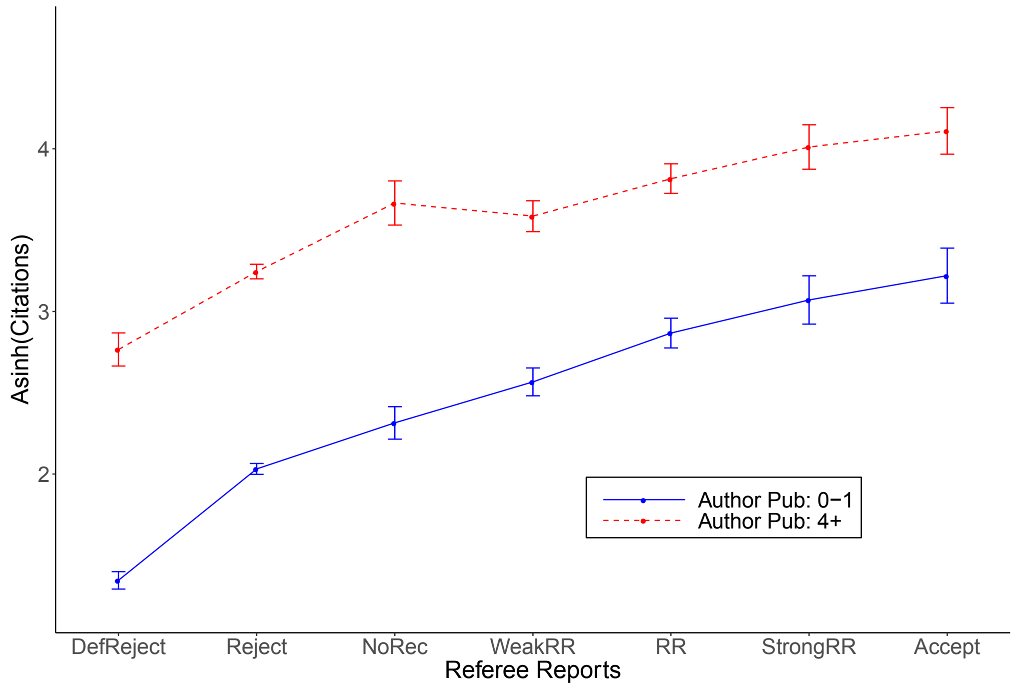
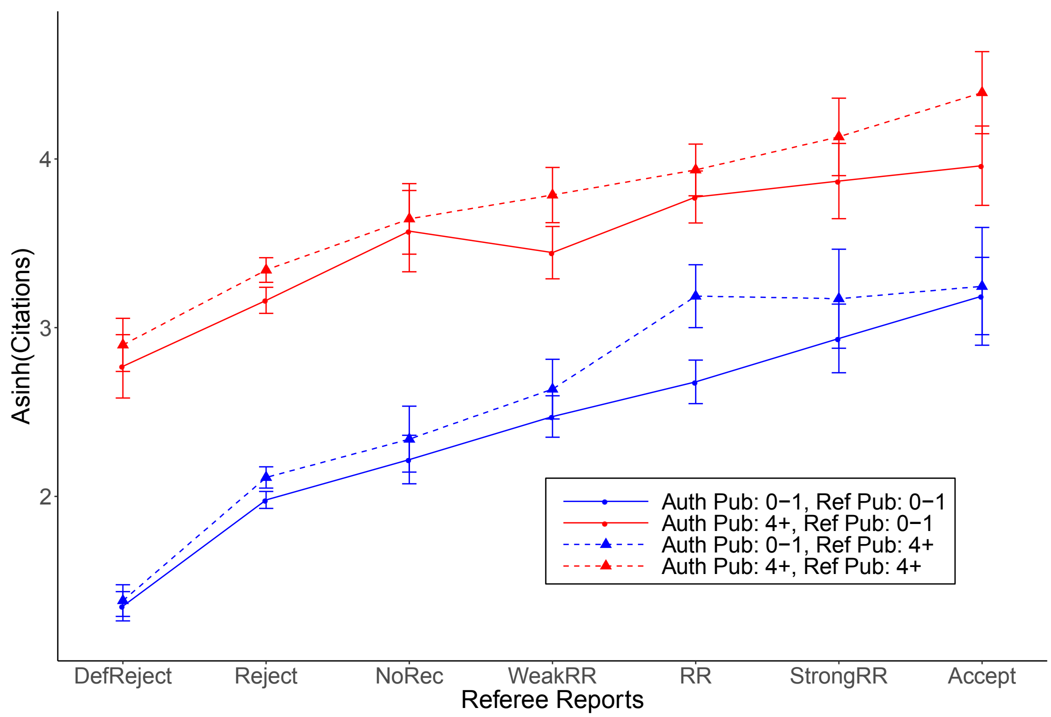


Figure 9b. Split by Prolific and Non-Prolific Referees



Notes: Figure 9 shows the weighted Arsinh(citations) for a paper receiving a given recommendation. Figure 9a shows the results separately for authors with 0-1 recent publications and authors with at least 4 recent publications, while figure 9b splits these two categories further into whether the report was provided by a referee with 0 -1 recent publications or by a referee with at least 4 recent publications. The unit of observation is a referee report, and observations are weighted by the number of referee reports for the paper to ensure that each paper receives equal weight. Standard errors are clustered at the paper level.

Figure 10. Evidence on Citation Discounting from Survey of Economists
Figure 10a. Assessed Relative Citations versus Actual Citation Ratio, Theoretical Cases

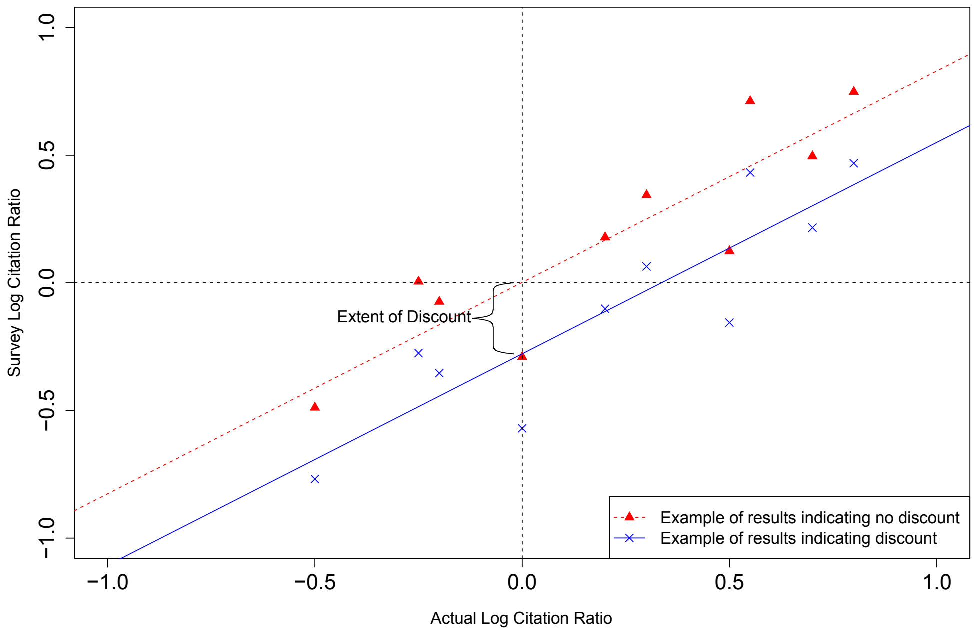
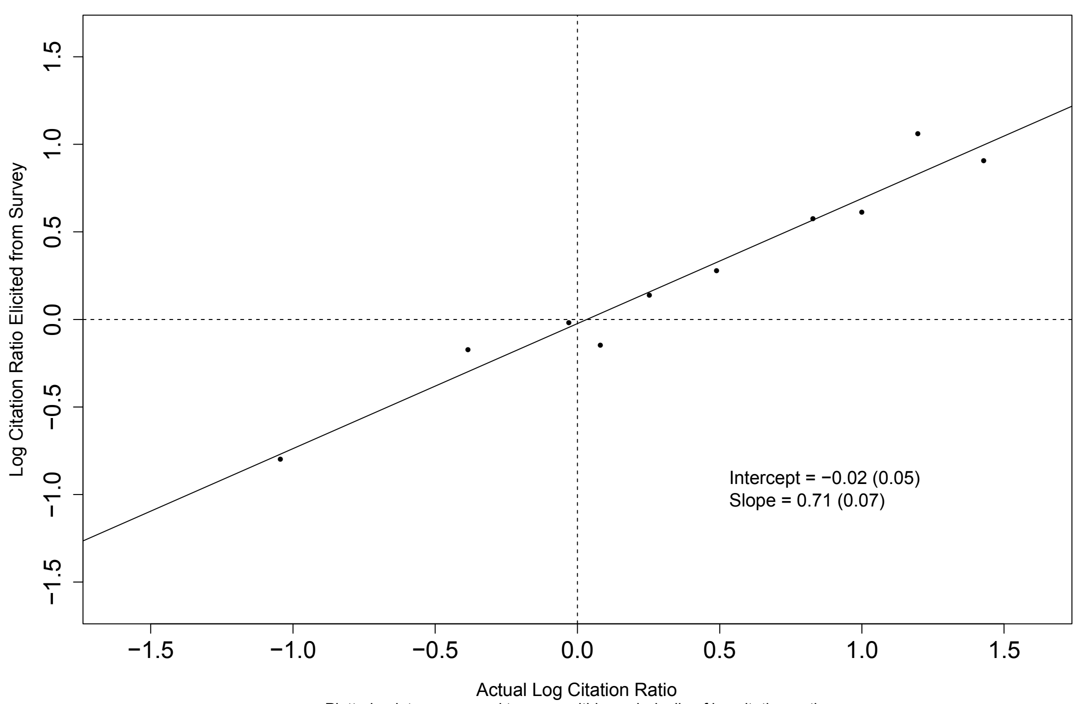
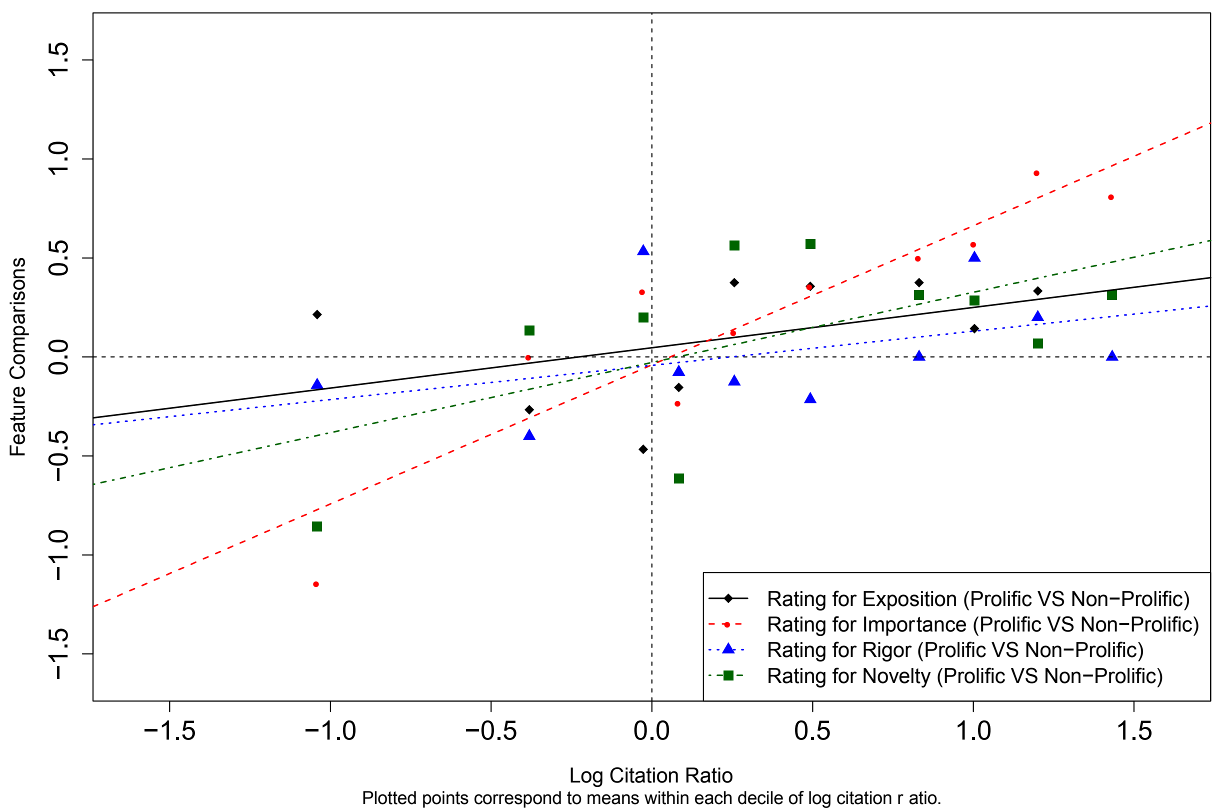


Figure 10b. Assessed Relative Citations versus Actual Citation Ratio, Decile Bins



Plotted points correspond to mean within each decile of log citation r atio.

Figure 10c. Qualitative Assessment of Paper Pairs, Decile Bins



Notes: Figures 10a-c show a few key results from the survey, where we asked respondents to compare pairs of papers that were similar except that one was by a prolific author and the other was by a non-prolific author at the approximate time of submission. Figure 10a is created using simulated data, illustrating two possible scenarios. One possibility, illustrated by the red points (corresponding to deciles of the actual log citation ratio of the paper pairs used for the survey) and the best fit line through the points, is that subjects on average indicate that relative citations merited by the paper by the prolific versus that by the non-prolific authors was about right, in which case citations roughly correspond to quality (i.e. citations for prolific authors need not be discounted relative to those for non-prolific authors to obtain an unbiased measure of quality). Another possibility, illustrated by the blue points and line, is that subjects may on average indicate that the paper by the prolific author receive too many citations relative to the paper by the non-prolific author. In this case, citations by prolific authors will need to be discounted relative to citations by non-prolific authors in order to obtain an unbiased measure of quality. The negative of the estimated intercept indicates the extent to which citations for prolific authors are inflated. Figure 10b shows the actual survey results, which are more in line with the former interpretation, since the estimated intercept is statistically insignificant and close to zero. We winsorized the top and bottom 2% of survey responses of the log citation ratio which subjects thought was justified (as per our pre-analysis registration). Figure 10c displays the results the section of our survey where we asked subjects to compare papers on a scale of -2 to 2 on four dimensions – exposition, importance, rigor and novelty. We plot the average comparisons separately for these four dimensions as a function of the actual log citation ratio between each paper pair, after converting responses so that positive values indicate an evaluation in favor of the paper by the prolific author. The positive slope of all four lines indicate that papers with more citations were also typically judged more positively on each of the four dimensions. The estimated intercepts being close to zero suggests that similar papers by prolific and non-prolific authors that receive similar citations are typically comparable on these four dimensions, a finding that is in line with figure 10b.

Table 1. Summary Statistics For All Submissions and Non-Desk-Rejected Papers

	All Papers					Non-Desk-Rejected Papers				
Journals in Sample:	All	QJE	REStat	JEEA	REStud	All	QJE	REStat	JEEA	REStud
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Google Scholar Citations										
Percentile (Journal/Year)	50.0	50.0	50.0	49.9	50.0	59.5	65.4	58.6	53.6	59.0
	(28.9)	(28.9)	(28.9)	(28.9)	(28.9)	(28.0)	(26.3)	(28.3)	(29.0)	(27.6)
Asinh Citations	2.11	2.19	2.03	2.23	1.99	2.74	3.21	2.62	2.47	2.57
	(1.86)	(1.95)	(1.76)	(1.82)	(1.81)	(1.84)	(1.88)	(1.76)	(1.83)	(1.78)
Editorial Decisions										
Not Desk-Rejected	0.58	0.40	0.46	0.76	0.80	1.00	1.00	1.00	1.00	1.00
Received R&R Decision	0.08	0.04	0.12	0.11	0.08	0.15	0.11	0.26	0.14	0.13
Author Publications in 35 high-impact journals										
Publications: 0	0.46	0.46	0.48	0.45	0.44	0.32	0.24	0.38	0.39	0.30
Publications: 1	0.17	0.16	0.20	0.18	0.15	0.17	0.16	0.20	0.18	0.16
Publications: 2	0.11	0.10	0.11	0.12	0.11	0.13	0.12	0.13	0.13	0.12
Publications: 3	0.08	0.08	0.07	0.09	0.09	0.11	0.11	0.10	0.11	0.12
Publications: 4-5	0.09	0.10	0.08	0.08	0.11	0.14	0.17	0.11	0.10	0.15
Publications: 6-7	0.04	0.05	0.03	0.04	0.05	0.07	0.09	0.05	0.05	0.07
Publications: 8+	0.04	0.05	0.02	0.03	0.05	0.07	0.10	0.04	0.04	0.07
Number of Authors										
1 author	0.38	0.38	0.30	0.37	0.42	0.31	0.26	0.27	0.34	0.35
2 authors	0.39	0.38	0.41	0.41	0.38	0.42	0.42	0.43	0.42	0.42
3 authors	0.19	0.19	0.23	0.18	0.17	0.21	0.24	0.24	0.19	0.19
4+ authors	0.05	0.06	0.06	0.03	0.04	0.05	0.08	0.06	0.04	0.04
Field of Paper										
Development	0.05	0.06	0.05	0.04	0.04	0.05	0.06	0.05	0.04	0.04
Econometrics	0.07	0.04	0.11	0.04	0.09	0.06	0.02	0.09	0.03	0.09
Finance	0.07	0.09	0.04	0.04	0.07	0.06	0.08	0.03	0.04	0.07
Health, Urban, Law	0.05	0.07	0.05	0.03	0.03	0.05	0.08	0.05	0.03	0.03
History	0.01	0.02	0.01	0.01	0.01	0.01	0.02	0.01	0.01	0.01
International	0.06	0.07	0.05	0.06	0.06	0.06	0.07	0.05	0.06	0.05
Industrial Organization	0.05	0.05	0.05	0.05	0.05	0.05	0.04	0.05	0.05	0.06
Lab/Experiments	0.02	0.03	0.01	0.03	0.02	0.03	0.03	0.01	0.03	0.03
Labor	0.11	0.13	0.11	0.11	0.08	0.12	0.18	0.11	0.12	0.09
Macro	0.10	0.11	0.07	0.10	0.12	0.10	0.09	0.07	0.10	0.11
Micro	0.11	0.12	0.05	0.10	0.13	0.11	0.12	0.05	0.10	0.13
Public	0.05	0.06	0.03	0.05	0.05	0.05	0.06	0.03	0.05	0.05
Theory	0.09	0.08	0.02	0.07	0.17	0.10	0.06	0.02	0.07	0.19
Unclassified	0.06	0.08	0.05	0.05	0.05	0.05	0.07	0.05	0.05	0.04
Missing Field	0.11	0.02	0.30	0.23	0.02	0.10	0.01	0.33	0.20	0.01
Referee Recommendations										
Fraction Definitely Reject						0.12	0.13	0.10	0.11	0.14
Fraction Reject						0.54	0.60	0.44	0.50	0.56
Fraction with No Rec'n						0.06	0.03	0.06	0.10	0.05
Fraction Weak R&R						0.10	0.09	0.13	0.11	0.10
Fraction R&R						0.10	0.08	0.16	0.11	0.09
Fraction Strong R&R						0.04	0.03	0.07	0.04	0.03
Fraction Accept						0.03	0.03	0.05	0.04	0.03
Years	2003-13	2005-13	2006-13	2003-13	2005-13	2003-13	2005-13	2006-13	2003-13	2005-13
Number of Observations	29,868	10,824	5,767	4,942	8,335	15,177	4,195	2,391	3,280	5,311

Notes: Table presents information on mean characteristics of all submitted papers (columns 1-5), and for non-desk-rejected papers (columns 6-10). The sample of non-desk-rejected papers also excludes papers with only 1 referee assigned. The Google Scholar citation percentile is computed within a year-journal cohort of submissions. To avoid ties, we randomly jitter the citations before calculating percentiles. Author publications are based on publications in a set of 35 high-impact journals (Appendix Table 1) in the 5 years prior to submission. In case of multiple authors, the measure is the maximum for all coauthors. Field is based on JEL codes at paper submission. Indicators of fields for a paper that lists N codes are set to 1/N. For example, a paper with JEL codes that match labor and theory will be coded 0.5 for labor and 0.5 for theory.

Table 2. Models for Realized Citations and Revise-and-Resubmit Decision

	OLS Models for Asinh of Google Scholar Citations			Probit Models for Receiving Revise-and-Resubmit Decision		
	(1)	(2)	(3)	(4)	(5)	(6)
Fractions of Referee Recommendations						
Reject	0.83		0.67	0.87		0.87
	(0.06)		(0.06)	(0.16)		(0.16)
No Recommendation	1.26		1.02	2.79		2.79
	(0.10)		(0.10)	(0.17)		(0.18)
Weak R&R	1.78		1.50	3.16		3.16
	(0.08)		(0.08)	(0.16)		(0.17)
R&R	2.37		1.96	4.64		4.62
	(0.08)		(0.08)	(0.17)		(0.18)
Strong R&R	2.76		2.34	5.58		5.56
	(0.11)		(0.11)	(0.19)		(0.20)
Accept	2.78		2.39	5.39		5.35
	(0.12)		(0.12)	(0.20)		(0.20)
Author Publications in 35 high-impact journals						
1 Publication		0.40	0.27		0.26	0.06
		(0.04)	(0.04)		(0.04)	(0.06)
2 Publications		0.66	0.50		0.36	0.17
		(0.05)	(0.05)		(0.05)	(0.06)
3 Publications		0.88	0.65		0.56	0.27
		(0.05)	(0.05)		(0.05)	(0.06)
4-5 Publications		1.11	0.85		0.68	0.33
		(0.05)	(0.05)		(0.05)	(0.06)
6-7 Publications		1.29	0.98		0.81	0.43
		(0.06)	(0.06)		(0.05)	(0.08)
8+ Publications		1.39	1.05		0.90	0.49
		(0.06)	(0.06)		(0.06)	(0.08)
Number of Authors						
2 authors		0.19	0.22		-0.12	-0.05
		(0.04)	(0.03)		(0.03)	(0.05)
3 authors		0.24	0.31		-0.16	-0.02
		(0.04)	(0.04)		(0.04)	(0.06)
4+ authors		0.40	0.45		-0.05	0.07
		(0.07)	(0.06)		(0.06)	(0.09)
Indicators for Field of Paper	No	Yes	Yes	No	Yes	Yes
Indicators for Journal-Year	Yes	Yes	Yes	Yes	Yes	Yes
R^2 / pseudo R^2	0.20	0.20	0.26	0.48	0.07	0.49

Notes: See notes to Table 1. The sample for all models is 15,177 non-desk-rejected papers with at least two referees assigned. All models include indicators for journal-year cohort. Dependent variable for OLS models in columns 1-3 is asinh of Google Scholar citations. Dependent variable in probit models in columns 4-6 is indicator for receiving revise and resubmit decision. Robust standard errors in parentheses.

Table 3. Models for Alternative Measures of Citations

	OLS Model for Asinh(GS Citations)	OLS Model for GS Citation Percentile	Probit Model for Top Decile of GS Citations	OLS Model for Log(1+ GS Citations)	OLS Model for Asinh(GS Citations)	OLS Model for Asinh (SSCI Citations)	OLS Model for SSCI Citation Percentile
Sample Years	**All Years**	**All Years**	**All Years**	**All Years**	**2006-2010**	**2006-2010**	**2006-2010**
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Fractions of Ref. Recommendations							
Reject	0.67	10.87	0.39	0.57	0.72	0.27	6.12
	(0.06)	(0.91)	(0.07)	(0.05)	(0.08)	(0.05)	(1.30)
No Recommendation	1.02	16.22	0.68	0.88	1.10	0.75	14.55
	(0.10)	(1.58)	(0.10)	(0.09)	(0.14)	(0.10)	(2.28)
Weak R&R	1.50	23.64	0.94	1.30	1.54	0.82	15.41
	(0.08)	(1.27)	(0.09)	(0.07)	(0.11)	(0.08)	(1.87)
R&R	1.96	30.81	1.31	1.71	2.10	1.32	24.90
	(0.08)	(1.23)	(0.09)	(0.07)	(0.11)	(0.08)	(1.86)
Strong R&R	2.34	36.30	1.58	2.06	2.53	1.88	33.03
	(0.11)	(1.71)	(0.11)	(0.10)	(0.15)	(0.12)	(2.50)
Accept	2.39	36.85	1.55	2.11	2.63	2.24	40.14
	(0.12)	(1.76)	(0.11)	(0.11)	(0.15)	(0.14)	(2.59)
Author Publications in 35 high-impact journals							
1 Publication	0.27	4.20	0.18	0.23	0.27	0.11	1.67
	(0.04)	(0.66)	(0.05)	(0.03)	(0.06)	(0.04)	(0.97)
2 Publications	0.50	7.90	0.33	0.43	0.57	0.28	5.54
	(0.05)	(0.73)	(0.05)	(0.04)	(0.06)	(0.05)	(1.08)
3 Publications	0.65	10.18	0.40	0.56	0.64	0.27	5.57
	(0.05)	(0.77)	(0.05)	(0.04)	(0.07)	(0.05)	(1.18)
4-5 Publications	0.85	13.14	0.53	0.74	0.88	0.51	9.99
	(0.05)	(0.71)	(0.05)	(0.04)	(0.06)	(0.05)	(1.10)
6-7 Publications	0.98	14.76	0.68	0.86	0.97	0.49	8.98
	(0.06)	(0.91)	(0.06)	(0.05)	(0.08)	(0.07)	(1.44)
8+ Publications	1.05	15.64	0.72	0.93	1.04	0.65	10.71
	(0.06)	(0.92)	(0.06)	(0.05)	(0.09)	(0.08)	(1.51)
Number of Authors							
2 authors	0.22	3.62	0.08	0.18	0.31	0.07	1.72
	(0.03)	(0.54)	(0.04)	(0.03)	(0.05)	(0.04)	(0.79)
3 authors	0.31	5.26	0.14	0.26	0.45	0.10	2.72
	(0.04)	(0.66)	(0.04)	(0.04)	(0.06)	(0.04)	(0.99)
4+ authors	0.45	7.45	0.29	0.39	0.47	0.05	-0.11
	(0.06)	(1.01)	(0.06)	(0.06)	(0.10)	(0.08)	(1.73)
No. of Observations	15,177	15,177	15,177	15,177	8,208	8,208	8,208
R^2 / pseudo R^2	0.26	0.19	0.15	0.27	0.24	0.19	0.11

Notes: See notes to Tables 1 and 2. The samples for this table includes non-desk-rejected papers with at least two referees assigned. All models include journal-year dummies and controls for field(s). Models in columns 1-5 use Google Scholar (GS) citations. Models in columns 6-7 use SSCI Citation counts. Since SSCI only counts citations in published papers, we restrict the sample to submissions from 2006-2010 to allow time for papers to accumulute citations in published works. Robust standard errors in parentheses.

Table 4. Models for Citations and R&R with Additional Measures of Author and Institutional

	OLS Models for Asinh of GS Citations			Probit Models for R&R Decision		
	Full Sample	Full Sample	JEEA/REStud	Full Sample	Full Sample	JEEA/REStud
	(1)	(2)	(3)	(4)	(5)	(6)
Fractions of Referee Recommendations						
Reject	0.67	0.66	0.66	0.87	0.87	0.86
	(0.06)	(0.06)	(0.07)	(0.16)	(0.16)	(0.23)
No Recommendation	1.02	0.97	0.83	2.79	2.76	2.75
	(0.10)	(0.10)	(0.13)	(0.18)	(0.18)	(0.26)
Weak R&R	1.50	1.44	1.40	3.16	3.14	3.49
	(0.08)	(0.08)	(0.11)	(0.17)	(0.17)	(0.25)
R&R	1.96	1.89	1.78	4.62	4.61	5.00
	(0.08)	(0.08)	(0.11)	(0.18)	(0.18)	(0.26)
Strong R&R	2.34	2.24	2.11	5.56	5.52	6.08
	(0.11)	(0.11)	(0.16)	(0.20)	(0.20)	(0.29)
Accept	2.39	2.25	2.15	5.35	5.30	5.63
	(0.12)	(0.12)	(0.17)	(0.20)	(0.20)	(0.30)
Author Publications in 35 high-impact journals up to 5 years before submission						
1 Publication	0.27	0.25	0.25	0.06	-0.01	-0.02
	(0.04)	(0.04)	(0.05)	(0.06)	(0.06)	(0.08)
2 Publications	0.50	0.43	0.51	0.17	0.03	-0.09
	(0.05)	(0.05)	(0.06)	(0.06)	(0.07)	(0.09)
3 Publications	0.65	0.52	0.50	0.27	0.07	-0.07
	(0.05)	(0.05)	(0.07)	(0.06)	(0.07)	(0.10)
4-5 Publications	0.85	0.63	0.63	0.33	0.06	-0.10
	(0.05)	(0.05)	(0.07)	(0.06)	(0.08)	(0.10)
6-7 Publications	0.98	0.63	0.59	0.43	0.07	-0.11
	(0.06)	(0.07)	(0.10)	(0.08)	(0.10)	(0.14)
8+ Publications	1.05	0.58	0.64	0.49	0.04	-0.17
	(0.06)	(0.08)	(0.11)	(0.08)	(0.10)	(0.14)
Author Publications in Top 5 Journals						
1 Publication		0.29	0.20		0.21	0.19
		(0.04)	(0.05)		(0.05)	(0.07)
2 Publications		0.44	0.28		0.26	0.25
		(0.06)	(0.08)		(0.07)	(0.10)
3+ Publications		0.57	0.36		0.44	0.41
		(0.06)	(0.09)		(0.08)	(0.11)
Author Publications in 35 high-impact journals up to 6-10 years before submission						
1-3 Publications		-0.11	-0.14		0.15	0.30
		(0.03)	(0.05)		(0.05)	(0.07)
4+ Publications		0.06	0.03		0.16	0.34
		(0.05)	(0.06)		(0.06)	(0.09)
Rank of Authors' Institution						
US: 1-10			0.51			0.25
			(0.05)			(0.06)
US: 11-20			0.43			0.30
			(0.06)			(0.08)
Europe: 1-10			0.35			0.05
			(0.05)			(0.07)
Rest of the World: 1-5			-0.26			0.23
			(0.12)			(0.18)
Number of Observations	15,177	15,177	8,591	15,177	15,177	8,591
R^2 / pseudo R^2	0.26	0.27	0.25	0.49	0.49	0.51

Notes: See notes to Tables 1 and 2. The sample includes non-desk-rejected papers with at least two referees assigned. All models include controls for field(s) and dummies for journal-year. Ranking of authors's institutions for US institutions are taken from Ellison (2013), while the rankings for Europe and the rest of the world are taken from the QS 2014 rankings. Information on authors' institutions are only available for REStud and JEEA. Robust standard errors in parentheses.

Table 5. Referee Discounting of Author Publications in Citation Regressions

	OLS Models for Asinh of Google Scholar Citations					
Data Set:	Editorial Express Submissions				Published Papers from Econlit	
Sample:	All non-Desk-rejected Submissions		Submission with R&R		Publications in Our 4 Journals, 2008-15	Publications in Top-5 Journals, 1997-2012
	(1)	(2)	(3)	(4)	(5)	(6)
Author Publications in 35 high-impact journals						
1 Publication	0.40	0.27	0.23	0.18	0.03	0.00
	(0.04)	(0.04)	(0.12)	(0.12)	(0.14)	(0.07)
2 Publications	0.66	0.50	0.20	0.18	0.10	0.13
	(0.05)	(0.05)	(0.14)	(0.13)	(0.16)	(0.07)
3 Publications	0.88	0.65	0.58	0.53	0.51	0.20
	(0.05)	(0.05)	(0.13)	(0.13)	(0.14)	(0.08)
4-5 Publications	1.11	0.85	0.78	0.72	0.43	0.27
	(0.05)	(0.05)	(0.11)	(0.11)	(0.13)	(0.07)
6-7 Publications	1.29	0.98	0.59	0.55	0.47	0.32
	(0.06)	(0.06)	(0.14)	(0.14)	(0.16)	(0.08)
8+ Publications	1.39	1.05	0.88	0.82	0.61	0.46
	(0.06)	(0.06)	(0.15)	(0.15)	(0.15)	(0.08)
Fractions of Referee Recommendations						
Reject		0.67		-0.56		
		(0.06)		(0.41)		
No Recommendation		1.02		-0.39		
		(0.10)		(0.44)		
Weak R&R		1.50		-0.27		
		(0.08)		(0.41)		
R&R		1.96		0.17		
		(0.08)		(0.40)		
Strong R&R		2.34		0.45		
		(0.11)		(0.41)		
Accept		2.39		0.37		
		(0.12)		(0.41)		
Indicators for Number of Authors	Yes	Yes	Yes	Yes	Yes	Yes
Indicators for Field of Paper	Yes	Yes	Yes	Yes	Yes	Yes
Indicators for Journal-Year	Yes	Yes	Yes	Yes	Yes	Yes
Sample Size	15,177	15,177	2,209	2,209	1,534	5,266
R^2	0.20	0.26	0.25	0.27	0.26	0.32

Notes: See notes to Table 1. The sample for models in Columns 1 and 2 is 15,117 non-desk-rejected papers with at least two referees assigned. The models in Columns 3 and 4 include only papers which ultimately received an invitiation to Revise and Resubmit. The sample in Column 5 includes the sample of papers published in one of the 4 journals considered in the years 2008-2015. This sample, obtained from Econlit, matches approximately the sample of papers receiving an R&R invitation, assuming a 2-year delay between submission and publication. The sample in Column 6, also from Econlit, includes all papers published in the traditional jop-5 economics journals between 1997 and 2012, assuming also a 2-year delay between submission and publication. The dependent variable is asinh of Google Scholar citations. Robust standard errors in parentheses.

Table 6. Within-Pair Models for Assessment of Relative Quality of Papers

Panel A: Relationship Between Preferred Citation Ratio and Actual Citation Ratio

	Models for Log of Elicited Citation Ratio from Survey Respondents:				
	Full Sample	**Full Sample (Weighed Least Squares)**	**Pairs with Log(Relative Citations) in [0.5, 0.5]**	**Responses by PhD Students and Non-Prolific Faculty**	**Responses by Prolific Faculty**
	(1)	(2)	(3)	(4)	(5)
Log of Actual Citation Ratio	0.71	0.70	0.57	0.64	0.74
	(0.07)	(0.07)	(0.19)	(0.09)	(0.10)
Constant	-0.02	-0.01	-0.03	0.02	-0.10
	(0.05)	(0.05)	(0.06)	(0.06)	(0.08)
No. of Pairs of Papers Evaluated	148	148	65	76	34
R-squared	0.53	0.56	0.09	0.50	0.63

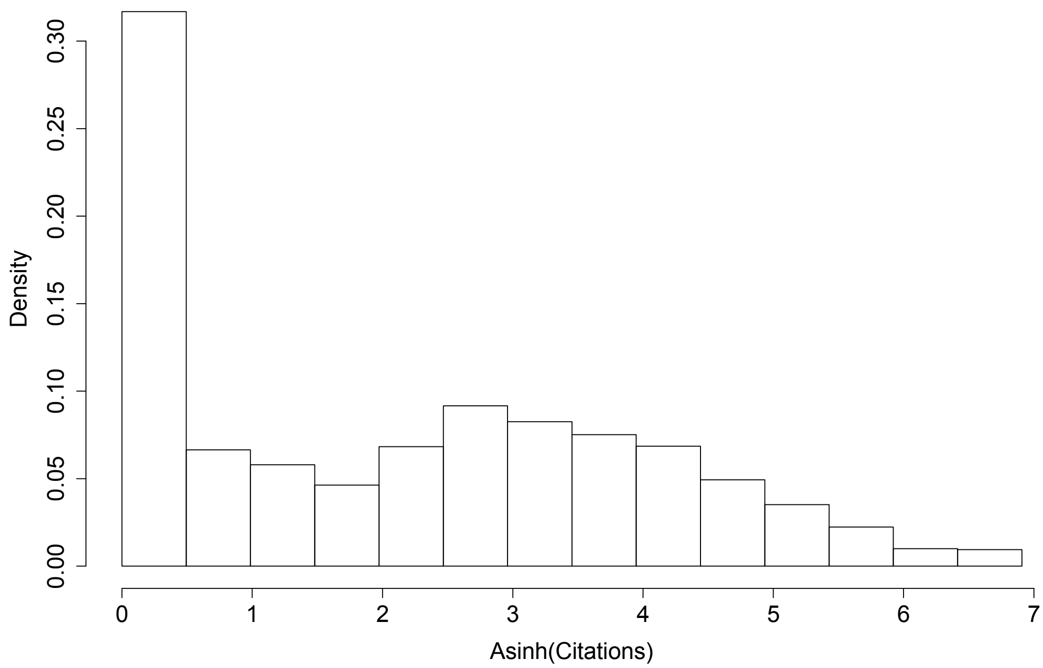
Notes: Table reports regression models (fit by OLS for models in columns 1,3,4,5 and weighted least squares for model in column 2) in which the dependent variable is the log of the respondent's preferred citation ratio for the paper in a given pair written by the more prolific author, and the dependent variable is the log of the actual relative citation ratio. See text for derivation of preferred citation ratio. Sample includes respondent-pair observations for sample indicated in column heading. Weight for model in column 2 is the inverse number of responents who evaluated the specific pair of papers, so each distinct pair is equally weighted. Standard errors (clustered by paper pair) in parentheses.

Panel B: Relationship Between Relative Quality (in 4 Dimensions) and Relative Citations

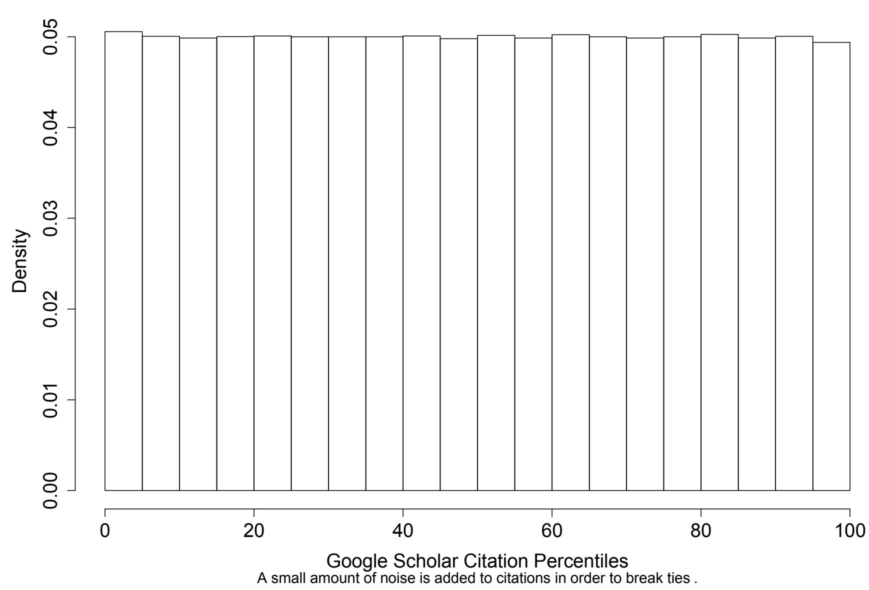
	Dimension of Relative Quality (5 point scale from -2 to 2):			
	Exposition	**Importance**	**Rigor**	**Novelty**
	(1)	(2)	(3)	(4)
Log of Actual Citation Ratio	0.20	0.70	0.17	0.35
	(0.18)	(0.14)	(0.16)	(0.16)
Constant	0.05	-0.04	-0.04	-0.03
	(0.14)	(0.13)	(0.13)	(0.17)
No. of Pairs of Papers Evaluated	148	148	148	148
R-squared	0.02	0.17	0.01	0.04

Notes: Table reports regression models fit by OLS in which the dependent variable is the log of the respondent's preferred citation ratio for the paper in a given pair authored by the more prolific author, and the dependent variable is the respondent's relative assessment of the quality of the paper in a given pair in the dimension indicated by the column heading on the log of the relative citation ratio. Respondents compare papers in a pair using a 5 point Likkert scale which is converted to a linear scale ranging from -2 to 2, with a more positive number indicating a preference for the paper by the prolific author. Sample includes 148 respondent-pair observations. Standard errors (clustered by paper pair) in parentheses.

Online Appendix Figure 1. Distribution of Citation Variables
Online Appendix Figure 1a. Distribution of Asinh of Citations



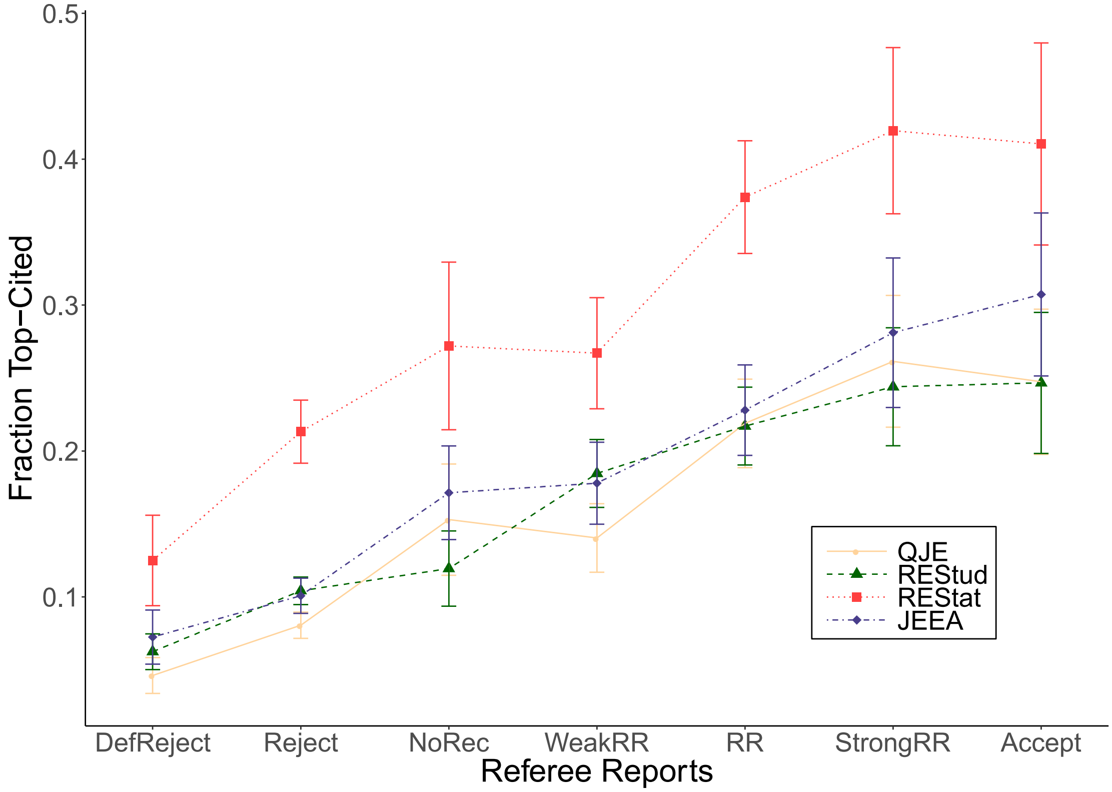
Online Appendix Figure 1b. Distribution of Percentile Citation



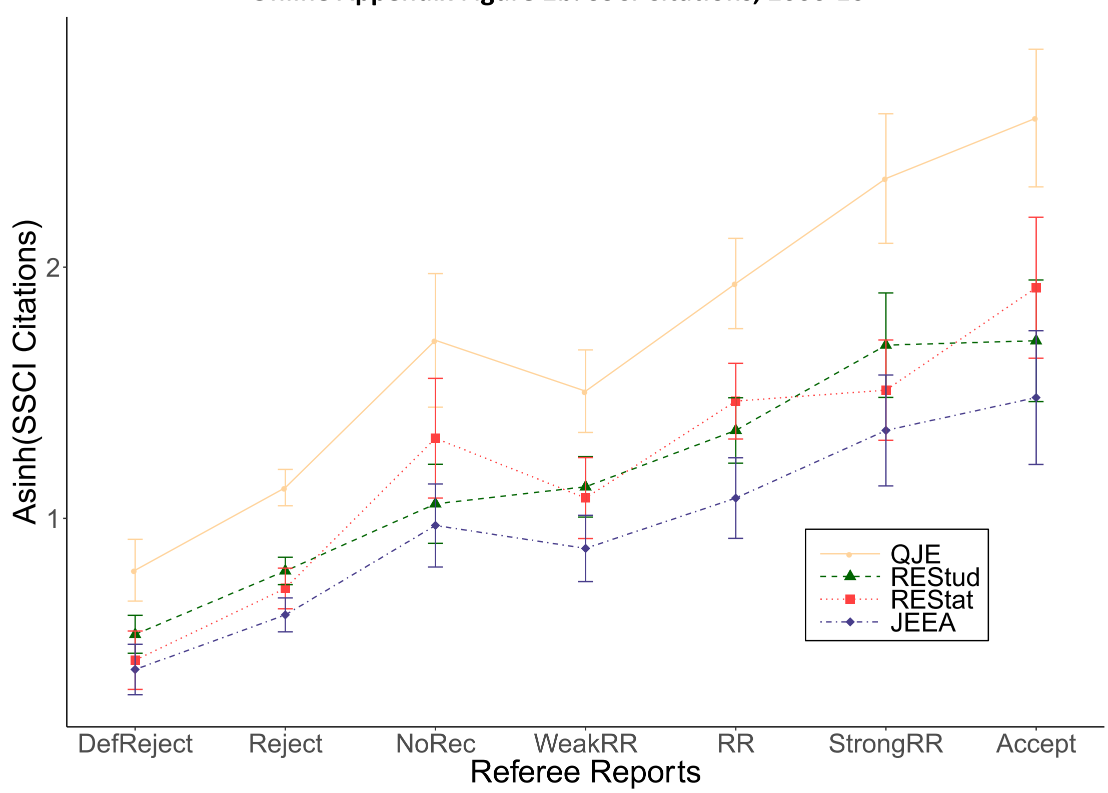
Notes: Online appendix figure 1 shows the distribution of the two main citation measures we use: the percentile citation by journal-year cohort and the Arsinh of Google Scholar citations. Panel 1a shows the distribution for the percentile measure, which is close to uniform given the definition of this variable, and the fact that we jitter sightly the citations to break ties. Panel 1b shows the distribution for Asinh(Citations), which exhibits bunching at the lower end (which is unsurprising given that almost 32 percent of papers in our sample have zero citations). Citation itations are top-coded at 500 (200 for REStud), which is about 6.9 after the Arsinh transformation.

Online Appendix Figure 2. Referee Recommendations and Citations, Robustness

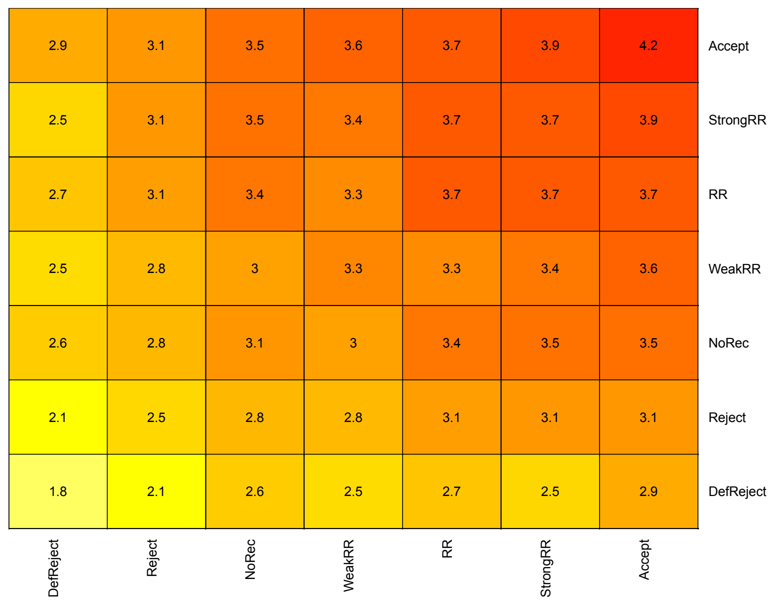
Online Appendix Figure 2a. Robustness using Fraction top-cited



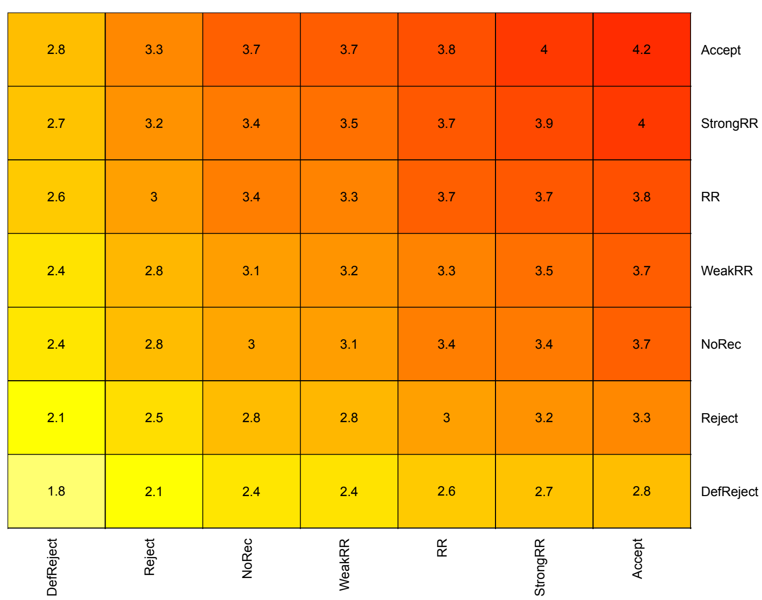
Online Appendix Figure 2b. SSCI Citations, 2006-10



Online Appendix Figure 2c. Heat Map for paper with 3 reports



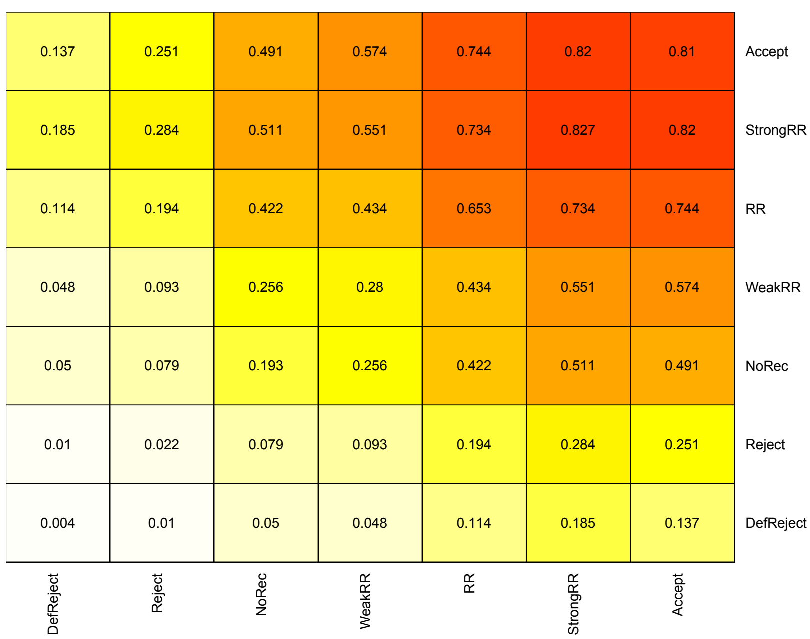
Online Appendix Figure 2d. Heat Map for paper with 3 reports, Model prediction



Notes: Online Appendix Figures 2a-d provide robustness checks of the correlation between referee reports and citations displayed visually in Figures 3a-d in the text. Panels a and b plot the weighted average citation measure for a paper receiving a given recommendation. The unit of observation is a referee report, and observations are weighted by the number of referee reports for the paper to ensure that each paper receives equal weight. Standard errors are clustered at the paper level. Panel a uses top-cited as the dependent variable, where a paper is defined as top-cited if its citations are in the top X% of its journal-year cohort (and X is defined as the percentage of papers receiving R&R's in that journal-year cohort). Panel b uses SSCI citations, and given that these take longer to accrue than Google Scholar citations, we restrict our attention to papers submitted between 2006 and 2010. Panels c and d display evidence at the paper level, focusing on papers with exactly 3 referee reports. Panel c shows a heat map of actual citations for all combinations of 3 reports whereas figure d does the same using predicted citations from a regression using only fraction of referee recommendations and year-journal fixed effects (Column 1 of Table 2). All possible combinations of 2 reports out of the 3 reports for each paper are considered, and darker colors in the heat map correspond to higher values of citation.

Online Appendix Figure 3a. R&R Probability, Heat Map for paper with 3 reports

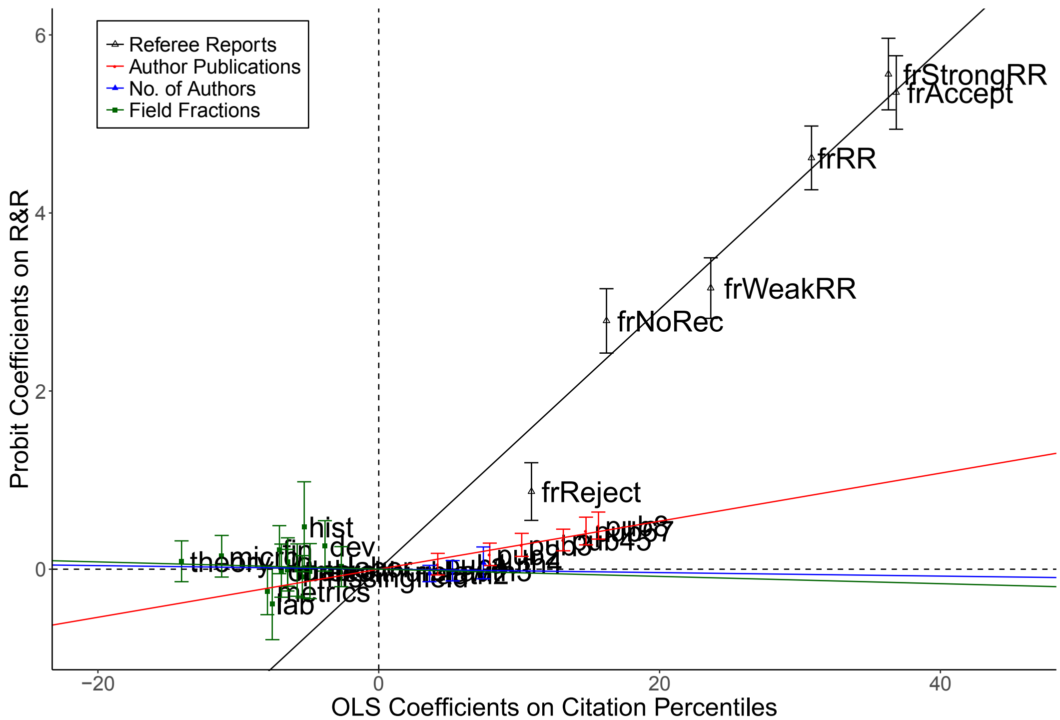
Online Appendix Figure 3b. Heat Map for paper with 3 reports, Model prediction



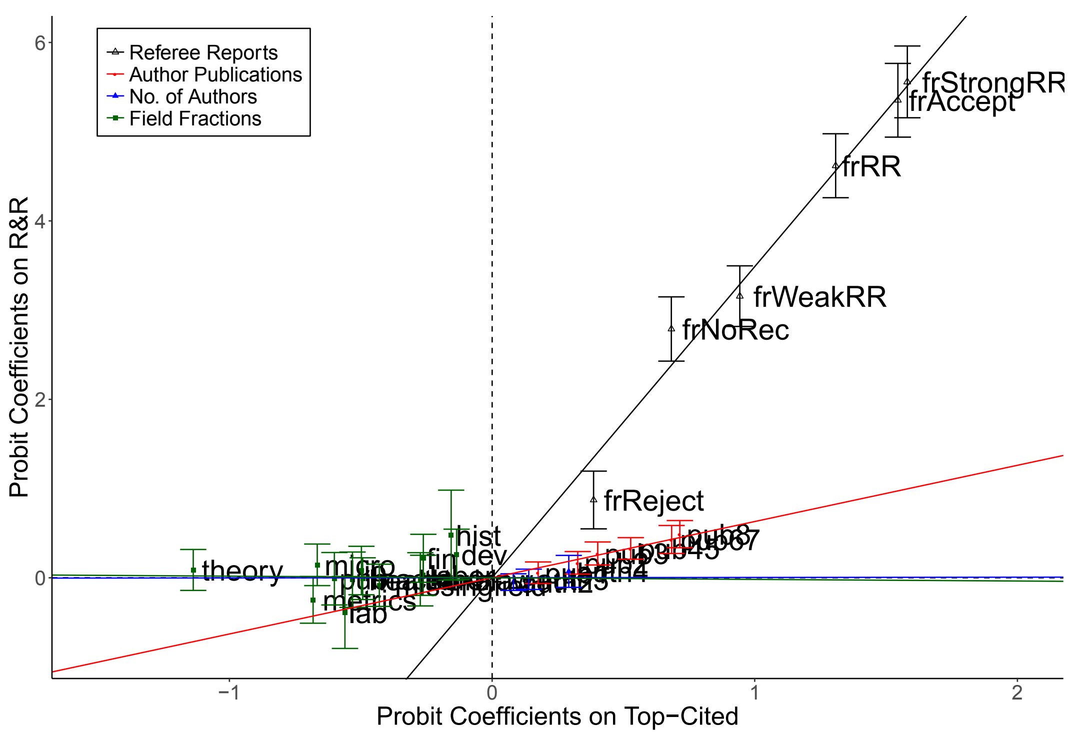
Notes: Online appendix figure 3 checks the robustness of the correlation between referee reports and editor R&R displayed visually in Figures 4b-c in the text. Unlike in Figures 4b-c which focus on papers with 2 reports, we focus here on papers with 3 reports. Panel a shows a heat map of actual R&R rates for all combinations of 2 reports whereas figure b does the same using predicted R&R probabilities from a regression using only fraction of referee recommendations and year-journal fixed effects (Column 3 in Table 2). All possible combinations of 2 reports out of the 3 reports for each paper are considered, and darker colors in the heat map correspond to higher R&R probabilities

Online Appendix Figure 4. The Relative Effect of Referee Recommendations and Paper Characteristics on Citations and the Probability of Revise and Resubmit, Robustness

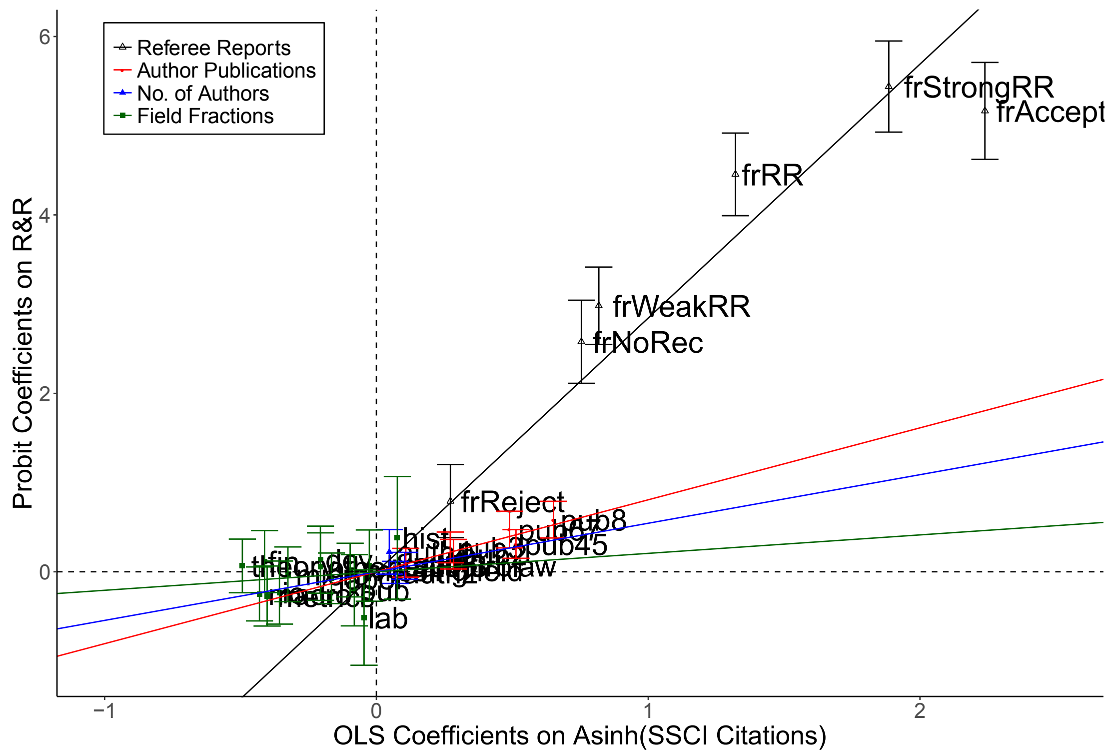
Online Appendix Figure 4a. Using Percentile Citations as Citation measure



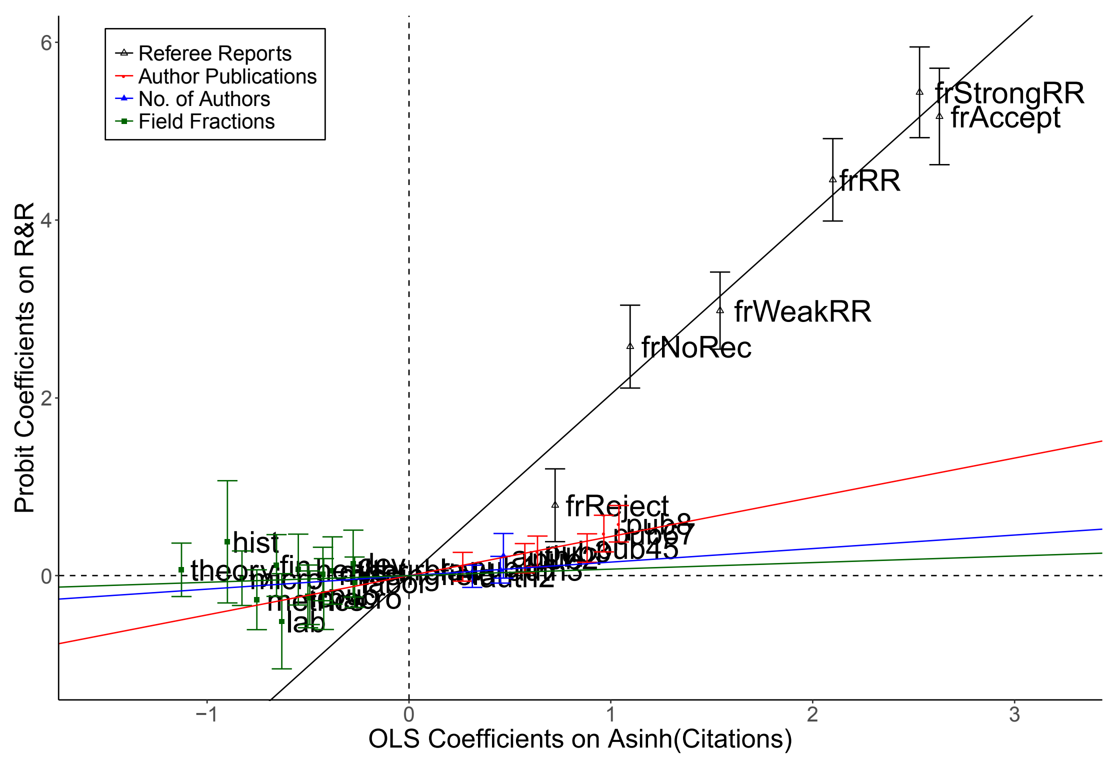
Online Appendix Figure 4b. Using top-cited as Citation measure



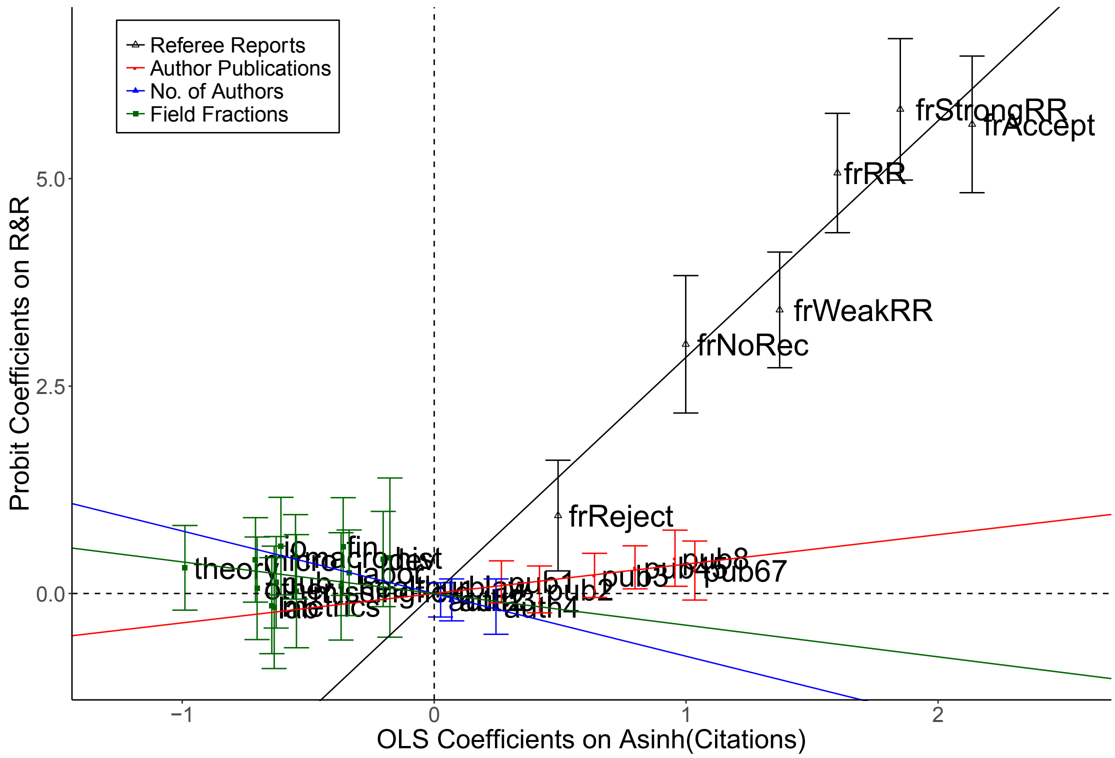
Online Appendix Figure 4c. Using Asinh(SSCI Citations), only years 2006-2010



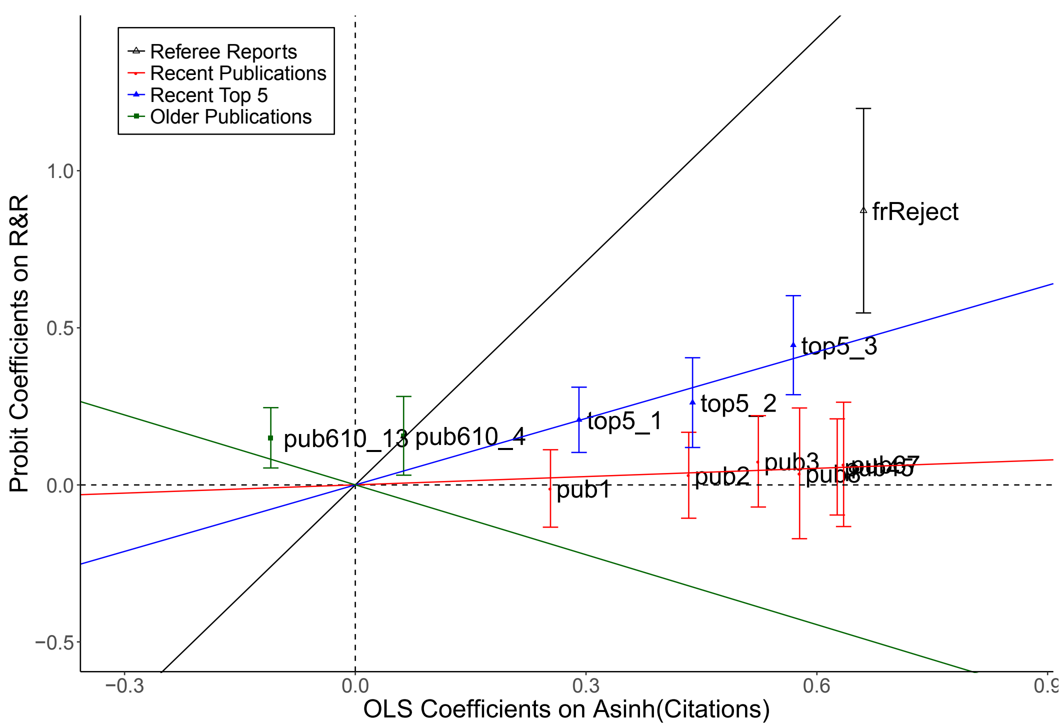
Online Appendix Figure 4d. Using Asinh(Google Scholar Citations), only years 2006-2010



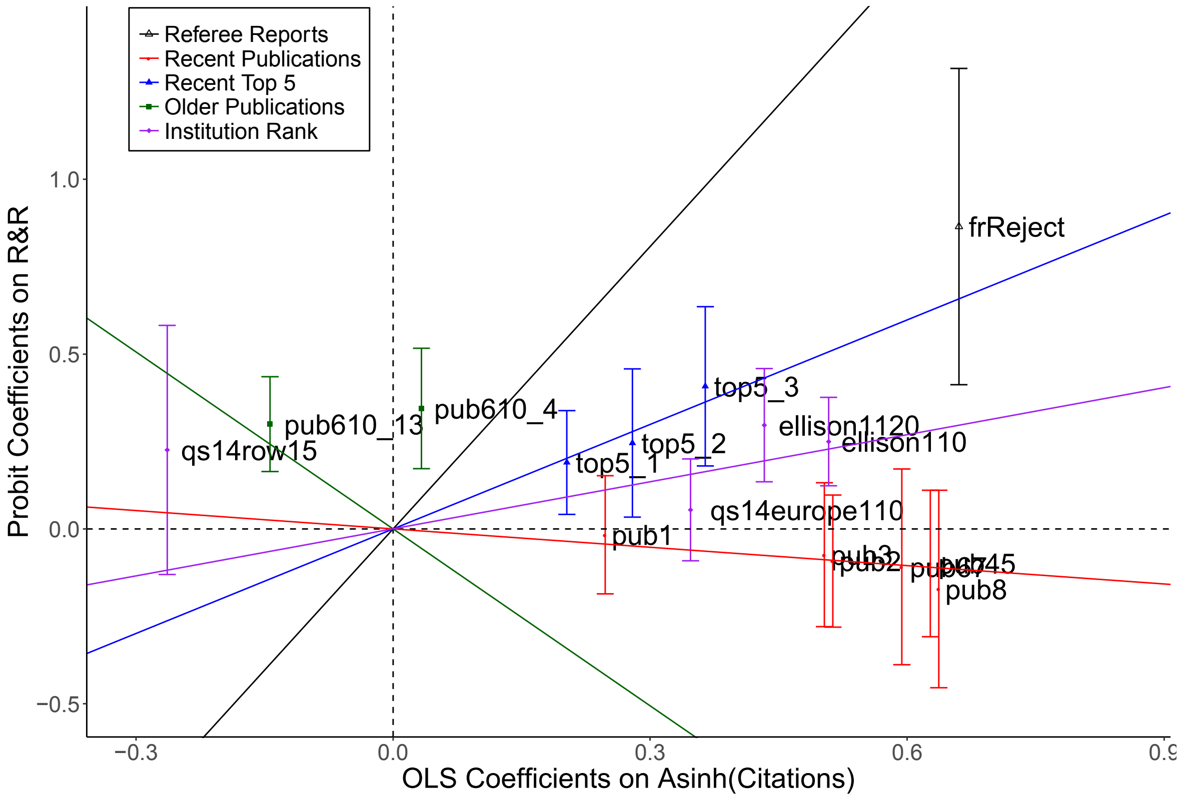
Online Appendix Figure 4e. Using Asinh(Google Scholar Citations), only years 2012-2013



Online Appendix Figure 4f. Including top 5 publications and older publications in regression (not showing coefficients for number of authors and fields)

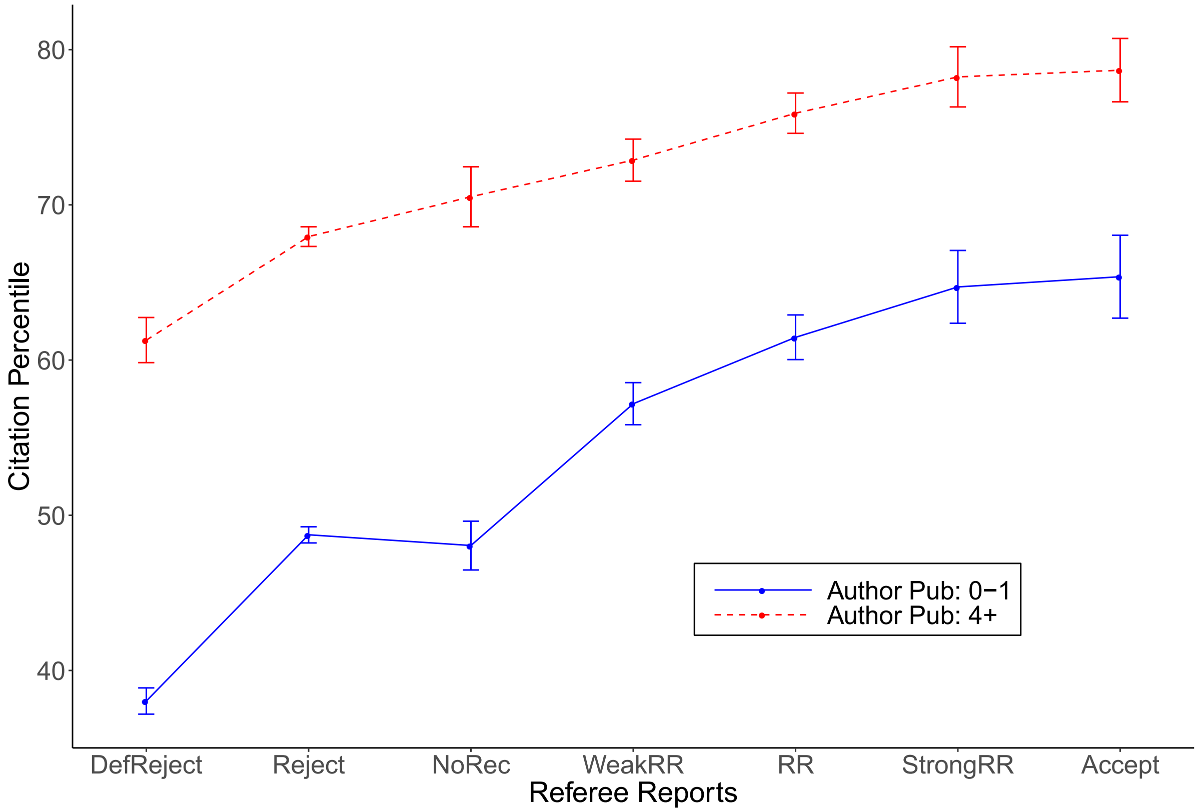


Online Appendix Figure 4g. Including top 5 publications, older publications, and institution rank in regression (JEEA and REStud papers only, not showing coefficients for number of authors and fields)



Notes: Online Appendix Figure 4a-g present a number of robustness checks for the patterns displayed in the coefficient plots of main figures 5a-b in the text. Instead of using the Asinh(Citations) as the dependent variable in the citation regression, panels a, b and c use citation percentile, top-cited and Arsinh(SSCI Citations) respectively. Panels d and e consider papers that were submitted during different periods of the sample – respectively, in 2006-2010 and in 2012-2013. Panels f and g explore the effect of including other measure of author prominence in the regressions – namely recent publications in the top 5 economic journals, older publications (6 to 10 before submission) and the rank of the author's institution. We only collected information on institutional rank for JEEA and REStud, hence the separate plot in panel g incorporating this measure. Across these robustness plots, the key pattern where the positive slope for the referee reports is steeper than for all the other groups variables is stable. In panels f and g, the slope for top 5 publications is steeper than that for other measures of institutional prominence, suggesting that editors' R&R decisions are affected more by authors' recent top 5 publications, than other measure of author prominence (relative to the impact of these variables on citations).

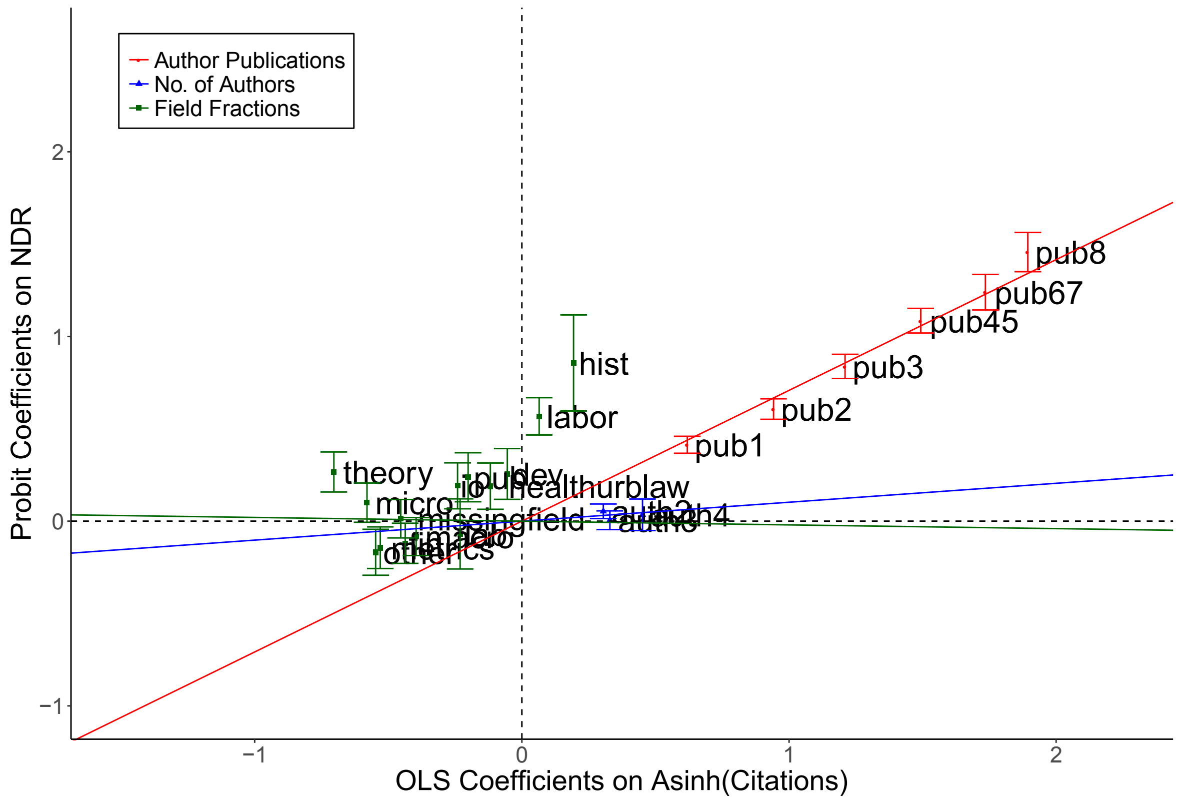
Online Appendix Figure 5. Discounting of Citations of Prolific Authors, Referees (Citation Percentiles)



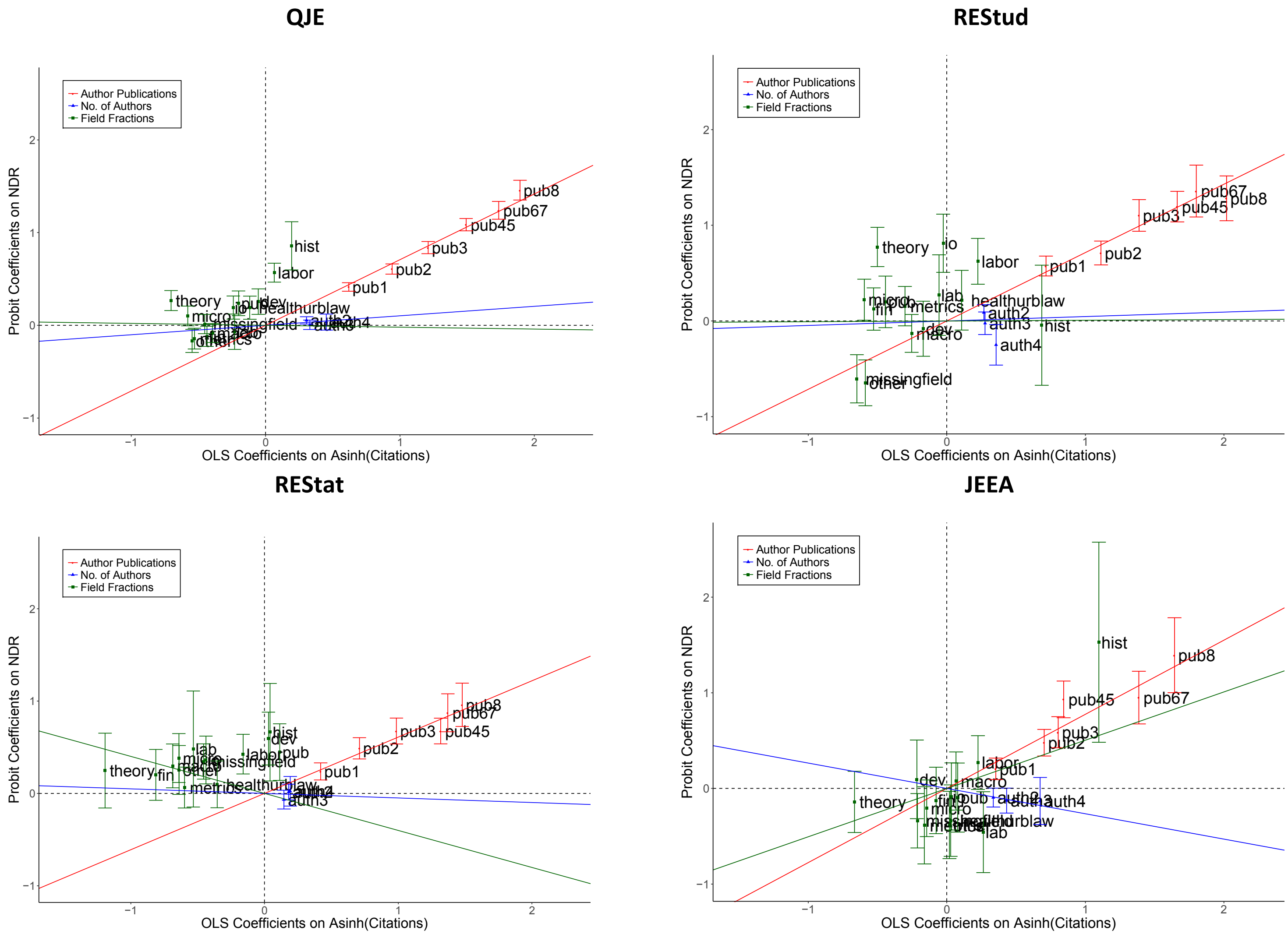
Notes: Online appendix figure 5 checks the robustness of the main figure 9a, using the citation percentile as the measure of citations instead of Asinh(Citations). The key result that conditional on referee recommendation, ex-post citations for prolific authors are higher on average, remains unchanged.

Online Appendix Figure 6. The Relative Effect of Referee Recommendations and Paper Characteristics on Citations and the Probability of Desk Rejection

Online Appendix Figure 6a. Pooled sample



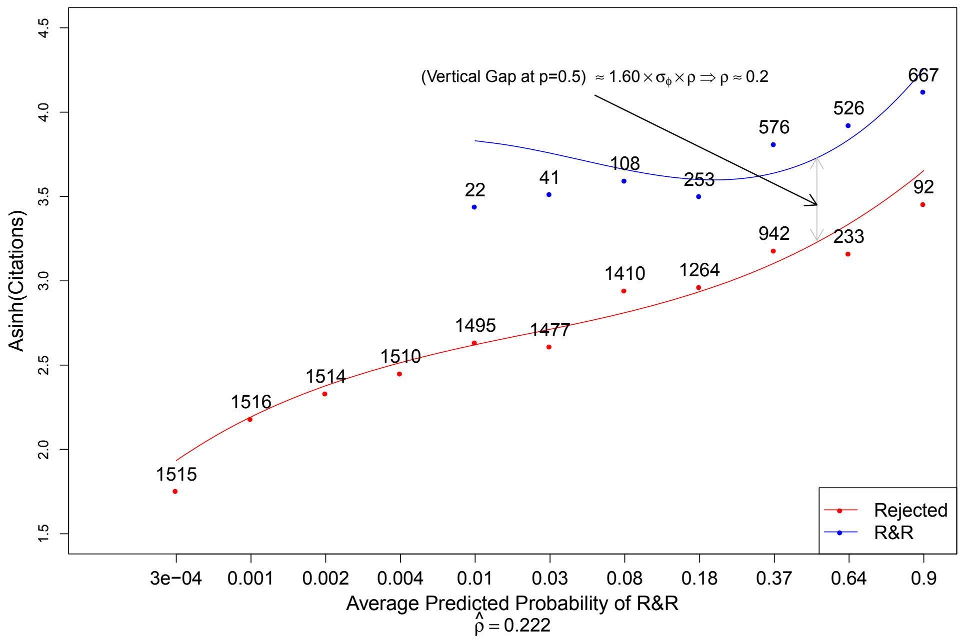
Online Appendix Figure 6b. By Journal



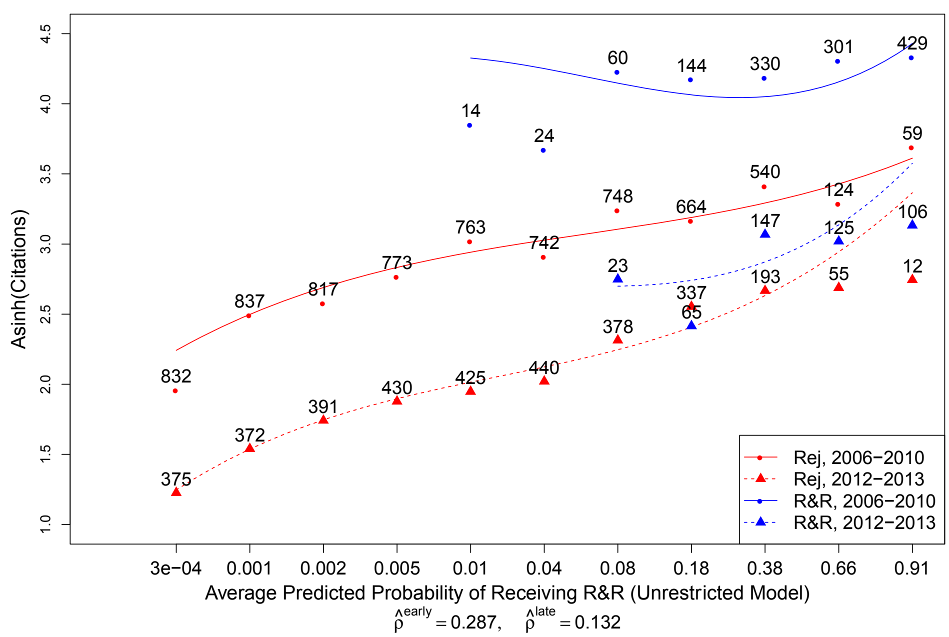
Notes: Online Appendix Figure 6a-b plot the coefficients from the main specifications of the citation and non-desk-rejection regressions (Online Appendix Table 6). Best fit lines through each group of coefficients are also shown (weighted by the inverse variance of the probit coefficient from the non-desk-rejection regression). Panel a shows the results for the pooled sample, whereas panel b shows the results separately by journal. While author publications, number of authors and fields are all predictive of ex-post citations, the editors' decision of whether to desk-reject is influenced more by author publications than by number of authors or fields, relative to the how much these variables predict citations (as the steeper line for author publications indicates).

Online Appendix Figure 7. The Relationship Between the Editor's Revise and Resubmit Decision and Realized Citations – Model Fit

Online Appendix Figure 7a. Lines showing model fit



Online Appendix Figure 7b. Separating bias due to publication, with lines showing model fit



Notes: Online Appendix Figures 7a-b shows the average Asinh(citations) by deciles of predicted probability of R&R where the top decile is further split into two ventiles, and is identical to figure 7 of the main text except for the smoothing lines. The smoothing lines in the online appendix are obtained via cubic fits to the predicted citations from the model (instead of the actual data, as in the main text). The plotted points still reflect actual citations.

Online Appendix Figure 8. Screenshots from an Example of the Survey

Berkeley
UNIVERSITY OF CALIFORNIA

Thank you for participating in our survey! We really appreciate you taking the time to read and evaluate the 4 papers we sent you.

First, we would like your opinion in comparing various features of the two papers:
Paper A: "Do Better Schools Matter? Parental Valuation of Elementary Education" by Sandra E Black (QJE, 1999), and
Paper B: "Using Maimonides' Rule to Estimate the Effect of Class Size on Scholastic Achievement" by Joshua D Angrist and Victor Lavy (QJE, 1999).

	Paper A is better.	Paper A is slightly better.	The two papers are about the same.	Paper B is slightly better.	Paper B is better.
Rigor (theoretical structure and/or research design)	○	○	○	○	○
Exposition (organization, clarity, detail, writing)	○	○	○	○	○
Importance of Contribution	○	○	○	○	○
Novelty	○	○	○	○	○

According to Google Scholar citations collected August 2016, Black, 1999 received 1295 citations and Angrist and Lavy, 1999 received 1667 citations.
In light of the 1295 citations accrued by Black, 1999 and your assessment above, please indicate whether you think that the number of citations for Angrist and Lavy, 1999 is

about right.

too high.

too low.

In light of the 1295 citations accrued by Black, 1999 and your assessment above, what do you think the appropriate number of citations for Angrist and Lavy, 1999 should be?

In light of your assessment of Black, 1999, please indicate whether you think that the number of citations for Black, 1999 (1295) is

about right.

too high.

too low.

Next, we would like your opinion in comparing various features of the two papers:
Paper C: "Schooling and Labor Market Consequences of School Construction in Indonesia: Evidence from an Unusual Policy Experiment" by Esther Duflo (AER, 2001), and
Paper D: "The Dynamics of Educational Attainment for Black, Hispanic, and White Males" by Stephen V Cameron and James J Heckman (JPE, 2001).

	Paper C is better.	Paper C is slightly better.	The two papers are about the same.	Paper D is slightly better.	Paper D is better.
Rigor (theoretical structure and/or research design)	○	○	○	○	○
Importance of Contribution	○	○	○	○	○

Novelty	○	○	○	○	○
Exposition (organization, clarity, detail, writing)	○	○	○	○	○

According to Google Scholar citations collected August 2016, Duflo, 2001 received 1269 citations and Cameron and Heckman, 2001 received 833 citations.
In light of the 1269 citations accrued by Duflo, 2001 and your assessment above, please indicate whether you think that the number of citations for Cameron and Heckman, 2001 is

- about right.
- too high.
- too low.

In light of the 1269 citations accrued by Duflo, 2001 and your assessment above, what do you think the appropriate number of citations for Cameron and Heckman, 2001 should be?

In light of your assessment of Duflo, 2001, please indicate whether you think that the number of citations for Duflo, 2001 (1269) is

- about right.
- too high.
- too low.

Notes: Online Appendix Figure 8 reproduces the survey, which was administered with the Qualtrics platform and displayed as one page. Omitted is only the final question on feedback.

Online Appendix Table 1. List of Journals Used for Prominence Measures

List of Journals	
American Economic Journal: Applied Economics	Journal of Economic Theory
American Economic Journal: Macroeconomics	Journal of Finance
American Economic Journal: Microeconomics	Journal of Financial Economics
American Economic Journal: Economic Policy	Journal of Health Economics
American Economic Review	Journal of Industrial Organization
Brookings Papers on Economic Policy	Journal of International Economics
Econometrica	Journal of Labor Economics
Economic Journal	Journal of Monetary Economics
Experimental Economics	Journal of Money, Credit and Banking
Games and Economic Behavior	Journal of Political Economy
International Economic Review	Journal of Public Economics
Journal of the European Economic Association	Journal of Urban Economics
Journal of Accounting and Economics	Quarterly Journal of Economics
Journal of American Statistical Association	The RAND Journal of Economics
Journal of Business and Economic Statistics	Review of Economics and Statistics
Journal of Development Economics	Review of Financial Studies
Journal of Econometrics	Review of Economic Studies
Journal of Economic Growth	

Online Appendix Table 2. Role of Fields for Citations and R&R Decision

	OLS Models for Asinh of Google Scholar Citations		Probit Models for Receiving Revise-and-Resubmit Decision	
	(1)	(2)	(3)	(4)
Fraction of All Fields Matched (Omitted Category: Theory)				
International	1.07	1.01	-0.03	-0.09
	(0.09)	(0.09)	(0.09)	(0.12)
Lab/Experiments	0.91	0.42	-0.19	-0.48
	(0.15)	(0.14)	(0.15)	(0.20)
Labor	0.64	0.77	-0.18	-0.06
	(0.08)	(0.07)	(0.08)	(0.10)
Health, Urban, Law	0.58	0.59	-0.19	-0.11
	(0.10)	(0.10)	(0.10)	(0.14)
Development	0.71	0.76	0.04	0.17
	(0.11)	(0.11)	(0.10)	(0.13)
History	0.49	0.52	0.53	0.39
	(0.22)	(0.20)	(0.18)	(0.24)
Public	0.54	0.56	-0.04	-0.10
	(0.11)	(0.10)	(0.10)	(0.14)
Industrial Organization	0.46	0.53	-0.12	-0.01
	(0.10)	(0.10)	(0.10)	(0.13)
Finance	0.51	0.52	-0.06	0.13
	(0.10)	(0.09)	(0.09)	(0.12)
Macro	0.59	0.56	-0.02	-0.10
	(0.08)	(0.08)	(0.08)	(0.10)
Field Missing	0.59	0.58	-0.10	-0.17
	(0.09)	(0.08)	(0.08)	(0.11)
Micro	0.30	0.23	0.00	0.06
	(0.09)	(0.08)	(0.09)	(0.11)
Unclassified	0.45	0.55	-0.25	-0.11
	(0.11)	(0.10)	(0.11)	(0.14)
Econometrics	0.45	0.40	-0.16	-0.34
	(0.09)	(0.09)	(0.09)	(0.12)
Controls for Referee Reports	No	Yes	No	Yes
Controls for Author Pubs.	No	Yes	No	Yes
Indicators for Journal-Year	No	Yes	No	Yes
R^2 / pseudo R^2	0.11	0.261	0.04	0.49

Notes: See notes to Tables 1 and 2. The sample for this table includes 15,177 non-desk-rejected papers with at least two referees assigned. Dependent variable for OLS models in columns 1-2 is asinh of Google Scholar citations. Dependent variable in probit models in columns 3-4 is indicator for receiving revise and resubmit decision. Robust standard errors in parentheses.

Online Appendix Table 3. Predictors of Citations and of Revise-and-Resubmit Decision, By Journal

Specification:	OLS				Probit			
Dependent Variable:	Asinh of Google Scholar Citations				Indicator for Revise-and-Resubmit Decision			
Sample:	QJE	REStud	REStat	JEEA	QJE	REStud	REStat	JEEA
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Fractions of Referee Recommendations								
Reject	0.59	0.83	0.75	0.39	0.96	0.74	0.8	1.36
	(0.10)	(0.09)	(0.15)	(0.13)	(0.34)	(0.27)	(0.32)	(0.49)
No Recommendation	1.25	1.05	1.13	0.69	3.34	2.58	2.71	3.32
	(0.26)	(0.17)	(0.25)	(0.19)	(0.39)	(0.32)	(0.35)	(0.52)
Weak R&R	1.61	1.72	1.29	1.18	2.84	3.66	2.84	3.64
	(0.16)	(0.13)	(0.18)	(0.18)	(0.35)	(0.30)	(0.32)	(0.52)
R&R	2.01	2.16	1.86	1.66	4.42	5.35	4.17	5.02
	(0.17)	(0.14)	(0.17)	(0.18)	(0.36)	(0.32)	(0.33)	(0.52)
Strong R&R	2.67	2.61	2.07	1.94	5.07	6.46	5.22	6.31
	(0.25)	(0.21)	(0.21)	(0.25)	(0.41)	(0.36)	(0.38)	(0.57)
Accept	2.48	2.29	2.23	2.34	5.34	5.58	4.8	6.42
	(0.27)	(0.23)	(0.21)	(0.27)	(0.42)	(0.38)	(0.38)	(0.57)
Author Publications in 35 high-impact journals								
Publications: 1	0.37	0.3	0.25	0.17	0.14	0.11	-0.07	0.01
	(0.08)	(0.07)	(0.09)	(0.09)	(0.14)	(0.11)	(0.12)	(0.13)
Publications: 2	0.42	0.64	0.49	0.44	0.32	-0.01	0.17	0.33
	(0.09)	(0.08)	(0.11)	(0.09)	(0.15)	(0.12)	(0.12)	(0.12)
Publications: 3	0.73	0.65	0.68	0.51	0.49	0.14	0.25	0.34
	(0.10)	(0.08)	(0.12)	(0.10)	(0.15)	(0.12)	(0.14)	(0.14)
Publications: 4-5	0.9	0.95	0.85	0.55	0.55	0.09	0.25	0.62
	(0.09)	(0.08)	(0.12)	(0.11)	(0.14)	(0.11)	(0.13)	(0.13)
Publications: 6-7	1.05	0.96	0.93	0.93	0.58	0.21	0.41	0.78
	(0.11)	(0.10)	(0.16)	(0.14)	(0.16)	(0.14)	(0.19)	(0.16)
Publications: 8+	1.02	1.17	0.88	1	0.76	0.28	0.10	0.86
	(0.10)	(0.10)	(0.17)	(0.16)	(0.15)	(0.13)	(0.19)	(0.17)
Number of Authors								
2 authors	0.14	0.23	0.17	0.33	-0.26	0.06	0.08	-0.09
	(0.07)	(0.06)	(0.08)	(0.07)	(0.10)	(0.08)	(0.10)	(0.09)
3 authors	0.21	0.31	0.28	0.42	-0.24	0.11	0.15	-0.21
	(0.08)	(0.07)	(0.10)	(0.09)	(0.11)	(0.10)	(0.12)	(0.13)
4+ authors	0.36	0.55	0.23	0.61	-0.10	0.05	0.38	-0.17
	(0.11)	(0.13)	(0.15)	(0.16)	(0.15)	(0.17)	(0.18)	(0.23)
Indicators for Field of Paper	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Indicators for Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Number of Observations	4,195	5,311	2,391	3,280	4,195	5,311	2,391	3,280
R-squared	0.27	0.24	0.26	0.22				
Pseudo-R^2					0.46	0.52	0.48	0.51

Notes: The sample for each journal includes all non-desk-rejected papers with at least two referees assigned. Robust standard errors in parentheses.

Online Appendix Table 4. Predictors of Citations and of R&R, By Number of Reports

Specification:	OLS			Probit		
Dependent Variable:	Asinh of Citations			Indicator for Revise-and-Resubmit Decision		
No. of Reports Received:	2 Reports	3 Reports	4+ Reports	2 Reports	3 Reports	4+ Reports
	(1)	(2)	(3)	(5)	(6)	(7)
Fractions of Referee Recommendations						
Reject	0.66	0.69	0.51	0.53	1.17	1.6
	(0.07)	(0.11)	(0.31)	(0.23)	(0.26)	(0.47)
No Recommendation	0.83	1.01	1.66	2.38	3.24	4.13
	(0.13)	(0.20)	(0.45)	(0.26)	(0.31)	(0.55)
Weak R&R	1.42	1.45	1.19	2.85	3.63	4.25
	(0.11)	(0.14)	(0.36)	(0.25)	(0.27)	(0.52)
R&R	1.7	1.93	1.87	4	5.42	5.79
	(0.11)	(0.14)	(0.33)	(0.25)	(0.29)	(0.51)
Strong R&R	2.17	2.22	2.59	5.16	6.18	6.46
	(0.16)	(0.19)	(0.40)	(0.29)	(0.32)	(0.58)
Accept	2.32	2.11	2.5	4.96	5.9	6.69
	(0.17)	(0.21)	(0.43)	(0.29)	(0.32)	(0.60)
Author Publications in 35 high-impact journals						
Publications: 1	0.31	0.23	0.15	0.03	0.04	-0.02
	(0.06)	(0.07)	(0.14)	(0.10)	(0.09)	(0.16)
Publications: 2	0.52	0.45	0.36	0.17	0.16	-0.08
	(0.06)	(0.08)	(0.15)	(0.10)	(0.10)	(0.17)
Publications: 3	0.62	0.59	0.62	0.32	0.11	0.27
	(0.07)	(0.08)	(0.14)	(0.10)	(0.10)	(0.16)
Publications: 4-5	0.9	0.71	0.79	0.43	0.19	0.32
	(0.07)	(0.07)	(0.14)	(0.10)	(0.10)	(0.15)
Publications: 6-7	0.97	0.92	0.79	0.73	0.21	0.20
	(0.09)	(0.09)	(0.16)	(0.13)	(0.12)	(0.18)
Publications: 8+	1.03	0.99	0.82	0.46	0.35	0.41
	(0.10)	(0.09)	(0.15)	(0.14)	(0.12)	(0.17)
Number of Authors						
2 authors	0.31	0.15	0.03	-0.12	-0.07	0.09
	(0.05)	(0.06)	(0.12)	(0.07)	(0.07)	(0.13)
3 authors	0.41	0.21	0.26	-0.07	0.05	0.10
	(0.06)	(0.07)	(0.13)	(0.09)	(0.09)	(0.15)
4+ authors	0.42	0.45	0.4	0.18	-0.17	0.30
	(0.10)	(0.11)	(0.16)	(0.15)	(0.15)	(0.19)
Indicators for Field of Paper	Yes	Yes	Yes	Yes	Yes	Yes
Indicators for Year	Yes	Yes	Yes	Yes	Yes	Yes
Number of Observations	7,238	5,414	1,800	7,238	5,414	1,795
R-squared	0.24	0.27	0.33			
Pseudo-R^2				0.57	0.48	0.37

Notes: Robust standard errors in parentheses.

Online App. Table 5. Predictors of Citations and R&R Decision, 2006-10 vs. 2012-

Specification:	**OLS**		**Probit**	
Dependent Variable:	**Asinh of Citations**		**Indicator for Revise-and-Resubmit Decision**	
Sample:	**Papers Submitted from 2006-2010**	**Papers Submitted from 2012-2013**	**Papers Submitted from 2006-2010**	**Papers Submitted from 2012-2013**
	(1)	(2)	(3)	(4)
Fractions of Referee Recommendations				
Reject	0.72	0.49	0.79	0.94
	(0.08)	(0.09)	(0.20)	(0.32)
No Recommendation	1.1	1	2.58	3
	(0.14)	(0.25)	(0.23)	(0.41)
Weak R&R	1.54	1.37	2.98	3.42
	(0.11)	(0.14)	(0.22)	(0.34)
R&R	2.1	1.6	4.45	5.07
	(0.11)	(0.14)	(0.23)	(0.35)
Strong R&R	2.53	1.85	5.44	5.83
	(0.15)	(0.22)	(0.25)	(0.41)
Accept	2.63	2.13	5.17	5.65
	(0.15)	(0.23)	(0.26)	(0.41)
Author Publications in 35 high-impact journals				
Publications: 1	0.27	0.27	0.10	0.14
	(0.06)	(0.07)	(0.08)	(0.13)
Publications: 2	0.57	0.42	0.2	0.05
	(0.06)	(0.08)	(0.08)	(0.14)
Publications: 3	0.64	0.64	0.27	0.22
	(0.07)	(0.08)	(0.09)	(0.13)
Publications: 4-5	0.88	0.8	0.31	0.32
	(0.06)	(0.08)	(0.08)	(0.13)
Publications: 6-7	0.97	1.03	0.47	0.28
	(0.08)	(0.11)	(0.10)	(0.17)
Publications: 8+	1.04	0.96	0.58	0.43
	(0.09)	(0.10)	(0.10)	(0.17)
Number of Authors				
2 authors	0.31	0.03	-0.01	-0.08
	(0.05)	(0.06)	(0.06)	(0.10)
3 authors	0.45	0.07	0.07	-0.08
	(0.06)	(0.07)	(0.07)	(0.12)
4+ authors	0.47	0.25	0.22	-0.16
	(0.10)	(0.10)	(0.12)	(0.16)
F.e. for Field of Paper	Yes	Yes	Yes	Yes
F.e. for Journal-Year	Yes	Yes	Yes	Yes
Number of Observations	8,208	3,893	8,208	3,893
R-squared	0.24	0.20		
Pseudo-R^2			0.50	0.49

Notes: The sample includes all non-desk-rejected papers with at least two reports submitted separately between 2006-2010 (columns 1 and 3), and 2012-2013 (columns 2 and 4). Columns 1 and 3 compare the citation model for papers submitted during the earlier and later time periods, while columns 2 and 4 make the same comparison for the model of editors' R&R decisions. Robust standard errors in parentheses.

Online Appendix Table 6. Average Editor Value Added

Specification:	OLS		
Dependent Variable:	**Asinh of Citations**		
Sample:	**All Papers**	**Papers Submitted from 2006-2010**	**Papers Submitted from 2012-2013**
	(1)	(2)	(3)
Indicator for R&R	0.65	0.87	0.30
(Average Editor Value Added)	(0.06)	(0.07)	(0.10)
1st Decile of Pr(R&R)	1.76	1.96	1.22
	(0.04)	(0.06)	(0.08)
2nd Decile of Pr(R&R)	2.17	2.48	1.54
	(0.04)	(0.06)	(0.08)
3rd Decile of Pr(R&R)	2.34	2.58	1.74
	(0.04)	(0.06)	(0.07)
4th Decile of Pr(R&R)	2.45	2.76	1.89
	(0.04)	(0.06)	(0.07)
5th Decile of Pr(R&R)	2.63	3.02	1.94
	(0.04)	(0.06)	(0.07)
6th Decile of Pr(R&R)	2.62	2.92	2.02
	(0.04)	(0.06)	(0.07)
7th Decile of Pr(R&R)	2.94	3.25	2.33
	(0.04)	(0.06)	(0.07)
8th Decile of Pr(R&R)	2.94	3.2	2.48
	(0.05)	(0.06)	(0.08)
9th Decile of Pr(R&R)	3.17	3.37	2.71
	(0.05)	(0.07)	(0.09)
90-95th Percentile of Pr(R&R)	3.24	3.4	2.71
	(0.07)	(0.10)	(0.13)
95-100th Percentile of Pr(R&R)	3.47	3.49	2.82
	(0.08)	(0.10)	(0.16)
Observations	15,177	8,208	3,893
R-squared	0.72	0.76	0.68

Notes: The sample for column (1) includes all 15,177 non-desk-rejected papers with at least two referees assigned, whereas columns 2 and 3 only use subsamples of these papers submitted from 2006-2010 and 2012-2013 respectively. The probability of R&R which is used to define the dependent variables in these regressions are calculated from the fitted values of the probit specification in column (6) of table 2 (fit to all papers for column (1), and only to the 8,208+3,893=12,101 papers submitted between 2006-2010 or 2012-2013 for columns (2) and (3). Robust standard errors in parentheses. These regressions correspond approximately to figures 7a and 7b in the paper, and the coefficient on the indicator for R&R represents the average value of the gap between the lines for reject and R&R.

Online Appendix Table 7. Predictors of Citations and Desk Rejection

Specification:	**OLS Regression**	**Probit**
Dependent Variable:	**Asinh of Citations**	**Indicator for Paper Not Desk Rejected**
	(1)	(2)
Author Publications in 35 high-impact journals		
Publications: 1	0.62	0.41
	(0.03)	(0.02)
Publications: 2	0.94	0.61
	(0.03)	(0.03)
Publications: 3	1.21	0.84
	(0.04)	(0.03)
Publications: 4-5	1.49	1.09
	(0.04)	(0.03)
Publications: 6-7	1.74	1.24
	(0.05)	(0.05)
Publications: 8+	1.89	1.46
	(0.05)	(0.05)
Number of Authors		
2 authors	0.3	0.05
	(0.02)	(0.02)
3 authors	0.33	0.01
	(0.03)	(0.03)
4+ authors	0.45	0.03
	(0.05)	(0.04)
Indicators for Field of Paper	Yes	Yes
Indicators for Journal-Year Cohort	Yes	Yes
Number of Observations	29,868	29,868
R-squared	0.23	
Pseudo-R^2		0.21

Notes: This table reports the result of regressions on all papers in our sample. Each regression also includes fixed effects for each journal-year cohort. Robust standard errors in parentheses.

Online Appendix Table 8. Excluding Papers with Missing Google Scholar Citations

Specification:	**OLS Regression**	**OLS Regression**	**Probit**	**Probit**
Dependent Variable:	**Asinh of Citations**		**Indicator for Paper Not Desk Rejected**	**Indicator for Paper Receiving a R&R**
Sample	**All Papers**	**Papers that were not desk-rejected**	**All Papers**	**Papers that were not desk-rejected**
	(1)	(2)	(3)	(4)
Fractions of Referee Recommendations				
Reject		0.56		0.79
		(0.06)		(0.16)
No Recommendation		1.03		2.7
		(0.10)		(0.18)
Weak R&R		1.34		3.07
		(0.08)		(0.17)
R&R		1.74		4.52
		(0.07)		(0.18)
Strong R&R		2.15		5.46
		(0.10)		(0.20)
Accept		2.07		5.23
		(0.11)		(0.21)
Author Publications in 35 high-impact journals				
Publications: 1	0.44	0.21	0.32	0.04
	(0.03)	(0.04)	(0.04)	(0.06)
Publications: 2	0.72	0.42	0.46	0.10
	(0.03)	(0.04)	(0.04)	(0.07)
Publications: 3	0.93	0.53	0.72	0.24
	(0.04)	(0.05)	(0.04)	(0.07)
Publications: 4-5	1.16	0.69	0.88	0.3
	(0.04)	(0.04)	(0.04)	(0.06)
Publications: 6-7	1.44	0.88	1.01	0.38
	(0.05)	(0.05)	(0.05)	(0.08)
Publications: 8+	1.61	0.99	1.17	0.44
	(0.05)	(0.05)	(0.05)	(0.08)
Number of Authors				
2 authors	0.19	0.16	-0.12	-0.04
	(0.02)	(0.03)	(0.03)	(0.05)
3 authors	0.26	0.25	-0.19	-0.01
	(0.03)	(0.04)	(0.04)	(0.06)
4+ authors	0.42	0.4	-0.01	0.11
	(0.05)	(0.06)	(0.06)	(0.09)
Indicators for Field of Paper	Yes	Yes	Yes	Yes
Indicators for Journal-Year Cohort	Yes	Yes	Yes	Yes
Number of Observations	24,012	13,581	24,012	13,581
R-squared	0.23	0.29		
Pseudo-R^2			0.11	0.48

Notes: This table reports the results of the main regressions in the paper, excluding observations for which Google Scholar citations were missing. In the main specifications, these observations were retained and assigned zero citatons.