What do Editors Maximize? Evidence from Four Economics Journals*

David Card
UC Berkeley and NBER

Stefano DellaVigna
UC Berkeley and NBER

August 2017

Abstract

We study editorial decision-making using anonymized submissions at four leading economics journals: the *Journal of the European Economic 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, systematic gaps between citations and quality, and a direct impact of publication on citations. Empirically, we find that referee recommendations are strong predictors of citations, and that editors follow the recommendations quite closely. We also show that papers by highly-published authors get more citations, conditional on the referees' recommendations and the publication status, suggesting either that referees impose a higher bar for these authors, or that prolific authors are over-cited. In their revise-and-resubmit decisions, editors only partially offset the referees' opinions, effectively discounting the citations of more prolific authors by up to 80%. We find similar patterns for the desk-reject decision. To attempt to disentangle the two explanations for this discounting, we conduct a survey of specialists, asking them to assess the relative quality of matched pairs of papers by more and less prolific authors. The respondents' assessments closely match relative citations and rule out large biases in citations as a measure of relative quality. We conclude that referees and editors at these journals set a higher bar for more prolific authors, implicitly helping younger and less established authors.

*We thank Daron Acemoglu, Pierre Azoulay, Esther Duflo, Glenn Ellison, Joey Engelberg, Patricia Funk, Joshua Gans, Matthew Gentzkow, Campbell Harvey, David Hirshleifer, Nagore Iriberri, Lawrence Katz, Chris Parsons, Imran Rasul, Laszlo Sandor, Jesse Shapiro, Scott Stern, Vera te Velde, Ivo Welch, and Fabrizio Zilibotti for their comments and suggestions. We are also grateful to the audiences at Columbia University, MIT, UC Berkeley, UC Santa Cruz, the Annual Meeting of WEAI and at the NBER Summer Institute on Innovation for useful comments. We thank Alden Cheng, Jared Grogan, Johannes Hermle, Kaushik Krishnan, Patricia Sun, and Brian Wheaton for outstanding 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*, and to those who responded to our survey. The survey analyzed in this paper was approved by UC Berkeley IRB, protocol 2016-08-9029 and pre-registered as trial AEARCTR-0001669.

1 Introduction

Editorial decisions at top academic journals help shape the careers of young researchers and 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 or reject it? Do referees and editors set the same standard for established scholars as for younger or less prolific authors?

We address these questions using anonymized data on nearly 30,000 recent submissions to the *Quarterly Journal of Economics*, the *Review of Economic Studies*, the *Journal of the European Economic Association*, and the *Review of Economics and Statistics*. Our data set includes information on the field(s) of each paper, the recent publication records of the authors, whether the paper was desk rejected or sent to referees, summary recommendations of the referees, and the editor's reject or revise-and-resubmit decision. All submissions, regardless of the editor's decision or ultimate publication status, are matched to citations from Google Scholar and the Social Science Citation Index.

These unique data allow us to significantly advance our understanding of the editorial decision process at scientific journals. Most previous research has focused on published papers or aggregated submissions (e.g., Laband and Piette, 1994; Ellison, 2002; Hofmeister and Krapf, 2011; Card and DellaVigna, 2013; Brogaard, Engelberg and Parsons, 2014). While such studies offer many 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 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 the referees' recommendations, the characteristics of a paper and its authors, and their own private information to select which papers to invite for revision. As a starting point we assume that editors maximize the expected quality of published papers and that quality is revealed by citations – i.e., citation-maximizing behavior. While highly restrictive, these assumptions provide a useful benchmark. On one hand, *impact factors,* which count citations to the articles in a journal, are salient to publishers, editors and authors.[1] On the other hand, citations are also important determinants of career advancement (Ellison, 2012) and salaries (Hamermesh, Johnson and Weisbrod, 1982; Hilmer, Ransom and Hilmer, 2015).

Nevertheless, there are at least three major limitations of this benchmark. First, the mere publication of a paper in a prestigious journal may raise its citations, introducing a bias in citations

[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.

as measure of quality. Second, editors may favor certain fields or types of authors, resulting in higher and lower quality thresholds for different papers.[2] Third, even in absence of editorial preferences, citations may be systematically biased as a measure of quality by a tendency to cite well-known authors (Merton, 1968) or by differences in citing practices across fields.

We incorporate all three features in our modeling framework and econometric specifications. First, we allow for a direct impact of the editor's revise and resubmit (R&R) decision on ultimate citations. Using differences in R&R rates across editors at the same journal (analogous to the *judges design* used in many recent studies) we develop an instrumental variables (IV) procedure that allows us to separate the mechanical effect of R&R status on citations from the signal contained in the editor's decision. We also develop, under weaker assumptions, bounds for the impact of this mechanical effect on our key results.

Second, we allow referees and editors to hold preferences for or against certain types of papers, relaxing the "citation maximizing" objective. Finally, we allow for the possibility that papers by certain authors (or in certain fields) may receive more citations, holding quality constant. Using only information on citations and editorial decisions we cannot distinguish between editorial preferences for certain types of papers and systematic gaps between citations and quality. Thus, in the final section of the paper we present data from a survey of experts in which we aim to quantify the relative bias in citations versus quality for matched pairs of papers.

We focus our main empirical analysis on the R&R decision for non-desk-rejected papers. These are typically reviewed by 2 to 4 referees who provide summary evaluations ranging from *Definitely Reject* to *Accept*. Our first finding is that referee recommendations are strong predictors of citations: a paper unanimously classified as *Revise and Resubmit* by the referees has on average 240 log points more citations than a paper they unanimously agree is *Definitely Reject*. 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, a second key finding is that the referee recommendations are *not* sufficient statistics for expected citations. In particular, submissions from authors with more publications receive substantially more citations, controlling for referee recommendations. For example, papers by authors with 6 or more recent publications (in a set of general interest and field journals) have on average 100 log points more citations than papers with similar referee ratings by authors with no recent publications. This gap is essentially unchanged when we use our IV framework to measure the effect of R&R status, and is only slightly smaller under the extreme bound that treats R&R status as exogenous. We conclude that either referees are significantly tougher on more prolific authors (i.e., a bias arising from referee preferences) or that submissions from more prolific authors receive many more citations conditional on their quality (i.e., a bias in citations as a measure of quality).

Looking at the R&R decision, our third finding is that editors are heavily influenced by the referees' recommendations: the summary referee recommendations alone explain over 40 percent of

[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. (forthcoming) find strong evidence of favoritism in elections to the Chinese Academies of Engineering and Science.

the variation in the R&R decision.[3] Moreover, the relative weights that editors 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.

While the editors largely follow the referees, our fourth finding is that they also have substantial private information about the future citations of papers they handle, over and above the information contained in the summary referee recommendations (and other observable paper characteristics). In our econometric model, the reliability of this information is revealed by the coefficient on the control function that is included in our citation models to address the endogeneity of the R&R decision. Interpreted in this light, the correlation of the editor's signal with the unobserved determinants of future citations is around 0.20.

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 accept fewer papers from more prolific authors than would be predicted from a citation-maximizing perspective – our fifth and perhaps most surprising finding. In fact, editors at all four journals undo only a quarter or less of the bias against more prolific authors exhibited by referees. Either editors agree with referees in their preference for less prolific authors, or they agree with referees that the papers by more prolific authors get too many citations, conditional on their quality.

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.

Although our main focus is on the R&R decision, we also analyze the desk rejection (DR) decision. Desk rejections are increasingly common in economics – accounting for about 50% of submissions in our sample – yet there is little evidence on how DR decisions are made. Our sixth finding is that editors have substantial private information about paper quality at the DR stage. Conditional on 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 72 log points more citations on average than papers that are desk rejected. Since both groups of papers are ultimately rejected, this comparison bypasses any concern about endogenous publication effects. As at the R&R stage, editors discount the expected citations that will accrue to papers by more prolific authors. Indeed, desk-rejected papers by prolific authors have higher average citations than non-desk-rejected papers by authors with no previous publications.

Our finding that referees *and* editors act as if they under-value citations to papers by more prolific authors runs counter to a long strand of research suggesting that the editorial process is biased in favor of more prominent scholars. Nevertheless, a recent review argues that evidence for such bias is limited (Lee et al., 2013). Moreover, Blank's (1991) analysis of a randomized comparison of blind versus non-blind refereeing at the *American Economic Review* showed that blind refereeing

[3] Blank (1991) and Welch (2014) similarly show that editorial decisions are highly related to the referees' opinions.

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) and Ellison (2011) 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]

To disentangle whether the higher citation bar for papers by prolific authors arises because reviewers and editors think these papers get too many citations or because of "affirmative action" for younger or less accomplished 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 a top five journals) in their field. One paper in each pair was written by author(s) with 4 or more publications in the years prior to an approximate submission date, while the other was written by author(s) with at most one recent publication. We provided respondents with the actual GS citations for each paper and asked them to assess the appropriate relative number of citations based on their judgment of the quality of the papers. We use the responses to infer the relative ratio of citations to quality for more versus less prolific authors, using a pre-registered specification. We emphasize that survey respondents were asked to evaluate the relative quality of papers, *not* to make R&R recommendations. Thus, we hope to abstract from any tendency to raise (or lower) the bar for prolific authors at the refereeing stage.

Our survey respondents do not think that prolific authors are over-cited. Their preferred relative citations for more prolific authors are only 2% below their actual relative citations (standard error = 5%), suggesting that relative citations are approximately proportional to relative quality. In light of this finding, we argue that referees' and editors' systematic discounting of expected citations for papers written by more prolific authors arise because they impose a higher bar for these authors, leaving room in the journals for younger and less established authors.

What implications do our findings have for the editorial process at top economics journals? Overall, publication decisions are largely driven by the referees, whose summary recommendations explain 40% or more of R&R outcomes. Editors also act on some additional private information that is highly correlated with ultimate citations, though we cannot tell whether this information arises from their own reading of the paper or from the detailed referee reports (which we do not see). In either case, however, editors appear to "add value" to the decision process. On average they also partly offset the higher bar imposed by the referees on more prolific authors, lowering the discount on their expected citations from about 100 log points to about 80 points. Editors play a more decisive role at the desk reject stage. Here, we find that their private information is also highly predictive of citations. Again, there is strong evidence of a higher bar for more prolific authors.

Are these results surprising? To provide some evidence, we collect forecasts as in DellaVigna

[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]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. Medoff (2006) finds that papers by authors from Harvard and the 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.

and Pope (forthcoming). In advance of a presentation of this paper we surveyed a group of editors and associate editors at the *Review of Economic Studies* and a set of faculty and graduate students at the University of Zurich. Overall, the forecasts by editors and faculty respondents are quite accurate, more so than those by graduate students. However, editors overestimate their ability to select highly-cited papers at the R&R and DR stages; other faculty *underestimate* their ability to do so. Editors are aware that they set a higher bar for papers by prolific authors at the DR stage, a pattern that other faculty do not anticipate. Editors also underestimate how much they follow the referees. In light of these results, we believe our findings do provide additional insight into the editorial process – even for editors themselves – and lay the groundwork for a deeper understanding of this process, at least in the upper tier journals in economics.

2 Model

To help organize our empirical analysis we begin by developing a stylized model of the editorial decision process. We focus initially on the R&R stage. Then we move to the earlier desk rejection stage, which shares many of the same features apart from the input of referees. For simplicity we ignore any stages after the 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:

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

where for simplicity we treat the unobserved component of quality, ϕ_q, as normally distributed with mean 0 and standard deviation σ_q. Notice that we allow observable paper characteristics to help forecast paper quality conditional on the referee assessments. That is, we do not assume that the referee recommendations efficiently incorporate both the private information extracted by the referees from reading the paper and the publicly available information contained in x_1.[7]

[6] For simplicity 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 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) and is related to referee "matching" (Hamermesh, 1994). 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, so for simplicity we ignore it.

[7] Equation (1) could be motivated by a model in which referees observe the quality of a paper with some error - similar to the model we pose below for editors - and then report to the editor. If referees act as Bayesians and report their best estimate of quality, conditional on their signal and other public information contained in x_1, then we would expect $\beta_1 = 0$. If referees report their signal, however, then we would expect $\beta_1 \neq 0$. We thank Glenn Ellison for this point.

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

$$s = \phi_q + \zeta.$$

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

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

where $A = \sigma_q^2/(\sigma_q^2 + \sigma_\zeta^2)$. This is an optimally shrunk version of the editor's private signal, and is normally distributed with standard deviation $\sigma_v = A^{1/2}\sigma_q$ and correlation $\rho_{vq} = A^{1/2}$ with ϕ_q. 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.[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:

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

where $\tau_1 = 0$ corresponds to the situation where the editor cares only about expected quality. As in a standard random preference model (McFadden, 1973) 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:

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

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

We assume that cumulative citations (c) to a paper, which are observed some time after the editor's decision, reflect a combination of quality and other factors summarized in η :[9]

[8] Assuming that 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 T. If $\log q$ is normally distributed with mean M and variance V conditional on (s,x) then expected quality is $\exp(M+V/2)$, which will exceed a given threshold T if and only if $M \geq \tau_0 \equiv \log T - V/2$. We have found little evidence of heteroskedasticity in the residual from a regression of log citations on measures of x_1 and x_R, though this does not necessarily imply that v is homoskedastic.

[9] As we discuss in Section 5.2, this can be easily generalized to $\log c = \theta(\log q + \eta)$, 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 $\theta = 1$.

$$\log c = \log q + \eta.$$

The simplest assumption is that η depends only on the vintage of the paper: in this case citations form a perfect index of quality apart from an adjustment for the lag between the time the paper was evaluated and the time citations are measured. More generally, however, citations can also depend on factors like the field of a paper and the track record of the author(s) – variables included in the vector x_1 – as well as on the R&R decision made by the editor and other random factors captured in an error component ϕ_η :

$$\eta = \eta_0 + \eta_1 x_1 + \eta_{RR} RR + \phi_\eta. \tag{5}$$

The coefficient η_{RR} captures any mechanical "publication bias". Papers that receive an R&R verdict are likely to be published sooner (and in a higher-ranked journal) than those that are rejected, leading to a boost in citations and implying that $\eta_{RR} > 0$.

Combining equation (5) with equation (1) leads to a model for citations:

$$\begin{aligned} \log c &= \beta_0 + \eta_0 + (\beta_1 + \eta_1) x_1 + \beta_R x_R + \eta_{RR} RR + \phi_q + \phi_\eta \\ &= \lambda_0 + \lambda_1 x_1 + \lambda_R x_R + \lambda_{RR} RR + \phi \end{aligned} \tag{6}$$

where $\lambda_0 = \beta_0 + \eta_0$, $\lambda_1 = \beta_1 + \eta_1$, $\lambda_R = \beta_R$, $\lambda_{RR} = \eta_{RR}$, and $\phi = \phi_q + \phi_\eta$.

Clearly, when η is constant across papers (and thus $\eta_1 = \eta_{RR} = 0$) we can recover β_1 and β_R from a regression of citations on paper characteristics and referee recommendations, and potentially compare these coefficients to those estimated from the R&R probit model. 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. Moreover, OLS estimation of equation (6) poses a potential problem because RR status is endogenous, and will be positively correlated with the error component ϕ to the extent that editors' private signals are informative about quality.

Fortunately, the structure of the editorial process suggests a straightforward approach for recovering consistent estimate of the coefficients λ_1, λ_R and λ_{RR}. Specifically, assume that different editors have different quality thresholds for reaching an R&R decision (i.e., different values of the constant τ_0) but that the particular editor assigned to a paper has no effect on citations. In this case, following the approach used in many recent studies of judicial and administrative decision making (e.g., Maestas, Mullen and Strand, 2013; Dahl, Kostol and Mogstad, 2014; Aizer and Doyle, 2015), we can use the R&R rate for other papers handled by the same editor as a variable that shifts the threshold for R&R but has no independent effect on citations.

Rather than estimate equation (6) by two stage least squares, we use a control function approach (Heckman and Robb, 1985; Wooldridge, 2015; Brinch, Mogstad and Wiswall, forthcoming) which in principle identifies the average treatment effect of R&R status. We first fit a probit model for the R&R decision, including x_1, x_R and the instrumental variable z formed by the leave out mean R&R rate of the specific editor. We then form an estimate of the generalized residual r from the R&R

probit model:

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

where $\phi(\cdot)$ and $\Phi(\cdot)$ are the standard normal density and distribution functions, respectively, and

$$\pi(x,z) = \pi_0 + \pi_1 x_1 + \pi_R x_R + \pi_z z$$

is a linear index function of x and the instrumental variable z. Finally, we include $\widehat{r}$ (the estimate of r) in the citation model:

$$\log c = \lambda_0 + \lambda_1 x_1 + \lambda_R x_R + \lambda_{RR} RR + \lambda_r \widehat{r} + \phi'. \tag{7}$$

Equation (7) is a standard two-step selection-corrected model (Heckman 1976, 1979). The inclusion of the generalized residual from the R&R probit absorbs any endogeneity bias in the R&R decision. Moreover, the estimate of λ_r provides a measure of the correlation $\rho_{v\phi}$ between the editor's private signal (v) and the unobserved determinants of citations (ϕ) since *plim* $\widehat{\lambda}_r = \rho_{v\phi}\sigma_\phi$. In the special case where $\phi_\eta = 0$ (i.e., there is no additional noise in realized citations) $\rho_{v\phi} = \rho_{vq}$, and we can use the estimate of λ_r to estimate the informativeness of the editor's signal. Otherwise, the implied correlation will tend to under-estimate ρ_{vq} because citations contain an extra component of noise.

A possible concern with our procedure for estimating λ_{RR} is that the identity of the editor assigned to a paper affects citations, controlling for journal, field and other characteristics, or that the functional form of the control function we use is incorrect. In this case the estimated coefficients for x_1 and x_R, which are our main focus, will be potentially biased. To address these concerns we re-estimate equation (7) under two alternative procedures which we believe bracket the plausible range of values for the coefficient λ_{RR}. For an upper bound we estimate (7) without including $\widehat{r}$, thereby ignoring the likely endogeneity of RR. For a lower bound we set $\lambda_{RR} = 0$ and estimate (7) without any control for RR status.

Interpreting the Effects of Referee Recommendations and Paper Characteristics In our analysis, we estimate the R&R decision model and the citation model (equations (4) and (7)) and then compare the relative effects of paper characteristics on the probability of an R&R verdict and on citations. As a starting point, consider a benchmark model with 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 quality. If referees tend to give worse (better) recommendations to certain types of papers, controlling for quality, this will lead to a positive (negative) coefficient for the corresponding element of x_1 in equation (7). An estimate of 0.20 for the coefficient of a certain paper characteristic in equation (7), for example, means that the referees are essentially discounting expected citations to papers in this category by 20% in making their recommendations about publishing the paper.[10]

Assuming that the editor is a citation maximizer, however, he or she will take the referees' biases into account and weight papers with this characteristic more positively. 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 1 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 the 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 show, these restrictions are not 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 relative effect of x_1 and x_R on the probability of an R&R versus observed citations. In either case the referee recommendation variables will still affect citations and the R&R decision proportionally, so the coefficients $\widehat{\pi}_R$ and $\widehat{\lambda}_R$ 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. For a characteristic that leads the editor to impose a higher (lower) R&R threshold, the corresponding pair of coefficients $(\widehat{\pi}_{1k}, \widehat{\lambda}_{1k})$ will fall below (above) the reference line plotting the $\widehat{\pi}_R$ coefficients against the $\widehat{\lambda}_R$ coefficients. Similarly, for a paper characteristic that leads to more (less) citations conditional on the paper quality, the corresponding pair of coefficients will fall below (above) the reference line. The two alternative explanations for any non-proportional effects can only be distinguished if we measure the relationship between quality and citations, which our survey of expert readers attempts to uncover.

Regardless of the source of non-proportionality, we can devise a simple metric that summarizes the effective degree of discounting by editors. Specifically, consider the k^{th}element of the vector x_1. Let π_{1k} represent the coefficient of this characteristic in the R&R probit model. Under a citation maximizing benchmark, $\tau_{1k} = \eta_{1k} = 0$ and the coefficient of this characteristic in the log citation model would be:

$$\lambda^*_{1k} = \pi_{1k}\sigma_v \tag{8}$$

We can estimate λ^*_{1k} using the estimate of π_{1k} from the probit model and an estimate of σ_v based on the relative effect of the referee recommendations on citations and the R&R decision. The gap $(\lambda_{1k} - \lambda^*_{1k})$ between the actual coefficient in the citation model and the predicted coefficient with $\tau_{1k} = \eta_{1k} = 0$ represents the extent to which citations to papers with the k^{th} characteristic are over- or undervalued by the editor. Notice that if $\pi_{1k} = 0$ then $\lambda^*_{1k} = 0$: the editor agrees with the referees and applies the same discount factor. But for characteristics with a positive (negative) coefficient in the R&R probit model, the editor is applying a smaller (larger) discount than the referees.

[10]This is most easily seen by moving $\lambda_1 x_1$ to the left hand side of equation (7) so the dependent variable is $\log c - \lambda_1 x_1$, which can be interpreted as "discounted citations".

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

$$\log q = \alpha_0 + \alpha_1 x_1 + \omega_q \tag{9}$$

where ω_q is a normally distributed error component with mean 0 and standard deviation σ_{ω_q}. Based on an initial reading of the paper, the editor observes a signal

$$s_0 = \omega_q + \varepsilon$$

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

$$\begin{aligned} E[\log q|x_1, s_0] &= \alpha_0 + \alpha_1 x_1 + A_0 s_0 \\ \text{where } \ A_0 &= \frac{\sigma^2_{\omega_q}}{\sigma^2_{\omega_q} + \sigma^2_\varepsilon}. \end{aligned}$$

Define $v_0 = A_0 s_0$: this is a normally distributed random variable with mean 0 and standard deviation $\sigma_{v_0} = A_0^{1/2}\sigma_{\omega_q}$ 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

$$E[\log q|x_1, s_0] = \alpha_0 + \alpha_1 x_1 + v_0 \geq \gamma_0 + \gamma_1 x_1$$

which has the same form as the decision rule at the R&R stage.[11] This rule leads to a simple probit model for the probability of non-desk-rejection ($NDR = 1$), conditional on the characteristics x_1:

$$P[NDR = 1|x_1] = \Phi[\frac{\alpha_0 - \gamma_0 + (\alpha_1 - \gamma_1)x_1}{\sigma_{v_0}}]. \tag{10}$$

Next we specify a model for citations, conditional on information available at the desk reject stage. We assume that the gap between citations and quality, conditional on x_1 and NDR status, can be written as:

$$\log c - \log q \quad = \quad \delta_0 + \delta_1 x_1 + \delta_{NDR} NDR + \omega_c. \tag{11}$$

which includes a constant, a random error component, and potential controls for x_1 and NDR status. As we argued at the R&R stage, it is plausible that δ_{NDR} is positive. Non-desk-rejected papers have

[11] 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

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

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 $E[\log q|x_1, s_0] > \gamma_0 + \gamma_1 x_1$.

some chance of being published in the current journal, whereas those that are desk rejected have to be submitted to other outlets. Other things equal, NDR papers are therefore more likely to be published sooner and to be published in a higher quality journal – both factors that could raise citations.

Combining equations (9) and (11) leads to a model for observed citations conditional on x_1:

$$\begin{aligned} \log c &= \alpha_0 + \delta_0 + (\alpha_1 + \delta_1)x_1 + \delta_{NDR} NDR + \omega \\ &= \psi_0 + \psi_1 x_1 + \psi_{NDR} NDR + \omega \end{aligned} \tag{12}$$

where $\psi_0 = \alpha_0 + \delta_0$, $\psi_1 = \alpha_1 + \delta_1$, $\psi_{NDR} = \delta_{NDR}$, and $\omega = \omega_q + \omega_c$. Since desk rejection is determined in part by the editor's signal, we expect that the error term will be positively correlated with NDR. As at the R&R stage, we address this using a control function approach. We first fit a probit model for NDR, including the observable paper characteristics and an instrumental variable z_0 based on the mean NDR rate of the editor on other papers. We then form an estimate of the generalized residual from this probit model, r_0, and estimate a selection-corrected citation model:

$$\log c = \psi_0 + \psi_1 x_1 + \psi_{NDR} NDR + \psi_r \widehat{r}_0 + \omega'.$$

The coefficient ψ_r captures the strength of the correlation between the residual in the NDR probit, which is based on the editor's signal, and the residual $\omega = \omega_q + \omega_c$. This will be larger, the better able is the editor to forecast quality, and the stronger the link between citations and quality (i.e., the smaller is the variance of the noise component ω_c).

Comparisons of Papers with the Same Probability of Desk Rejection At the desk rejection stage we do not have a set of variables, comparable to the referee recommendations, that provide a ready benchmark for gauging the effects of a given characteristic on expected citations and the probability of desk rejection. Nevertheless, it is possible to develop a simple test of citation-maximizing choice behavior based on comparisons of citations for non-desk-rejected papers with the same probability of NDR. Specifically, our model implies that expected citations should be similar for any two papers with the same probability of non-desk-rejection handled by a given editor.[12]

To see why, suppose that $\gamma_1 = \delta_1 = 0$. In this case that the probability of non-desk rejection by a given editor is:

$$p(x_1) = P[NDR = 1|x_1] = \Phi[\frac{\alpha_0 - \gamma_0}{\sigma_{v_0}} + \frac{\alpha_1}{\sigma_{v_0}} x_1]$$

where the editor's NDR threshold is captured by the constant α_0. Expected citations for a paper that is NDR by this editor are :

$$E\left[\log c|x_1, NDR = 1\right] = \alpha_0 + \delta_0 + \alpha_1 x_1 + \psi_{NDR} + \psi_r g(\frac{\alpha_0 - \gamma_0}{\sigma_{v_0}} + \frac{\alpha_1}{\sigma_{v_0}} x_1)$$

where $g(y) = \frac{\phi[y]}{\Phi[y]}$ is the standard "selection correction" term specified in Heckman (1979). Now

[12] This test is similar to the tests widely used in the law and economics literature to test for discrimination by police officers in deciding to stop people of different race groups, e.g., Knowles, Persico and Todd (2001).

consider any two papers that receive an NDR verdict from a given editor with the same value for $p(x_1)$. These papers must have the same value for the covariate index $\alpha_1 x_1$, and thus the same expected citations. The assumption of citation maximizing behavior therefore implies:

$$E\left[\log c | x_1, NDR = 1\right] = G(p(x_1)), \tag{13}$$

where $G(\cdot)$ is a strictly increasing continuous function. Equation (13) leads to a simple test for the citation maximizing hypothesis: we fit a model for the probability of NDR, then classify papers into cells based on their propensity to receive an NDR verdict, and compare average citations for papers handled by a given editor with different values of an individual covariate (such as the author's previous publication record). Under citation maximization, $p(x_1)$ is a sufficient statistic for expected citations among all papers with NDR=1, and there should be no difference in expected citations for papers in a given cell. If, instead, editors are using a different threshold for different authors (i.e., $\gamma_1 \neq 0$) or editors care about quality but citations are a biased measure of quality (i.e., $\delta_1 \neq 0$), then we expect to see differences in expected citations for papers with the same NDR propensity.

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), an (anonymized) identifier for the editor handling the paper[13], citation information from Google Scholar (GS) and the Social Science Citation Index (SSCI), and the editor's decisions regarding desk rejection and R&R status.[14]

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.[15] We constructed our data sets for the *Review of Economics and Statistics* (REStat) and the *Quarterly Journal of Economics* (QJE) in April 2015, and the data set for the *Review of Economic Studies* (REStud) in September 2015.

[13] As per our agreement with the journal, we did not store editor identifiers for REStat. In the analysis with editor fixed effects, we treat REStat as having just one editor. In that analysis, we also pool, within a journal, editors who handled very few papers, as the mean R&R rate would be very imprecisely estimated.

[14] The data set does not include any information on demographic features of the authors or referees, such as age or gender, and does not track authors or referees across papers.

[15] The top-code limit for citations 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.

The data set for the *Journal of the European Economic Association* (JEEA) was constructed over 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 16 months for the accrual of citations before citations are measured. 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. 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 any revision stages that occur after an initial R&R decision.[16]

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".[17] 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".[18]

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.[19] We then take the highest publication record of all co-authors, setting the count to 0 if we 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 significant 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

[16] We 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.

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

[18] 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.

[19] We also calculate the number of publications in these same journals in years 6-10 before submission, and the number of publications in the top 5 economics journals in the 5 years before submission.

QJE tend to come from the most prolific authors, followed by REStud, then REStat and JEEA.

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 GS, 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 SSCI, which are reported in GS 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. Thus, we add the citations accrued in working paper format and in the final publication, as long as the paper title is the same. Papers with no match in Google Scholar are coded as having zero citations.[20]

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.[21] Online 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 Online Appendix Figure 1b).

[20]In Online Appendix Table 7 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.

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

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 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 degree of referee informativeness 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.[22]

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.[23] 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

[22]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.

[23]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.

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 it 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 $\times$ year fixed effects and the fractions of referee reports in each category. As in the figures, the estimates show a strong correlation between referee evaluations and mean citations. The increases in the estimated coefficients between categories are substantially larger than the slopes in Figure 3a, reflecting the fact that the coefficients in the regression model measure the effect of all referees unanimously changing recommendations from one category to another, whereas the figure measures the effect of a single referee changing his or her recommendation.

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 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, when we compare the coefficients of the referee category variables for papers with 2, 3, and at least 4 referees, the coefficients are remarkably 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 a sufficient statistic 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 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, controlling for the referee recommendations. This means that papers by authors with 4-5 recent publications have to receive 2.3 ($=e^{0.85}$) times more citations, on average, than papers by authors with no recent publications to get the same evaluations from the referees. There are similar effects for papers with more co-authors and papers in more-cited fields.

Mechanical Publication Bias

So far, we have neglected the potential for a mechanical publication bias: papers that receive an R&R may accumulate more citations, conditional on quality, because the publication itself increases visibility, or provides a signal. This bias could lead us to overstate the impact of the determinants of citations. For example, positive referee recommendations may be correlated with citations not (only) because referees capture the paper quality, but because positive reports increase the probability that a paper obtains an R&R, which itself increases citations.

As we discussed in Section 2, under the assumptions of the model, we can address this issue with specification (7). We include an indicator for R&R, as well as a control function for the selection into the R&R stage, using as predictor the average R&R rate of an editor (with a leave-one-out mean). (This selection equation, which we discuss below, is reported in Column 9 of Table 2). The coefficient on the R&R indicator indicates the mechanical publication effect (in log points), while the coefficient on the control function provides a measure of the "value added" of the editor.

We display the estimate in Column 4 of Table 2. The estimate on the control function is statistically significant and positive at 0.32 (s.e. 0.08). Given the residual variance of citations ($\sigma_v \approx 1.6$), this implies a correlation of the unobservable determinants of the editor's decision with the unobserved component of citations of around 0.2. The coefficient on the R&R dummy of 0.06 (s.e. 0.14) indicates that the mechanical effect of an R&R is to increase citations by just 6 log points, though we cannot rule out an effect as large as 34 points. In interpreting this estimate we stress that many of the papers receiving an R&R in our sample were not published by the time we collected citations in mid 2015, and that not all journals in our sample would be expected to have a sizable publication effect relative to alternative outlets. We return to these points shortly.

Importantly, under this specification, the coefficients on the other variables—the referee recommendations, prior author publications, and the number of authors — are barely affected compared to a specification without controls for R&R status and the selection effect (Column 3). This suggests that the biases in these coefficients arising from any mechanical publication effects are small.

A reasonable objection is that our baseline specification relies on a set of modeling assumptions, including the implicit assumption that the identity of the editor handling a paper does not matter for citations, only for the probability of R&R. To address this concern we use the bounding approach described above. The specification in column 5 of Table 2 includes a control from R&R status but does not include the control function. In this specification, any higher citations for R&R'd papers are attributed to a mechanical publication effect, with no role for the editor's signal. Under this upper bound, the mechanical publication bias is estimated to be 57 log points. Even under this extreme assumption, however, the coefficients on the other key variables are only modestly affected: the estimated coefficients on the referee recommendations are about 20 percent smaller in magnitude, while the estimated coefficients of the author publication variables are 2-3 percent smaller. Moreover, the *relative* attenuation of the author publication variables is smaller, so any conclusions regarding the under-weighting of the author's publication record relative to the referee recommendations in reaching an R&R decision are actually stronger.

For completeness, Column 6 presents the lower bound, assuming no publication bias but including the control function. The estimate is nearly identical to the estimate in Column 3 without the control function.

In Table 3 we present additional evidence expanding on the benchmark specification (Column 4 of Table 2), reproduced in Column 1 of Table 3. To the extent that there is a mechanical publication effect, we expect it to be larger for papers submitted prior to 2010, since most of these papers should have been published for at least a year by the time we gathered citation information (in 2015) and had time to gain visibility. By comparison, papers submitted in 2011 or later that received a positive R&R decision are unlikely to have been published for long enough to receive a large citation boost. We also expect that the publication effect would be larger for higher impact journals. By contrast, we would not necessarily expect big differences in the correlation between the editor's signal and the unexplained component of citations along these dimensions.

Column 2 of Table 3 displays the evidence. The mechanical publication effect is, indeed, 49 log points larger for papers submitted earlier in our sample (up to 2010) than for papers submitted later. The mechanical publication effect is also larger for the highest-impact journal, the QJE, than for the other journals. Column 3 shows that there are no significant interactions with the control function term, suggesting that the informativeness of the editor's signal is roughly constant across these two dimensions.

The table also reports the coefficient estimates for two representative determinants of citations – the fraction of R&R recommendations, and an indicator for highly prolific authors. The estimates of these variables are essentially unaffected by the introduction of this more flexible model of the mechanical publication bias, compared to the benchmark model in Column 1.

We take these results as evidence that there are indeed positive effects of receiving an R&R decision on subsequent citations, particularly at the QJE and for papers reviewed further in the past. On average, though, this effect is not particularly large, nor does allowing for variation in the publication effect have any impact on our conclusions regarding the relative size of other determinants of citations. Further, as we documented above, even under the upper bound assumption that ignores any endogeneity in the R&R decision, the relative size of the estimated effects of the other determinants of citations are not much affected. Thus, in the rest of the paper we adopt as benchmark the specification with both publication bias and control function (Column 4 in Table 2).

The Revise and Resubmit Decision

We now 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 $paper \times referee$ 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 recommendations, rises to 25 percent with two *Weak R&R* recommendations, and to 80 percent or higher with two *R&R* recommendations. Similar patterns are present looking at all possible pairs of recommendations for papers with three referees (Online Appendix Figure 3a). Along similar lines, Welch (2014) compares referee recommendations and editorial decisions for an anonymous journal and shows that editorial decisions are highly related to the referees' opinions.

Columns 7-9 of Table 2 present the estimated coefficients for probit models that parallel the citation models, using only the referee recommendations (column 7), only the other controls (column 8), and finally both sets of variables and the editor leave-out-mean R&R rate (column 9). As might be expected given the patterns in Figure 4a, the model with only the referee recommendations and $journal \times submission year$ controls is remarkably successful, with a pseudo R^2 of 0.48.[24] The quality of fit is apparent in the comparison between Figure 4c, which plots predicted probabilities for each of the possible referee combinations for 2-referee papers, and Figure 4b, which shows the actual probabilities. The close fit of the model across the cells is also evident when we look at pairs of reports for papers with 3 referees (see Online Appendix Figures 3a-b).

Column 8 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 positive impact of 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 9 presents the full specification of equation (4) with both the referee variables and the other covariates. This specification also includes the editor leave-out-mean average R&R rate. The addition of these extra controls 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 referees' recommendations. As in column 8, the number of authors has no systematic effect. The leave-out-mean variable has a statistically significant effect (t=2.9): there is significant variation among the editors in their R&R rates.[25]

[24] 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.

[25] Our confidentiality agreement at REStat did not include access to editor information so for this journal we have no variation in editor's R&R rates. We simply include the leave out mean rate for all papers.

Coefficient Plots

With these results in hand, we turn to a comparison of the coefficients of various paper characteristics in the citation model and the R&R decision model. Our focus is on evaluating prediction P1, which states 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 9 of Table 2) against the corresponding coefficients from the citation model (column 4 of Table 2). For visual clarity, Figure 5a displays only the coefficients on the referee recommendation variables and the author publication variables, while Figure 5b shows all the coefficients. To aid in interpreting the results, 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 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 closely follows the views of the referees, and that 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. For example, measures of author publications have a much smaller effect on probability of R&R than would be expected given their impacts on citations. The degree of proportionality between the author publication variables in the R&R model and the citation model (shown by the red line in Figure 5a) is only about one fifth of the slope of the black line which shows the degree of proportionality between the referee recommendation coefficients in the models.

Indeed, comparing columns 4 and 9 of Table 2, the coefficients of the referee variables are about twice as big in the R&R model as in the citation model, implying 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 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, a fact that is clearly visible in Figure 5a.

The implication of these estimates is that editors are only partly offsetting the tendency of referees to discount the expected citations to papers by more prolific authors.We can quantify the size of this offset using the framework of equation (8). For example, consider the relative effect of a "revise and resubmit" opinion by the referees versus the effect of an author team with 4-5 recent publications. From our baseline citation model (column 4 of Table 2), these papers receive 85 log points more citations controlling for the referees' opinions. The implied bump in citations expected by editors, based on the coefficients in the R&R probit model (column 4 of Table 2) is 14 log points.[26] Thus, editors are "discounting" citations to papers by authors with 4-5 recent publications by $0.85 - 0.14 = 0.71$ (71 log points). We can perform this same calculation using the extreme

[26]Following equation (8), $\lambda^*_{4-5} = \widehat{\lambda}_{RR} \times \widehat{\pi}_{4-5}/\widehat{\lambda}_{4-5} = 1.97 \times 0.32/4.61 = 0.14$.

upper bound model for the publication effect in column 5 of Table 2: the implied degree of editor discounting is essentially the same (72 log points), illustrating the robustness of our conclusions to the treatment of publication bias. We can also estimate similar discount factors for other author groups. For example, the discount for authors with 6+ publications is 82 log points, versus the 101 log point effect implicit in the referees' opinions.

Figure 5b also displays the coefficients for the number of authors and the field of the paper. 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 fractions line up with their effects on citations (compare columns 2 and 4 of Online Appendix Table 2). Editors are putting much greater weight on referee recommendations relative to other variables that are also predictive of citations.

Do these patterns differ by journal? Online Appendix Figure 4 (based on the coefficients in Online Appendix Table 3) shows that several key patterns are common. First, within each group of variables the coefficients fall 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 impact on the R&R decision. This gap is particularly striking at REStud and REStat, where the editors appear to assign *no weight* to any variable other than the referee recommendations.[27] Since future citations are much lower for authors with few prior publications (conditional on the referee recommendations), the lack of attention to prior publications is consistent with the REStud's stated mission of supporting young economists.[28]

Visual Evidence on R&R and Rejects

As an additional piece of graphical evidence, in Figure 6 we plot the average citation rate for papers that receive an R&R and for those that are rejected. For each paper we predict the probability of a revise-and-resubmit decision using the specification in Column 9 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.

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. This gap, as we discussed above, captures the sum of any mechanical publication effect and the editor's value added. Interestingly, the vertical gap between R&Rs and rejects is wider to the left, as predicted by a model with informed editors: the editor has to receive a very positive signal for papers with relatively low observable

[27] Again, using the framework of equation (8) we can quantify the degree of discounting applied by editors at different journals to papers by more prolific authors. On average across all 4 journals editors discount the citations to authors with 4-5 recent publications by about 70 log points. This discount is about 65 log points at QJE, 90 log points at REStud, 74 log points at REStat, and 33 log points at JEEA.

[28] From the mission statement online: "[The] objective [of the Review] is to encourage research in theoretical and applied economics, especially by young economists".

quality in order to reach a positive R&R decision. Online Appendix Figure 8 displays the same data as in Figure 6 along with the predicted fit from our model, showing that the model does a good job of capturing the patterns in Figure 6.

Another salient feature of Figure 6 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 7^{th} 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 recommendations 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). This is consistent with the model and illustrates the informativeness of the referee variables that drive differences in predicted citations.

Other Citation Measures

A potential concern with the findings so far is that the results hinge on our use of the inverse hyperbolic sine transformation in modeling citations. To address this concern, in Table 4 we re-estimate the citation model using alternative transformations. Column 1 shows our base specification. Column 2 uses our percentile citation measure, which controls for differences in citations across journal-year cohorts flexibly by computing citation percentiles within cohorts. Column 3 is motivated by the hypothesis that editors focus on the probability that a paper becomes a "major hit". 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. Taking this point further, In Column 4, we use an indicator for a paper in the top 2% of citations in a journal-year cohort, proxying for "superstar" papers. We also consider a specification in column 5 using $\log(1 + citations)$ as an alternative to the *asinh* specification. Finally in column 7 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 6 shows a model for *asinh*(*GS citations*) fit to the 2006-2010 sample, which is similar to the baseline model in column 1.[29]

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

[29]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.

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 right tail of GS citations (columns 3 and 4), or the expected *asinh* or percentile of SSCI citations (columns 7-8). The one notable difference across specifications is that the estimated impact of the mechanical publication effect is much larger for the SSCI citations, as expected, given that in this case there is indeed an obvious mechanical confound.

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 years prior to submission. We now consider three additional measures of productivity. The first is a count of publications in the 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 a 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).

In Table 5 we augment our baseline models from Table 2 (reproduced in columns 1 and 4) with these additional measures. As shown 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 44 log points of citations, even conditional on all the other variables. Prior top-5 publications 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 5f).

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.[30] Since we only collected

[30]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.

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.[31]

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 51 log points, while having a coauthor at an 11-20 ranked US institution increases citations by 43 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 than the relative impact of the referee variables (see Online Appendix Figure 5g).

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 35 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 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.

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 (DR) decision, building on the simple model in Section 2.2. An empirical analysis of DR's 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.[32]

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 6). 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 6 are approximately 1.3 times larger than the coefficients in the model in column 4 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 $journal \times year$ fixed effects have a combined R^2 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 columns 3-4 of Online Appendix Table 6, show that editors use the prior publication record of authors in making their initial NDR decision, but put little

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

[32] 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.

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 7), 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 an important question because the desk rejection process is sometimes characterized as arbitrary or uninformed. Figure 7a 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.[33] On average, NDR papers receive about 80 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.[34]

The gap between NDR papers that are ultimately given an R&R and those that are rejected is also large – 140 log points. This gap reflects 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 ultimately receive R&Rs 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 72 log points. This gap is interesting because both sets of papers are rejected - thus, there is no mechanical publication effect biasing the comparison. Viewed this way, the editor's signal at the desk reject stage is relatively informative.

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 (13), 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.

[33] The gap between papers that are R&R'd and those that are rejected after review is larger than the corresponding gap in Figure 6 (for the same set of papers) because of the different ways of grouping papers along the x-axis – by probability of NDR in Figure 7a (based only on x_1) and by probability of R&R in Figure 6 (based on x_1 and x_R).

[34] Recall that according to our model the signal to total variance ratio is $A_0 = \rho_0^2$, where $\rho_1 = 0.31$ is the implied correlation of the editor's signal and the citation residual.

We present this comparison in Figure 7b, 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. At both stages there appears to be a higher bar for authors with a stronger publication record.

5 Interpretation and Survey

At all four journals in our sample referees and editors appear to impose a higher bar for papers by prolific authors. Figure 8a 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 7b). 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: referees evaluate papers as if the citations for prolific authors should be discounted by e^1.

Columns 1 and 2 of Table 6 quantify this discounting effect. When we include the controls for paper characteristics but not the referee recommendations (Column 1), papers from authors with 6+ publications have 134 log points higher citations, relative to authors with no recent publications. When we add in the referee recommendations (column 2), this gap is 101 log points.

In the R&R decision model in column 9 of Table 2 we saw that editors put positive weight on author's publications (given the referee opinions), "undoing" some of the bias against more prolific authors. We therefore expect a smaller citation effect for papers by more prolific authors when we limit the sample to papers that receive a positive R&R verdict. This is in fact the case, as shown by the specifications in columns 3 and 4 of Table 6, which are estimated on the subset of R&R papers. Among these papers the expected citation premium for papers from authors with 6+ previous publications, for example, falls from from around 100 log points to about 70 log points.[35]

To what extent does the citation advantage for papers of 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 papers receiving an R&R 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

[35] 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 6 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.

variables for these 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 using *asinh* GS citations as the dependent variable, 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 large.[36]

Finally, we constructed a third sample of papers published in the top 5 economics journals between 1997 and 2012, 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. The citation model for this sample (column 6) yields estimated author publication effects that are attenuated by about 50% relative to the effects in our R&R sample (column 3), but are still highly significant.

5.1 Interpretations

There are two main explanations for our key finding that referees and editors significantly underweight the expected citations of papers by more prolific authors. The first is that papers by prolific authors are *over-cited*, leading referees and editors to discount their citations accordingly. This could occur for several reasons. More prolific authors may have access to working paper series and other distribution channels that publicize their work, inflating citations. They may also have networks of colleagues and students who cite their work rather than related work by less prolific scholars. Finally, people may tend to cite the best known author when there are several possible alternatives - Merton's (1968) "Matthew effect."

An alternative interpretation is that citations are unbiased measures of quality, but referees and editors set 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 prevent established authors from publishing marginally acceptable papers (i.e., animus).[37] At least two pieces of evidence in the literature 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 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, we briefly discuss a third possibility that is sometimes raised in the editorial context: elite favoritism.[38] 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 in the same professional networks. If one takes citations as unbiased measures of quality, we clearly

[36]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.

[37]A related possibility is that editors impose a higher bar for prolific authors because they believe these authors will be less willing to revise their paper to accommodate the referees' and editors comments.

[38]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.

find substantial evidence against this 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) more prolific authors actually face a lower bar in the editorial process.

To test the elite favoritism hypothesis, we examined 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. We thus test whether the citation gap in Figure 8a differs when the referee has a strong publication record (and is therefore a potential member of the elite) or not.[39] Figure 8b displays no evidence of elite favoritism: the gap in citations between papers of prominent and non-prominent authors is the same whether the recommendation comes from a prolific or non-prolific referee. Interestingly, the seminal study by Zuckerman and Merton (1971) also found that more and less prominent referees tended to give similar assessments of papers by more and less prominent authors.

5.2 Survey Evidence on Quality vs. Citations

To distinguish between the two competing explanations for the down-weighting of expected citations for more prolific authors, we turn to survey evidence that allows us to separately measure quality and citations. We designed a survey of expert readers, asking them to compare paired papers in the same topical area, published in the same year in a similarly ranked (top) journal. The comparison was designed to mirror the R&R decision faced by a journal editor in selecting among submissions. It also mirrors our main empirical models, which include controls for field and fixed effects for journal-year cohorts. The comparison of papers *within* a field also makes the evaluation easier for survey respondents, and resembles the evaluation by referees who typically assess submissions in their field.

We selected paired sets of papers from articles published in one of the top-5 journals between 1999 and 2012, excluding AER P&P articles, notes, and comments. To better match the expertise of our survey population, we decided to focus on papers in 6 topical areas: (i) unemployment; (ii) taxation; (iii) crime; (iv) education; (v) family economics; and (vi) behavioral economics.[40] We also classified papers as mainly theoretical or mainly empirical, and paired similar papers within each field.

Following the same procedure as in our main analysis, we measure the publications of authors in the 35 high-impact journals in the 5 years prior to submission, assuming 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 none of the co-authors have more than 1 publication during this period. We note that some of the authors coded as non-prolific at the assumed submission year were clearly prolific in later years. This is as intended and reflects the procedure we used in our

[39] 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.

[40] 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.

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 author – published in one of the top-5 journals[41] 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 also exclude pairs with citations that were too imbalanced (a ratio of citations outside the interval from 0.2 to 5.0), and 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 sample included 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 are paired twice.

Survey Wording. The survey was administered on the Qualtrics platform, with all the questions displayed on one page (see Online Appendix Figure 9). The respondents were asked two main questions about each pair of papers. 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; and (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 appropriate ratio 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 included faculty and PhD students who specialize in the fields covered by the papers in the survey. The survey was administered in September and October 2016. Our analysis follows a pre-registered analysis plan, AEARCTR-0001669.

[41] 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*.

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:

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

We interpret $\widehat{R}_j$ as the respondent's assessment of the ratio of the quality of the paper by the prolific author in the j^{th} pair (q_{Pj}) to the quality of the paper by the non-prolific author (q_{Nj}), i.e.,

$$\widehat{R}_j = q_{Pj}/q_{Nj}.$$

Our model assumes that the relation between citations and quality is $\log c_{ij} = \log q_{ij} + \eta_{ij}$, where $i \in \{P, N\}$ and η_{ij} reflects non-quality-related determinants of citations for paper i in pair j. We assume that the within-pair gap in η_{ij} can be decomposed as $\eta_{Pj} - \eta_{Nj} = \eta_\Delta + e_j$, where η_Δ represents average excess (log) citations accruing to papers by more prolific authors and e_j is a random factor. It follows that

$$\log \widehat{R}_j = \log R_j - \eta_\Delta - e_j \tag{14}$$

Thus, we fit the simple regression model:

$$\log \widehat{R}_j = d_0 + d_1 \log R_j + \varepsilon_j. \tag{15}$$

According to our model we should estimate $d_0 = -\eta_\Delta$ and $d_1 = 1$.

A slightly more general model of the citation-quality relationship is $\log c_{ij} = \theta(\log q_{ij} + \eta_{ij})$, which allows a concave or convex mapping from quality to citations. It is straightforward to show that all the implications of the model in Section 2 remain unchanged when $\theta \neq 1$.[42] In this case, however, equation (14) becomes:

$$\log \widehat{R}_j = \frac{1}{\theta} \log R_j - \eta_\Delta - e_j$$

and the predicted value for the coefficient d_1 in equation (15) is $d_1 = 1/\theta$.

Figure 9a illustrates two possible patterns of results using simulated data. We bin papers into

[42] The only change is that the coefficients in the citation model, equation (6), take on the values $\lambda_0 = \theta(\beta_0 + \eta)$, $\lambda_1 = \theta(\beta_1 + \eta_1)$, $\lambda_R = \theta\beta_R$, and the residual in the citation model becomes $\theta(\phi_q + \phi_\eta)$. 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 $1/\theta\sigma_v$.

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 no quality discounting: the regression line runs through the origin. The continuous blue line shows the case where respondents think that papers by more prolific authors get 28 log points more citations than would be justified by quality, implying an intercept for the regression of $d_0 = -0.28$.

Figure 9b shows a bin-scatter of our actual data. Following our pre-analysis plan we winsorized 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 very close to 0. Panel A of Table 7 provides a series of estimates of the model specified by equation (15), with a simple OLS regression in Column 1 and a specification in Column 2 in which we weight the responses for a given paper pair by the inverse of the number of respondents who evaluated the pair. In Column 3 we limit the sample to pairs with more comparable citations ($-0.5 \leq \log R_j \leq 0.5$). These three specifications suggest that holding constant quality, papers by more prolific authors receive between 1 and 3 percent more citations than those of less prolific authors.

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 the 35 journals). Interestingly, any tendency to attribute excess citations to more prolific authors comes from prolific faculty, rather than from graduate students or faculty respondents with relatively few publications. This pattern provides no evidence of elite favoritism and suggests instead that the downweighting of citations to papers by relatively prolific authors 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 code from -2 to +2 so positive values correspond to a higher rating for the paper by the prolific author. As shown in Figure 9c, 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 7 presents regressions in which we relate the relative strength of the paper by the prolific author in a pair to the relative citation measure. Consistent with Figure 9c, only the model for "Importance" (column 2) has an R^2 above 0.05. Again, the key coefficient for our purposes is the constant, which (with a sign change) we interpret as an estimate of the excess citations received by more prolific authors, holding constant the relative quality in the particular domain. None of the estimated constants are large or even marginally significant, which is consistent with the result in Panel A. Overall, these results provide no evidence that papers by prolific authors receive more citations than those by non-prolific authors, controlling for their relative quality.

6 Comparing our findings to the predictions of experts

Are these results on editorial decision-making surprising? To provide some evidence, we collect forecasts as in DellaVigna and Pope (forthcoming). In the Fall of 2015, in advance of a presentation of this paper, we surveyed a group of editors and associate editors at the *REStud*, and a group of faculty and graduate students at the economics department of the University of Zurich.[43] Specifically, we emailed the link to an 11-question Qualtrics survey. We received 12 responses by editors and associate editors (editors for brevity) at the REStud and 13 faculty and 13 graduate students in Zurich. No draft had been available at the time and these were among the first presentations, making it very unlikely that the respondents could have known about the results.

Table 8 presents the responses to the most relevant questions (not in the same order in which they were asked). Overall, the forecasts by editors and faculty respondents are quite accurate, with an average absolute deviation in percentage points between the correct answer and the average forecast of 5.5 (editors) and 4.2 (faculty). In comparison, graduate students have a deviation of 8.2.

The patterns of deviation are also of interest. The first two questions elicit a measure of how well editors navigate the type 1-type 2 tradeoff at the R&R stage ("*What percent of papers with a Revise-and-Resubmit in the first round end up in the top 5 percent of citations (by the Google Scholar measure)?*") and at the desk-reject stage ("*What percent of desk-rejected papers end up in the top 5 percent of citations (by the Google Scholar measure)?*"). Editors overestimate their ability to pick top-cited papers at the R&R stage (average forecast of 32.5% versus the actual 18.1%), with a smaller, but directionally similar, pattern for desk-rejections (average forecast of 0.9% versus the actual 1.3%). The faculty and graduate students in Zurich, instead, err in the opposite direction, especially underestimating the informativeness of desk rejection decisions.

The next question provides a measure of the belief of whether prominent authors face a different threshold in the desk-rejection decision, compared to other authors: "*Consider all submissions with at least one 'prominent' coauthor that are desk- rejected. What percent of these papers end up in the top 5 percent of citations?*".[44] Editors appear aware that at the desk-reject stage they set a higher bar for papers by prominent authors (average forecast of 6.4% versus the actual 5.2%). The Zurich faculty, instead, give a similar answer for this question and the question for all desk rejections, suggesting that they do not expect prominent authors to be treated differently, conditional on citations. On one of our key findings, thus, there appears to be significant disagreement.

Next, we ask for a forecast of how predictive referee recommendations are of citations: "*How much higher is the percentile citation if a referee recommendation is positive versus if it is negative (for papers with 3 reports)?*" The respondents overestimate, on average, this informativeness.

Finally, we elicit a measure of how closely the editors follow the referees. To keep things simple, we ask for the share of papers with 3 reports that receive an R&R, as a function of the referee recommendations. The respondents are quite well calibrated overall, but they underestimate how closely editors follow the majority of referees. Editors expect that for papers with one positive and two negative referee recommendations, 21.3% percent get an R&R, while the true share is only 6.4%.

[43]The survey sent to the REStud editors refers only to the REStud findings, while the Zurich survey refers to all 4 journals.

[44]We unfortunately did not ask the parallel question for the R&R decision.

There is, thus, interesting variation in the deviations of the forecasts from the observed editorial patterns. We hope that this paper contributes to a more shared understanding of the editorial process among authors, referees, and editors.

7 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 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.

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 than those that are rejected, conditional on the referees' recommendations.

Nevertheless, there are important deviations. On the referee side, certain paper characteristics are strongly correlated with future citations, controlling for the referee recommendation. 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. In particular, referees appear to substantially discount the future citations that will be received by more prolific authors. At best, editors at the four journals only partially offset these tendencies.

We consider two main interpretations. Citations may be inflated measures of quality for prolific authors, leading referees and editors to discount citations accordingly. Alternatively, citations may be appropriate measures of quality but referees and editors may set a lower threshold for less prolific authors, perhaps reflecting a desire to help these authors. While our main analysis cannot separate the two interpretations, the results from a survey of economists asked to evaluate the quality of pairs of papers are most consistent with preference-based explanation that referees and editors are effectively easing entry into the discipline for younger and less established authors.

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. Another important set of questions concern the initial selection of referees and the dynamic process by which editors decide whether to reach a decision with the reports received so far, wait for more of the original referee(s) to respond, or recruit new referees. We hope that future research will be able to address these and other questions.

References

Aizer, Anna, and Joseph J. Doyle, Jr. 2015. "Juvenile Incarceration, Human Capital, and Future Crime: Evidence from Randomly Assigned Judges." *Quarterly Journal of Economics,* 130 (2): 759-803.

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.

Brinch, Christian N., Magne Mogstad, and Matthew Wiswall. Forthcoming. "Beyond LATE with a Discrete Instrument." *Journal of Political Economy.*

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.

Dahl, Gordon B., Andreas Ravndal Kostøl, and Magne Mogstad. 2014. "Family Welfare Cultures." *Quarterly Journal of Economics*, 129(4): 1711-1752.

DellaVigna, Stefano, and Devin Pope. Forthcoming. "Predicting Experimental Results: Who Knows What?" *Journal of Political Economy.*

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

Ellison, Glenn. 2011. "Is Peer Review in Declie?" *Economic Inquiry*, 49(3), 635-657.

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.*

Gans, Joshua and George Shepherd. 1994. "How Are The Mighty Fallen: Rejected Classic Articles By Leading Economists," *Journal of Economic Perspectives*, 8(1), 165-179.

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

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

Heckman, James J. 1976. "The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models." *Annals of Economic and Social Measurement*, 5(4), 475-492.

Heckman, James J. 1979. "Sample Selection Bias as a Specification Error." *Econometrica*, 47(1), 153-161.

Heckman, James J., and Richard Robb, Jr. 1985. "Alternative methods for evaluating the impact of interventions: An overview." *Journal of Econometrics*, 30(1): 239-267.

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.

Knowles, John, Nicola Persico, and Petra Todd. 2001. "Racial bias in motor vehicle searches: Theory and evidence." *Journal of Political Economy*, 109(1): 203-229.

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, and 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. 2017. "Expertise vs. Bias in Evaluation: Evidence from the NIH" *American Economic Journal: Applied Economics*, 9(2): 60-92.

Maestas, Nicole, Kathleen J. Mullen, and Alexander Strand. 2013. "Does Disability Insurance Receipt Discourage Work? Using Examiner Assignment to Estimate Causal Effects of SSDI Receipt." *American Economic Review*, 103(5): 1797-1829.

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.

Smart, Scott, and Joel Waldfogel. 1996. "A citation-based test for discrimination at economics and finance journals." NBER Working Paper 5460.

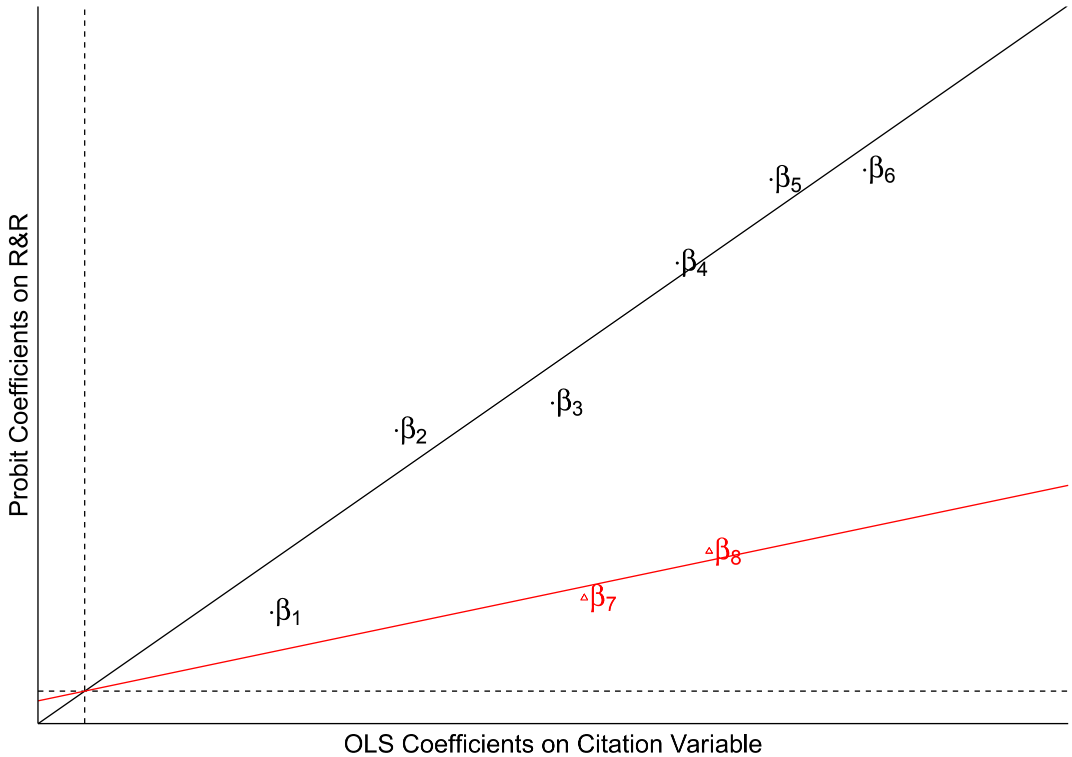
Vranceanu, Radu, Damien Besancenot, and Kim Huynh. 2011. "Desk rejection in an academic publication market model with matching frictions." ESSEC Working Paper.

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

Wooldridge, Jeffrey M. 2015. "Control Function Methods in Applied Econometrics." *Journal of Human Resources*, 50(2): 420-445.

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



Notes: Figure 1 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 2. Summary Statistics

Figure 2a. Distribution of Editorial Decisions

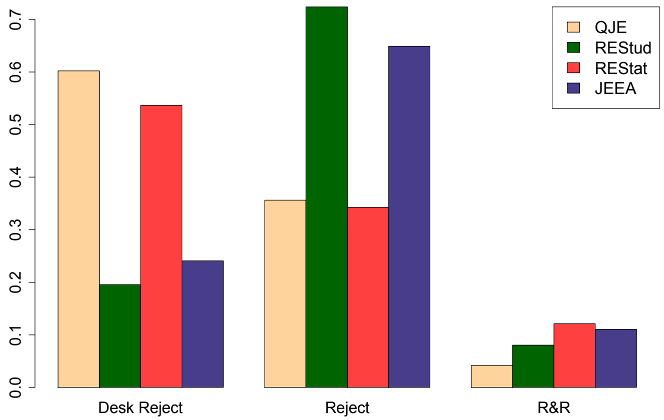


Figure 2b. Distribution of Referee Recommendations

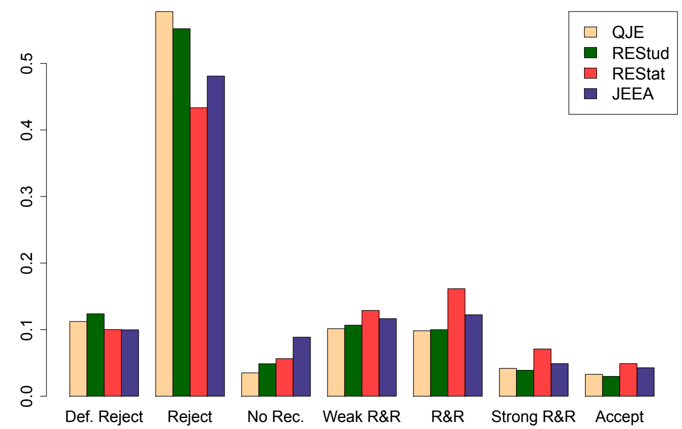
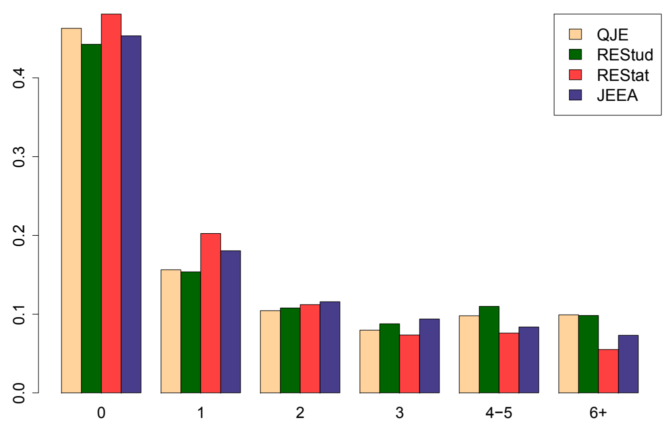


Figure 2c. Distribution of Author Prominence



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

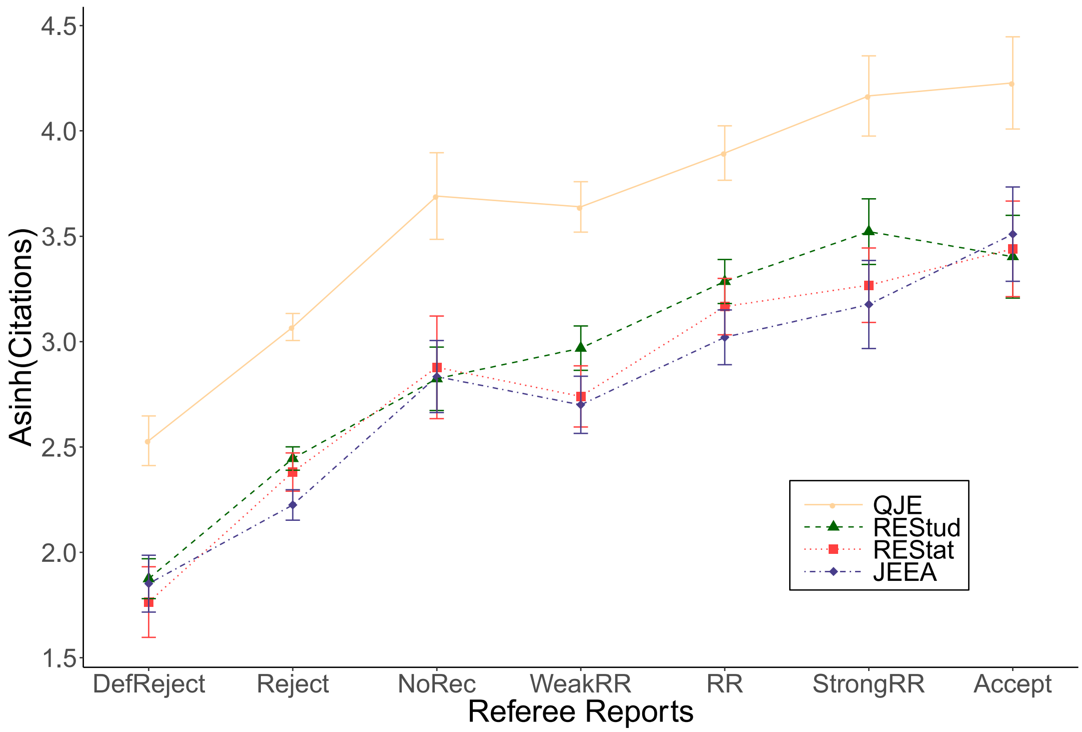


Figure 3b. Impact on Citation Percentile

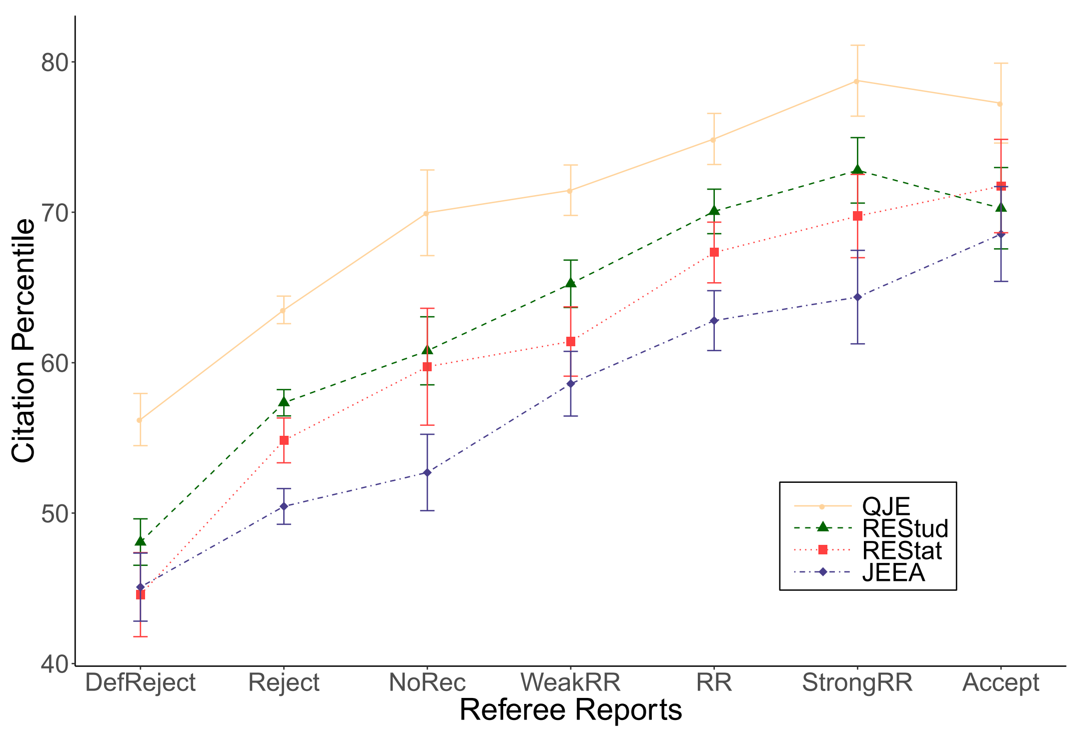


Figure 3c. Citation and Combination of Reports, Papers with 2 Reports, Data

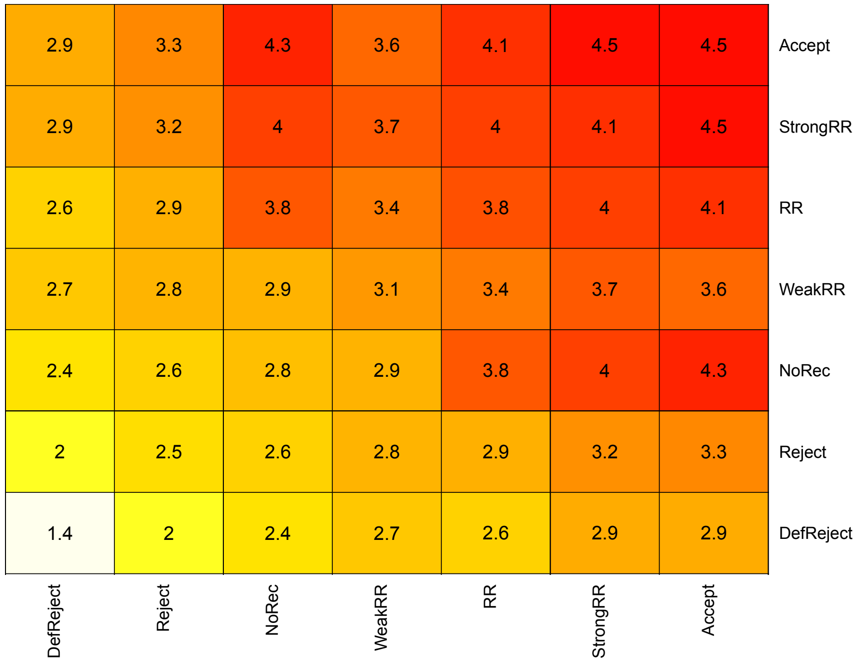
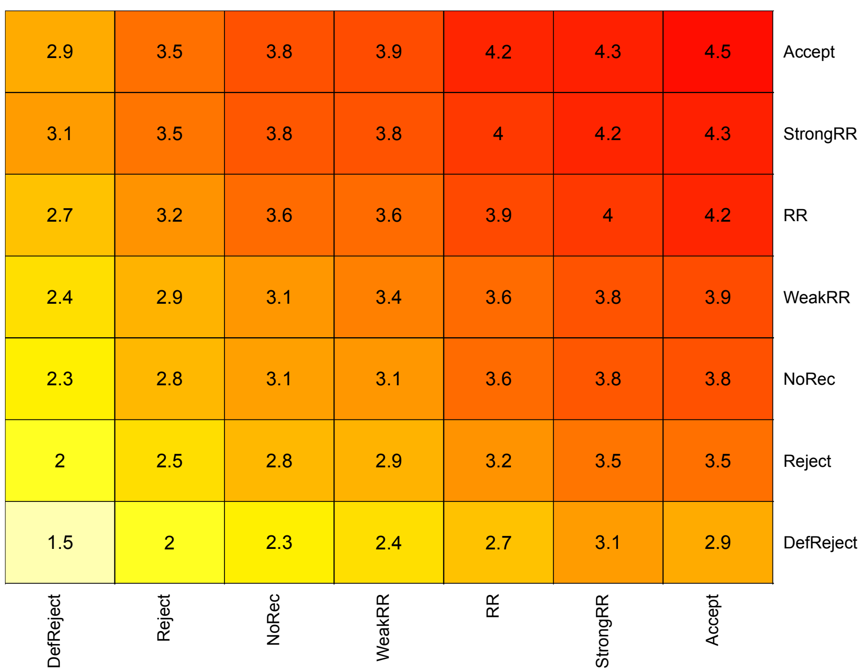


Figure 3d. Citation and Combination of Reports, Papers with 2 Reports, Model Prediction



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

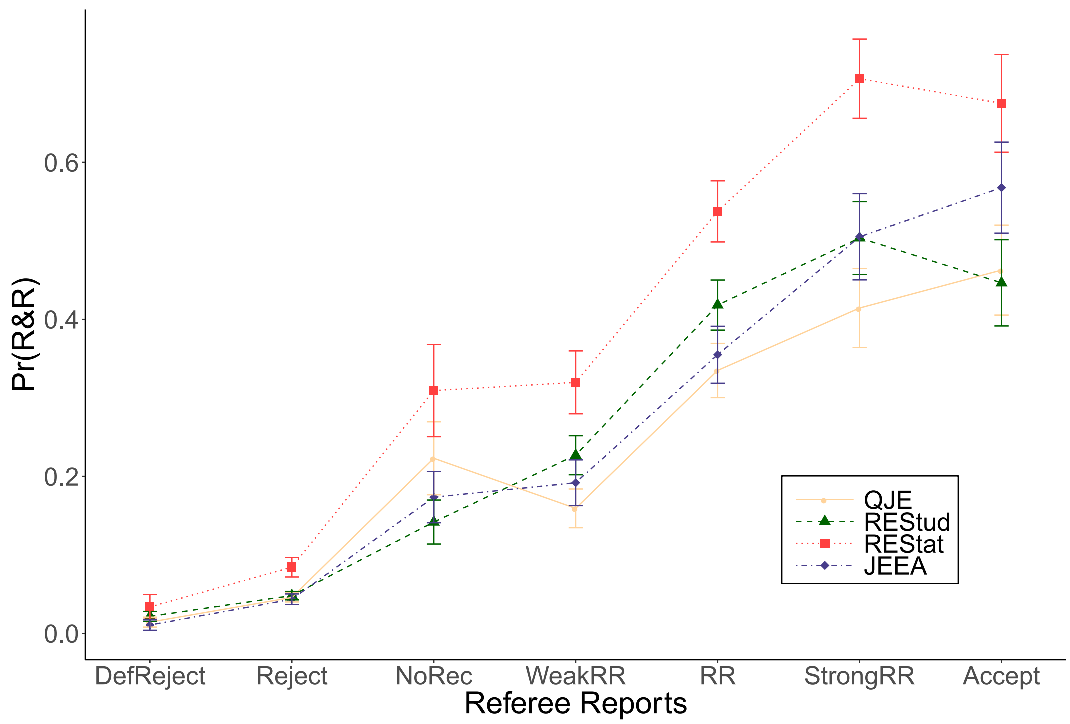


Figure 4b. Combinations of Referee Recommendations and R&R, 2-Report Papers

	DefReject	Reject	NoRec	WeakRR	RR	StrongRR	Accept
Accept	0.136	0.32	0.684	0.646	0.87	0.944	1
StrongRR	0.167	0.357	0.829	0.796	0.909	0.92	0.944
RR	0.065	0.156	0.434	0.669	0.793	0.909	0.87
WeakRR	0.052	0.052	0.22	0.242	0.669	0.796	0.646
NoRec	0.02	0.03	0.157	0.22	0.434	0.829	0.684
Reject	0.003	0.003	0.03	0.052	0.156	0.357	0.32
DefReject	0	0.003	0.02	0.052	0.065	0.167	0.136

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

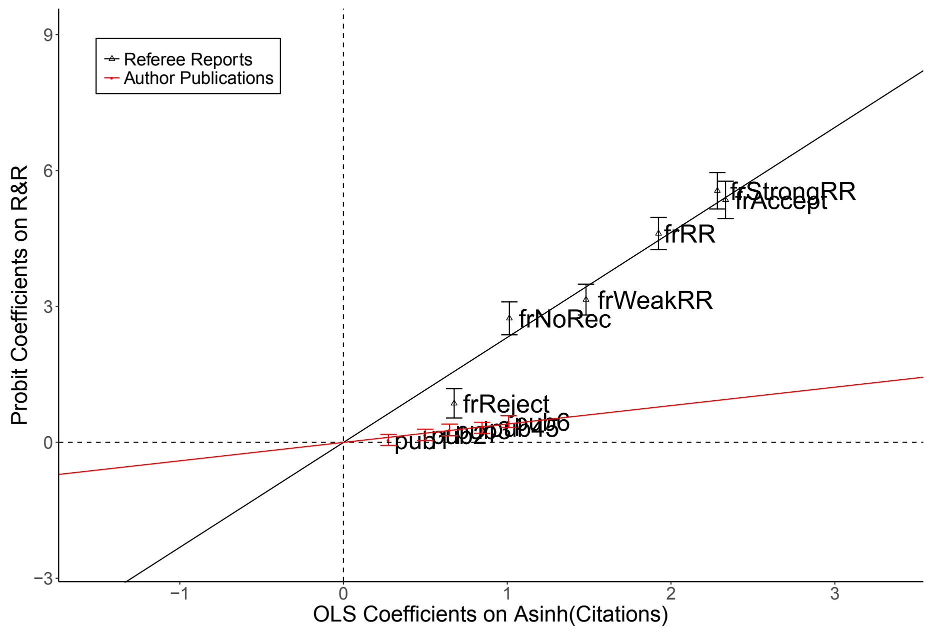
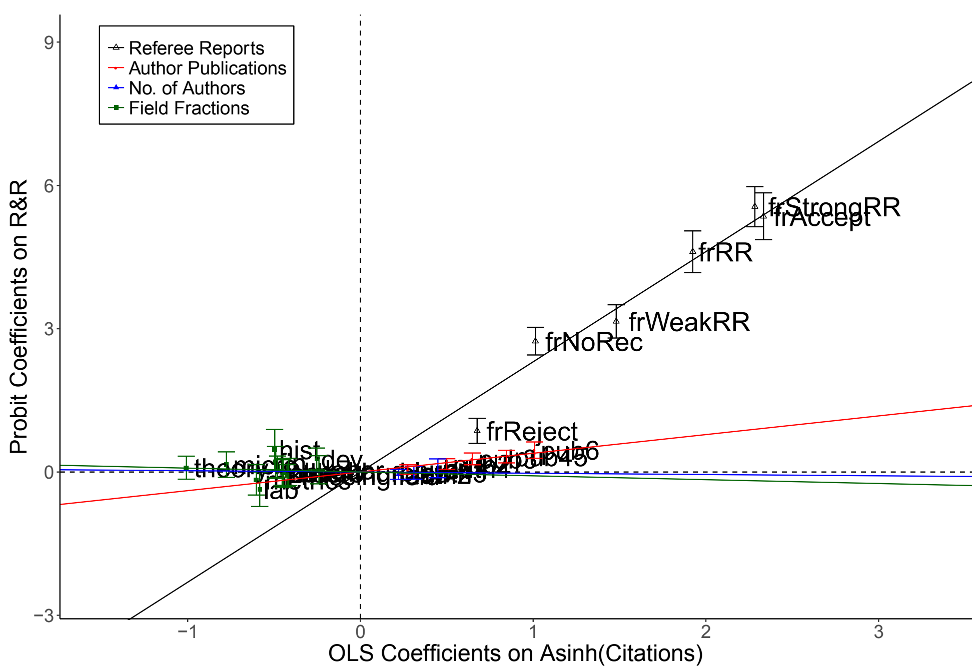
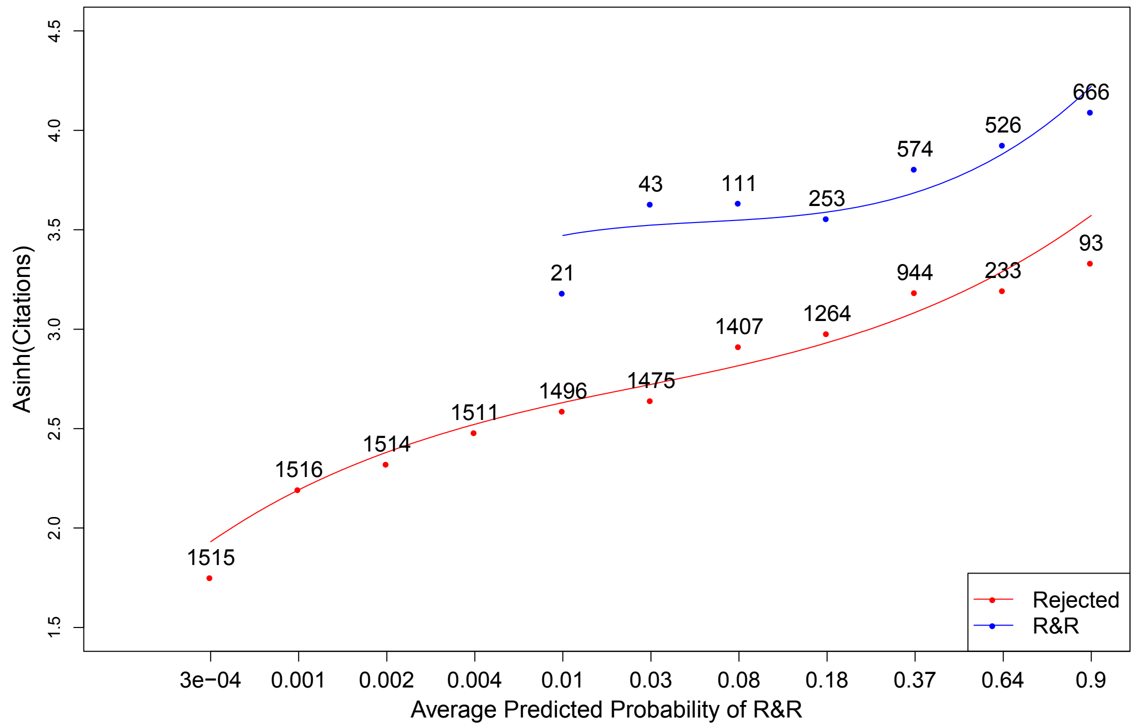


Figure 5b. Coefficients for Referee Recommendations, Author Publications, Number of Authors and Fields



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 Relationship Between the Editor's Revise and Resubmit Decision and Realized Citations



Notes: Figure 6 shows the average Asinh(citations) by deciles of predicted probability of R&R where the top decile is further split into two ventiles. Figure 6 considers separately papers that were rejected and those that the editor granted a revise-and-resubmit (using papers from the entire sample period). The smoothing lines are obtained via cubic fits to all data points.

Figure 7. The Relationship Between the Editor's Desk Rejection Decision and Realized Citations

Figure 7a. Plot by Quantiles in probability of avoiding desk-rejection

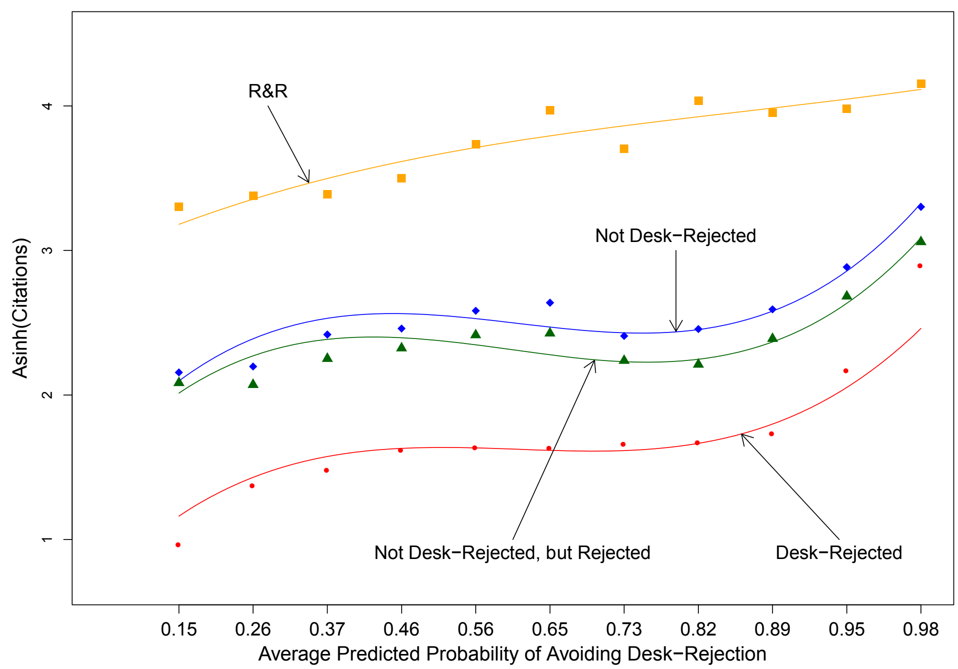
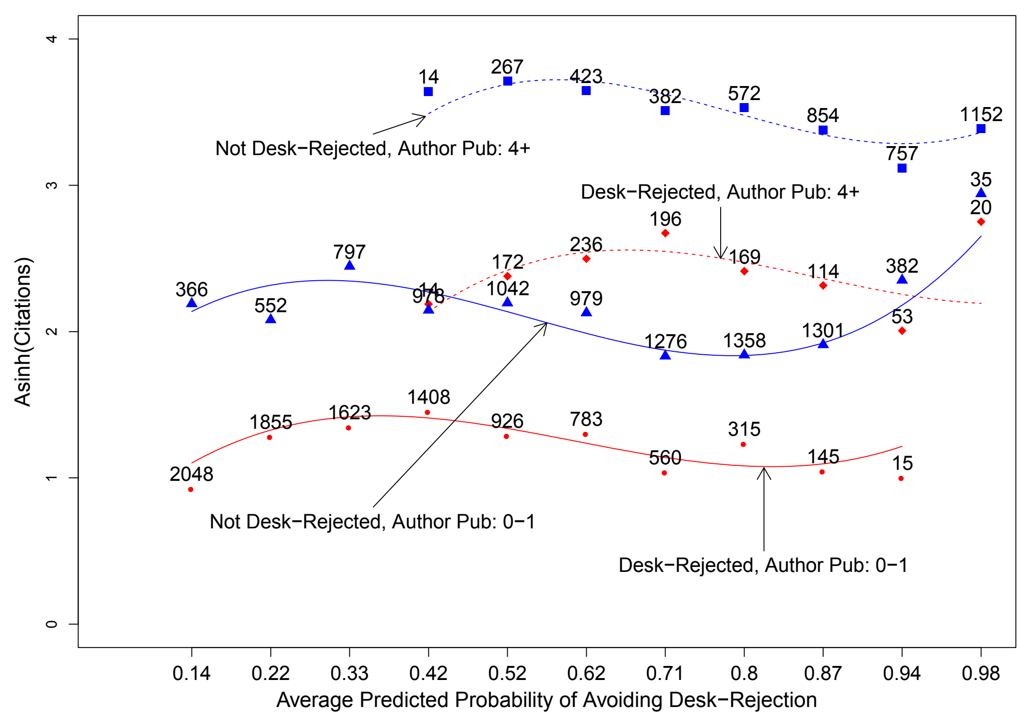


Figure 7b. Plot by Quantiles in probability of avoiding desk-rejection, split by author publications



Notes: Figure 7 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 7a 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 7b 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 8. Discounting of Citations of Prolific Authors, Referees

Figure 8a. All Referees

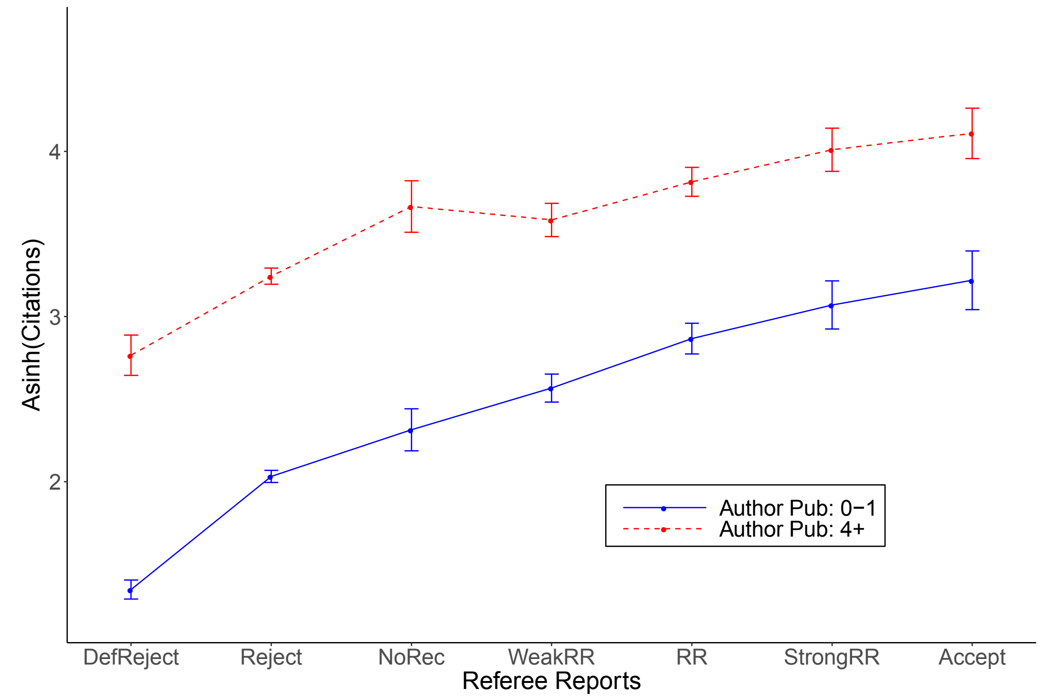
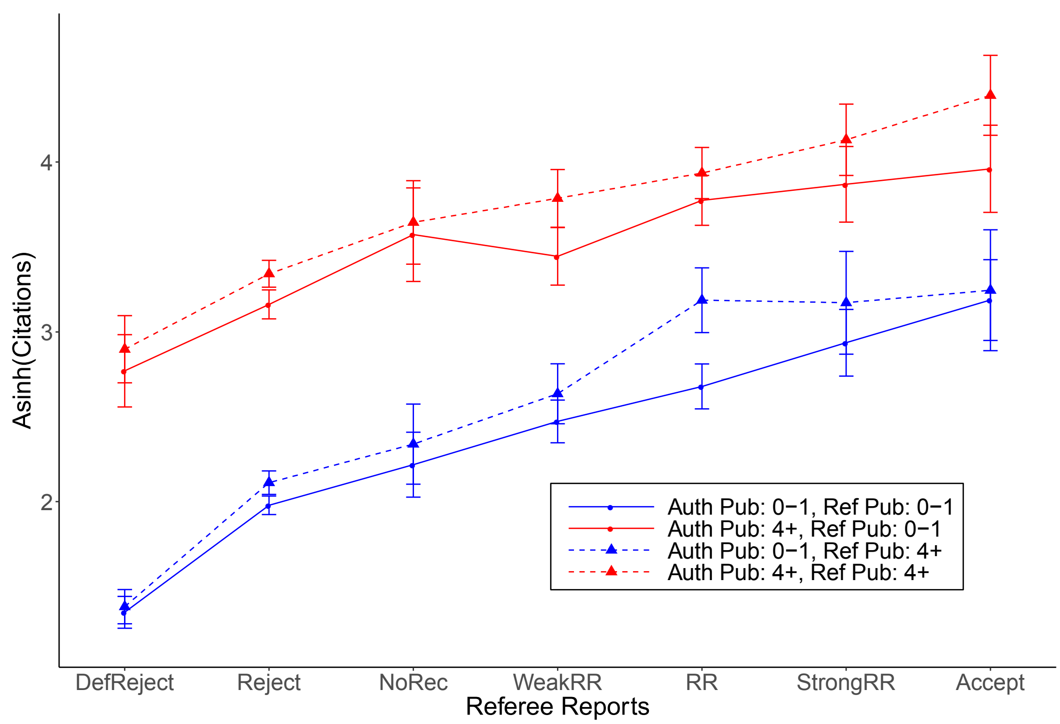


Figure 8b. Split by Prolific and Non-Prolific Referees



Notes: Figure 8 shows the weighted Arsinh(citations) for a paper receiving a given recommendation. Figure 8a shows the results separately for authors with 0-1 recent publications and authors with at least 4 recent publications, while figure 8b 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 9. Evidence on Citation Discounting from Survey of Economists

Figure 9a. Assessed Relative Citations versus Actual Citation Ratio, Theoretical Cases

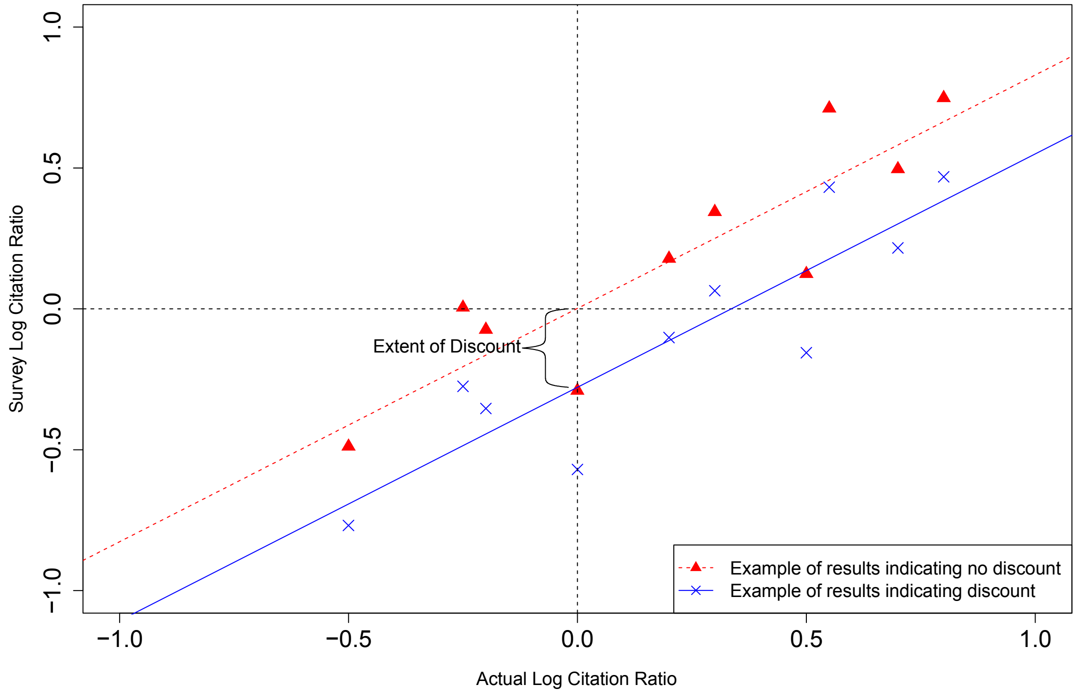
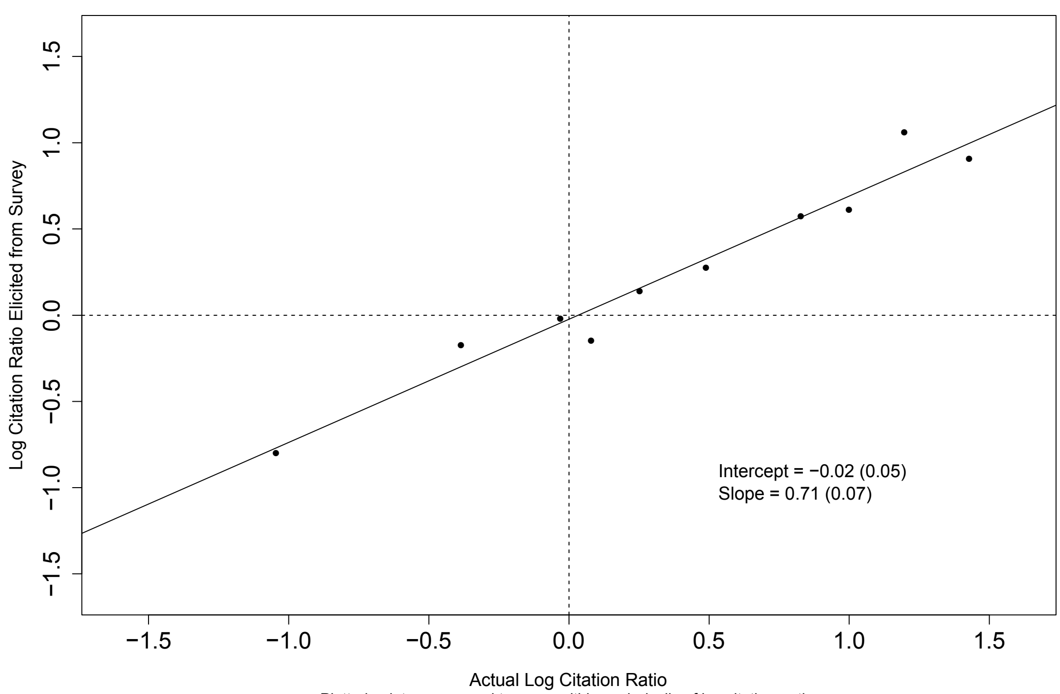
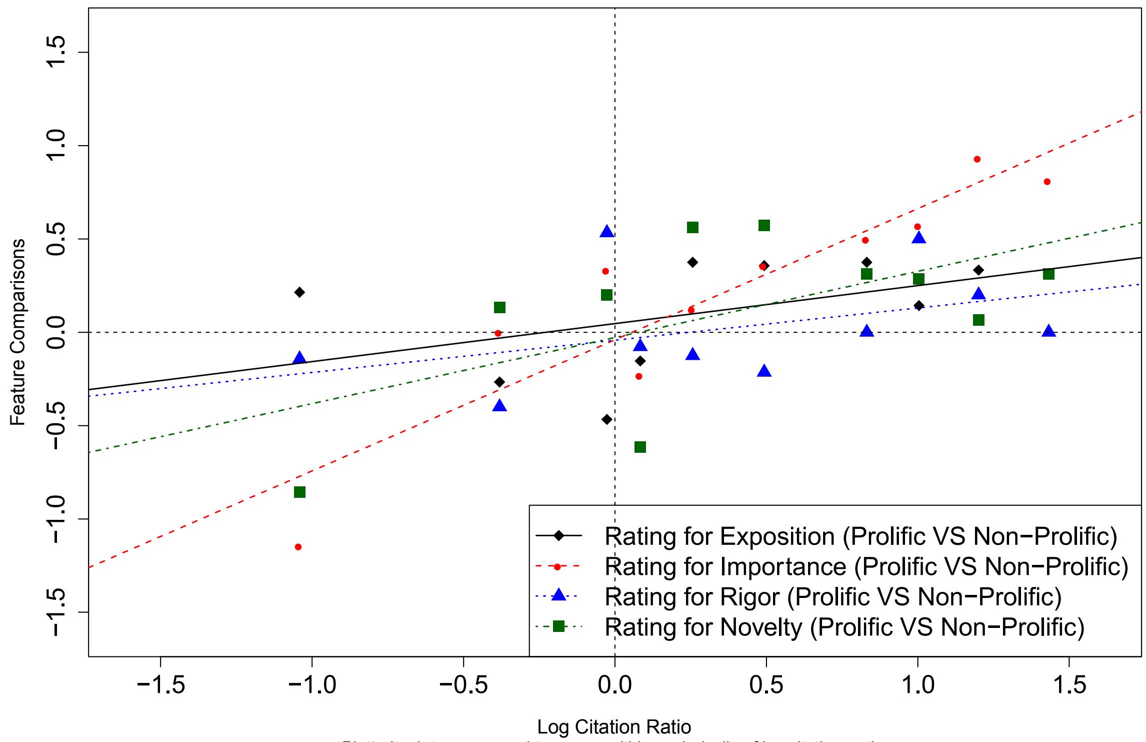


Figure 9b. Assessed Relative Citations versus Actual Citation Ratio, Decile Bins



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

Figure 9c. Qualitative Assessment of Paper Pairs, Decile Bins



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

Notes: Figures 9a-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 9a 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 9b 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 9c 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 9b.

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+	0.09	0.10	0.06	0.07	0.10	0.14	0.19	0.09	0.09	0.14
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)	(7)	(8)	(9)
Fractions of Referee Recommendations									
Reject	0.83		0.68	0.68	0.68	0.68	0.87		0.86
	(0.07)		(0.06)	(0.06)	(0.06)	(0.06)	(0.13)		(0.13)
No Recommendation	1.26		1.03	1.01	0.90	1.03	2.79		2.74
	(0.11)		(0.10)	(0.10)	(0.09)	(0.09)	(0.13)		(0.14)
Weak R&R	1.78		1.50	1.48	1.33	1.50	3.16		3.15
	(0.10)		(0.09)	(0.11)	(0.10)	(0.09)	(0.16)		(0.17)
R&R	2.37		1.97	1.92	1.57	1.97	4.64		4.61
	(0.09)		(0.08)	(0.14)	(0.08)	(0.07)	(0.21)		(0.21)
Strong R&R	2.76		2.34	2.28	1.79	2.34	5.58		5.55
	(0.18)		(0.15)	(0.25)	(0.17)	(0.15)	(0.20)		(0.20)
Accept	2.78		2.39	2.33	1.87	2.39	5.39		5.35
	(0.13)		(0.12)	(0.20)	(0.13)	(0.12)	(0.24)		(0.24)
Author Publications in 35 high-impact journals									
1 Publication		0.40	0.27	0.27	0.27	0.27		0.26	0.05
		(0.04)	(0.04)	(0.04)	(0.04)	(0.04)		(0.04)	(0.05)
2 Publications		0.66	0.50	0.50	0.49	0.50		0.36	0.17
		(0.04)	(0.04)	(0.04)	(0.04)	(0.04)		(0.05)	(0.06)
3 Publications		0.88	0.65	0.65	0.64	0.65		0.55	0.27
		(0.03)	(0.03)	(0.03)	(0.03)	(0.03)		(0.04)	(0.06)
4-5 Publications		1.11	0.85	0.85	0.83	0.85		0.67	0.32
		(0.06)	(0.05)	(0.05)	(0.05)	(0.05)		(0.05)	(0.07)
6+ Publications		1.34	1.01	1.01	0.98	1.01		0.86	0.46
		(0.06)	(0.05)	(0.06)	(0.05)	(0.05)		(0.06)	(0.08)
Number of Authors									
2 authors		0.19	0.22	0.22	0.22	0.22		-0.12	-0.05
		(0.04)	(0.04)	(0.04)	(0.04)	(0.04)		(0.05)	(0.05)
3 authors		0.24	0.31	0.31	0.31	0.31		-0.16	-0.02
		(0.05)	(0.04)	(0.05)	(0.05)	(0.04)		(0.04)	(0.07)
4+ authors		0.41	0.45	0.45	0.44	0.45		-0.04	0.08
		(0.07)	(0.07)	(0.07)	(0.07)	(0.07)		(0.08)	(0.10)
R&R Indicator				0.06	0.57				
(Mechanical Publ. Effect)				(0.14)	(0.05)				
Control Function for Selection				0.32		0.35			
(Value Added of the Editor)				(0.08)		(0.03)			
Editor Leave-out-Mean R&R									2.73
Rate									(0.94)
Controls for Field of Paper	No	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes
Indicators for Journal-Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R^2 / pseudo R^2	0.20	0.20	0.26	0.27	0.27	0.27	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 1-6 is asinh of Google Scholar citations. Dependent variable in probit models in columns 7-9 is indicator for receiving revise and resubmit decision. The control function for selection in columns 4 and 6 are calculated using predicted probabilities based on column 9. We code REStat as having a single editor because we lack information on editors for REStat. Standard errors clustered by editor in parentheses.

Table 3. Citation Models: Publication Bias vs. Editor Signal

	OLS Models for Asinh of Google Scholar Citations		
	(1)	(2)	(3)
Control for Publication Bias			
R&R Indicator	0.06	-0.34	-0.39
	(0.14)	(0.15)	(0.15)
R&R Indicator *		0.50	0.53
(Submission Up to 2010)		(0.07)	(0.10)
R&R Indicator * QJE		0.33	0.41
		(0.07)	(0.13)
R&R Indicator * REStud		0.05	0.09
		(0.07)	(0.10)
Control for Selection			
Control Function for Selection	0.32	0.32	0.36
	(0.08)	(0.08)	(0.09)
Control Function for Selection*			-0.03
(Submission Up to 2010)			(0.08)
Control Function for Selection*			-0.08
QJE			(0.08)
Control Function for Selection*			-0.04
REStud			(0.09)
Fractions of Referee Recommendations (Other Fractions Included, not Reported)			
R&R	1.92	1.92	1.92
	(0.14)	(0.13)	(0.13)
Author Prominence			
6+ Publications	1.01	1.00	1.00
	(0.06)	(0.06)	(0.06)
Control for Other Referee Recs, Author Prominence and No. of Authors	Yes	Yes	Yes
Controls for Field of Paper	Yes	Yes	Yes
Indicators for Journal-Year	Yes	Yes	Yes
R^2 / pseudo R^2			

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. The dependent variable is asinh of Google Scholar citations for all specifications. The control function for selection in all specifications of this table is calculated using predicted probabilities from column 9 of Table 2. We code REStat as having a single editor because we lack information on editors for REStat. Standard errors clustered by editor in parentheses.

Table 4. Models for Alternative Measures of Citations

	OLS Model for Asinh(GS Citations)	OLS Model for GS Citation Percentile	Probit Model for Top Group of GS Citations	Probit Model for Top 2% 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**	**All Years**	**2006-2010**	**2006-2010**	**2006-2010**
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Fractions of Ref. Recommendations								
Reject	0.68	10.87	0.29	0.31	0.57	0.72	0.28	6.18
	(0.06)	(0.98)	(0.08)	(0.12)	(0.05)	(0.08)	(0.04)	(1.16)
No Recomm.	1.01	16.49	0.55	0.57	0.86	1.11	0.60	12.18
	(0.10)	(1.34)	(0.10)	(0.19)	(0.09)	(0.12)	(0.13)	(2.37)
Weak R&R	1.48	23.98	0.78	0.71	1.27	1.56	0.60	12.05
	(0.11)	(1.58)	(0.13)	(0.16)	(0.11)	(0.11)	(0.09)	(1.97)
R&R	1.92	31.63	1.07	0.79	1.64	2.11	0.78	16.60
	(0.14)	(2.15)	(0.14)	(0.21)	(0.12)	(0.14)	(0.13)	(2.75)
Strong R&R	2.28	37.44	1.19	0.99	1.96	2.54	1.14	21.57
	(0.25)	(3.42)	(0.25)	(0.29)	(0.23)	(0.28)	(0.26)	(4.37)
Accept	2.33	37.94	1.33	1.19	2.01	2.65	1.54	29.44
	(0.20)	(2.67)	(0.22)	(0.29)	(0.18)	(0.16)	(0.22)	(2.95)
Author Publications in 35 high-impact journals								
1 Publication	0.27	4.20	0.17	0.24	0.23	0.26	0.11	1.64
	(0.04)	(0.60)	(0.05)	(0.14)	(0.03)	(0.05)	(0.03)	(0.90)
2 Publications	0.50	7.91	0.31	0.36	0.43	0.57	0.27	5.35
	(0.04)	(0.57)	(0.04)	(0.09)	(0.03)	(0.04)	(0.05)	(0.87)
3 Publications	0.65	10.20	0.39	0.49	0.56	0.64	0.25	5.30
	(0.03)	(0.49)	(0.05)	(0.09)	(0.03)	(0.03)	(0.04)	(0.61)
4-5 Publications	0.85	13.17	0.50	0.58	0.74	0.88	0.49	9.63
	(0.05)	(0.72)	(0.05)	(0.12)	(0.05)	(0.06)	(0.05)	(1.23)
6+ Publications	1.01	15.25	0.69	0.81	0.89	0.99	0.51	8.99
	(0.06)	(0.82)	(0.05)	(0.11)	(0.05)	(0.08)	(0.07)	(1.09)
Control for R&R	0.06	-1.18	0.21	0.30	0.10	-0.01	0.77	11.78
(Mech. Publ. Effect)	(0.14)	(2.48)	(0.11)	(0.19)	(0.12)	(0.18)	(0.15)	(2.89)
Control Function	0.32	5.65	0.16	0.13	0.27	0.46	0.08	2.20
(Ed. Value Added)	(0.08)	(1.27)	(0.07)	(0.09)	(0.07)	(0.10)	(0.09)	(1.51)
No. of Observations	15,177	15,177	15,177	15,177	15,177	8,208	8,208	8,208
R^2 / pseudo R^2	0.27	0.20	0.15	0.16	0.28	0.25	0.22	0.13

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, controls for field(s), and for number of authors. Models in columns 1-6 use Google Scholar (GS) citations. Models in columns 7-8 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. The control function for selection in all specifications of this table is calculated using predicted probabilities from column 9 of Table 2. We code REStat as having a single editor because we lack information on editors for REStat. Standard errors clustered by editor in parentheses.

Table 5. Models with Additional Measures of Author and Institutional Prominence

	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 (Other Fractions Included, not Reported)						
R&R	1.92	1.87	1.76	4.61	4.60	4.98
	(0.14)	(0.13)	(0.17)	(0.21)	(0.21)	(0.33)
Author Publications in 35 high-impact journals up to 5 years before submission						
1 Publication	0.27	0.25	0.25	0.05	-0.01	-0.03
	(0.04)	(0.04)	(0.06)	(0.05)	(0.05)	(0.08)
2 Publications	0.50	0.43	0.51	0.17	0.03	-0.10
	(0.04)	(0.05)	(0.05)	(0.06)	(0.07)	(0.10)
3 Publications	0.65	0.52	0.50	0.27	0.08	-0.07
	(0.03)	(0.04)	(0.06)	(0.06)	(0.07)	(0.11)
4-5 Publications	0.85	0.63	0.63	0.32	0.06	-0.10
	(0.05)	(0.06)	(0.07)	(0.07)	(0.08)	(0.11)
6+ Publications	1.01	0.61	0.61	0.46	0.06	-0.13
	(0.06)	(0.08)	(0.09)	(0.08)	(0.10)	(0.13)
Author Publications in Top 5 Journals						
1 Publication		0.29	0.20		0.21	0.19
		(0.04)	(0.06)		(0.05)	(0.06)
2 Publications		0.44	0.28		0.26	0.23
		(0.04)	(0.07)		(0.07)	(0.10)
3+ Publications		0.56	0.37		0.43	0.39
		(0.06)	(0.09)		(0.09)	(0.10)
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
		(0.05)	(0.05)		(0.06)	(0.06)
4+ Publications		0.06	0.03		0.16	0.35
		(0.07)	(0.08)		(0.07)	(0.07)
Rank of Authors' Institution						
US: 1-10			0.51			0.25
			(0.06)			(0.07)
US: 11-20			0.43			0.29
			(0.06)			(0.07)
Europe: 1-10			0.35			0.06
			(0.06)			(0.08)
Rest of the World: 1-5			-0.26			0.23
			(0.13)			(0.20)
Control for R&R	0.06	0.02	0.03			
(Mechanical Publ. Effect	(0.14)	(0.14)	(0.18)			
Control Function	0.32	0.33	0.26			
(Ed. Value Added)	(0.08)	(0.08)	(0.10)			
Editor Leave-out-Mean R&R				2.73	2.67	3.08
Rate				(0.94)	(0.95)	(0.97)
Number of Observations	15,177	15,177	8,591	15,177	15,177	8,591
R^2 / pseudo R^2	0.27	0.27	0.26	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. The control functions for selection in columns 1, 2, and 3 are calculated using predicted probabilities from column 4, 5, and 6 respectively. We code REStat as having a single editor because we lack information on editors for REStat. Standard errors clustered by editor in parentheses.

Table 6. 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)	(0.09)	(0.09)	(0.14)	(0.07)
2 Publications	0.66	0.50	0.20	0.18	0.10	0.13
	(0.04)	(0.04)	(0.10)	(0.10)	(0.16)	(0.07)
3 Publications	0.88	0.65	0.58	0.53	0.51	0.20
	(0.03)	(0.03)	(0.12)	(0.12)	(0.14)	(0.08)
4-5 Publications	1.11	0.85	0.78	0.72	0.43	0.27
	(0.06)	(0.05)	(0.10)	(0.10)	(0.13)	(0.07)
6+ Publications	1.34	1.01	0.74	0.69	0.55	0.39
	(0.06)	(0.05)	(0.12)	(0.12)	(0.14)	(0.07)
Fractions of Referee Recs.	No	Yes	No	Yes	No	No
Authors	Yes	Yes	Yes	Yes	Yes	Yes
Controls 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	4,235
R^2	0.20	0.26	0.25	0.26	0.26	0.35

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 top-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. Standard errors are clustered by editor in columns 1 to 4, and robust standard errors are used for column 5 and 6.

Table 7. 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 respondents 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

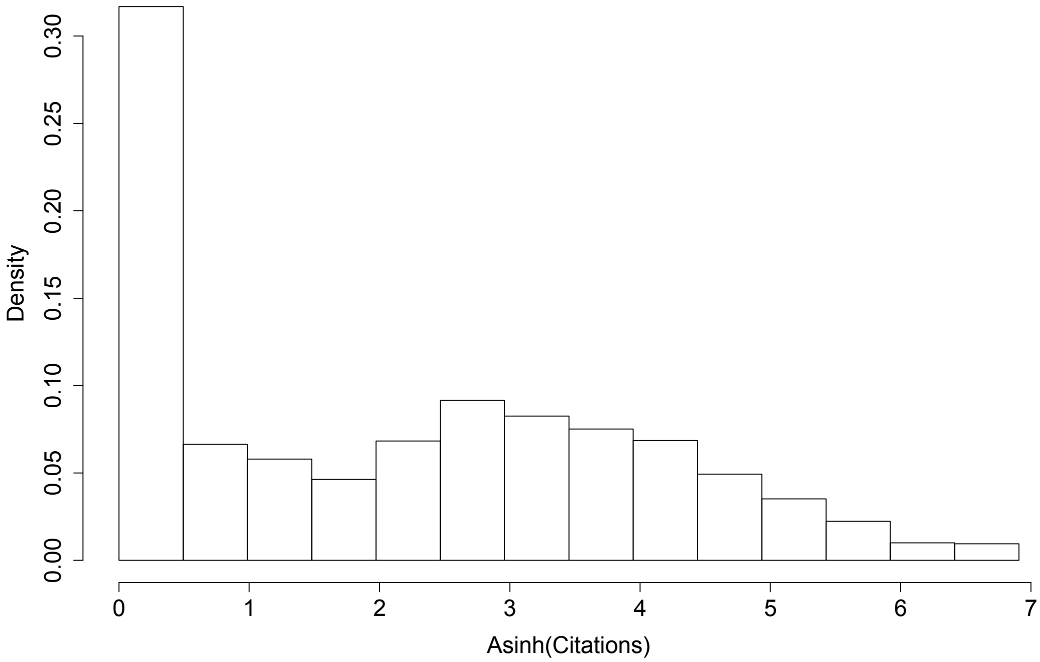
Notes: Table reports regression models fit by OLS in which the dependent variable is the log of the respondent's preferred citation ratio for the paper in a given pair authored by the more prolific author, and the dependent variable is the respondent's relative assessment of the quality of the paper 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.

Table 8. Expert Forecasts About Findings on Editorial Decisions

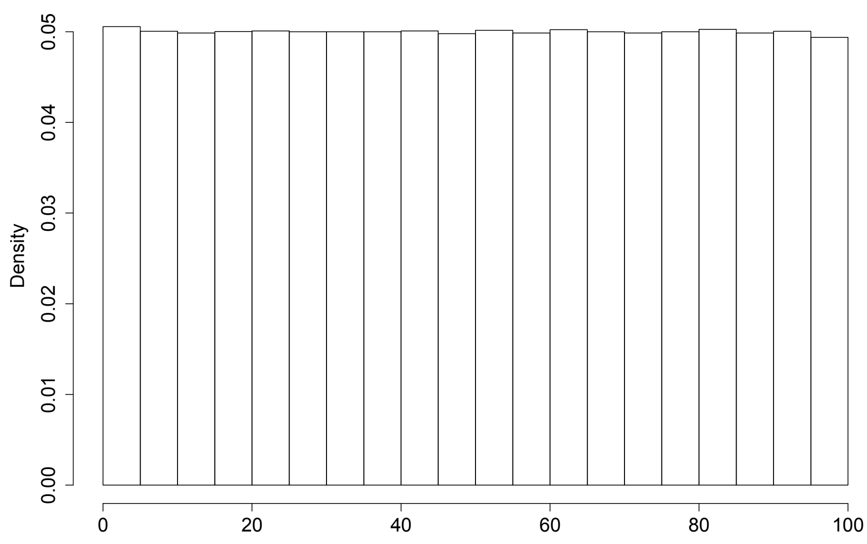
Sample of Experts:	REStud Meeeting		Economics Dept. at Univ. of Zurich		
Question:	Correct Answer (REStud Only)	Average Answer by Editors (N<=12)	Correct Answer (Four Journals)	Average Answer by Faculty (N<=13)	Average Answer by Grad. Stud. (N<=13)
Type 1 - Type 2 errors in R&R / Desk rejects					
What percent of papers with a Revise-and-Resubmit in the first round end up in the top 5 percent of citations (by the Google Scholar measure)?	*18.1%*	32.5%	*22.3%*	19.1%	17.6%
What percent of desk-rejected papers end up in the top 5 percent of citations (by the Google Scholar measure)?	*1.3%*	0.9%	*1.6%*	5.5%	8.7%
Effect of Author Prominence (in Desk Rejects)					
Consider all submissions with at least one “prominent” coauthor that are desk-rejected. What percent of these papers end up in the top 5 percent of citations?	*6.4%*	5.2%	*4.8%*	4.9%	12.7%
Informativeness of Referee Recommendations					
How much higher is the percentile citation if a referee recommendation is positive versus if it is negative (for papers with 3 reports)?	*11.50*	17.50	*10.5*	14.8	18.0
How Much Editors Follow Referee Recommendations					
For papers with three positive referee recommendations, what percent gets an R&R?	*100.0%*	92.5%	*95.9%*	89.7%	92.4%
For papers with two positive referee recommendations and one negative, what percent gets an R&R?	*59.7%*	65.8%	*53.8%*	56.2%	71.5%
For papers with one positive referee recommendations and two negative, what percent gets an R&R?	*6.4%*	21.3%	*5.9%*	14.8%	16.9%
For papers with three negative referee recommendations, what percent gets an R&R?	*0.0%*	2.6%	*0.3%*	3.9%	3.0%

Note: Participants In these surveys received a Qualtrics link to a survey in advance to a presentation by one of the authors of this paper. The first set of forecasts are from a survey of editors and associate editors at the annual board meeting of the *Review of Economic Studies* in Sep. 2015. The second set of forecasts took place before a presentation at the Economics Department in the University of Zurich in the Fall of 2015. The Table reports the exact wording of the questions (not all questions are included in this Table). "Prominent" authors are defined as those with at least 3 publications in the set of 35 journals in the 5 years leading up to (and including) the submission of the paper in question. The questions for the REStud survey pertain to the sample at REStud, whereas the questions for the Zurich survey are based on the entire sample (of all 4 journals pooled). Respondents were not forced to fill in an answer for all questions, so the reported averages are for all available responses to a given question.

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



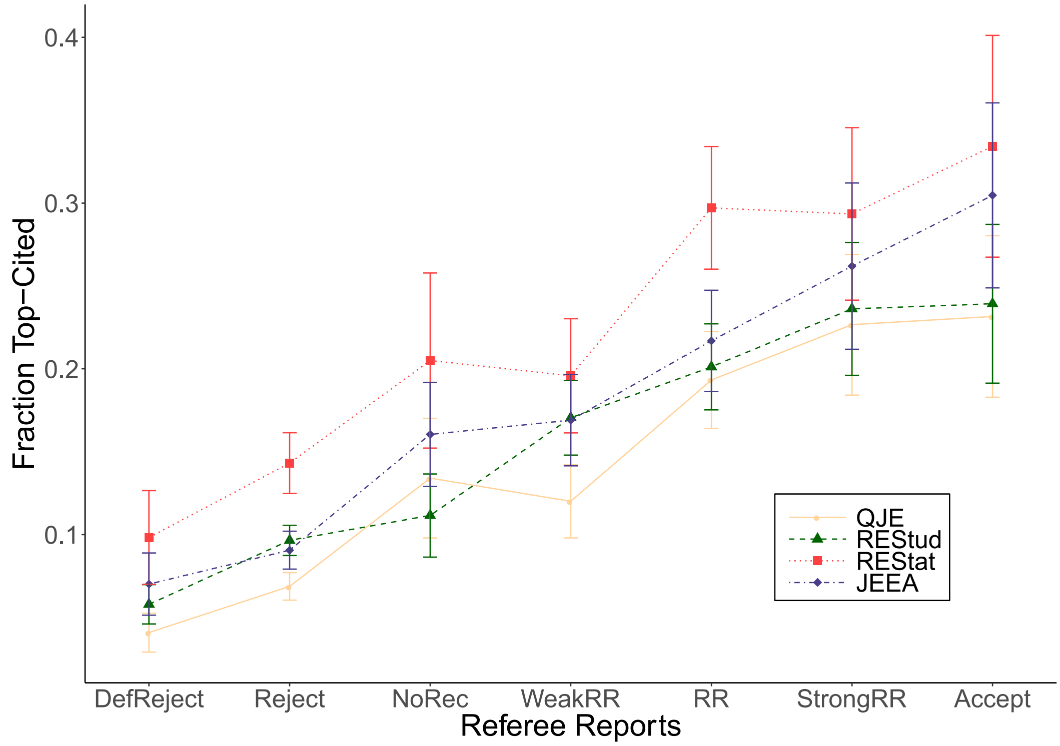
Online Appendix Figure 1b. Distribution of Percentile Citation



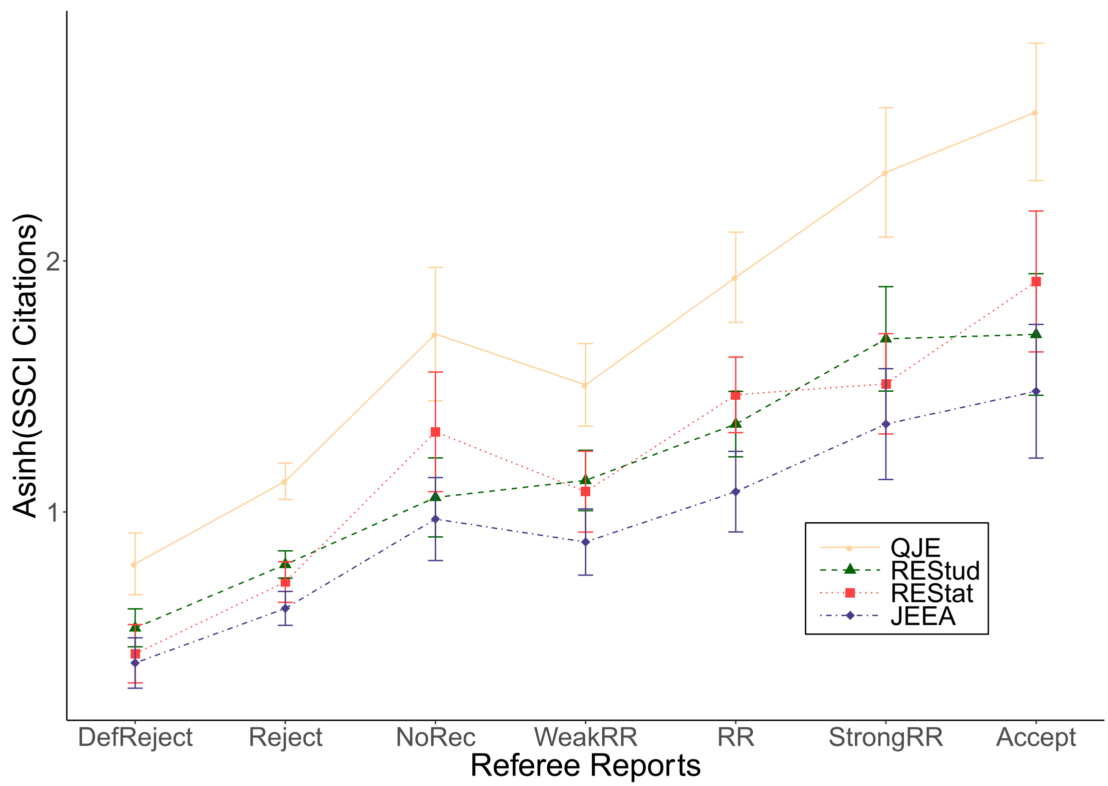
Notes: Online appendix figure 1 shows the distribution of the two main citation measures we use: the percentile citation by journal-year cohort and the Arsinh of Google Scholar citations. Panel 1a shows the distribution for the percentile measure, which is close to uniform given the definition of this 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

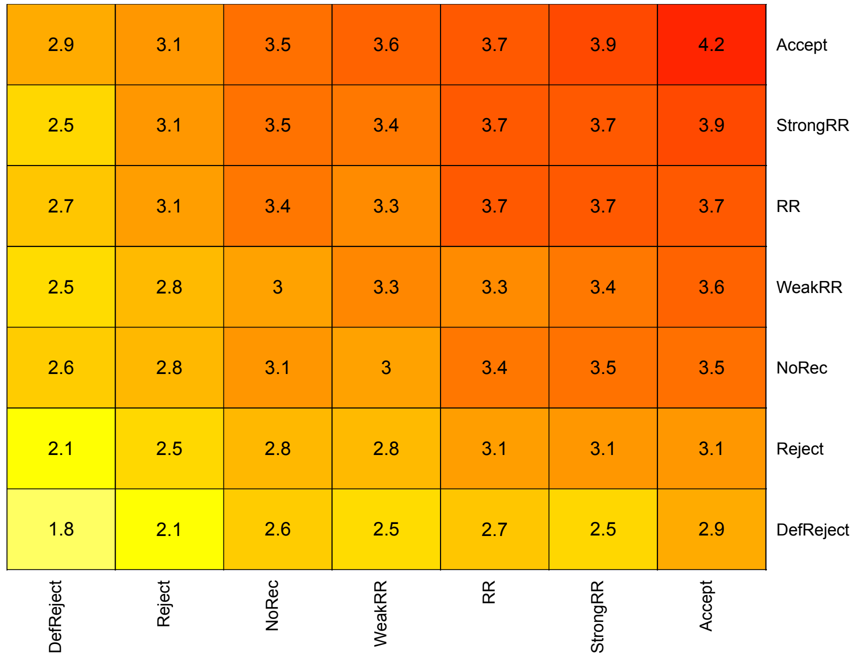
Online Appendix Figure 2a. Robustness using Fraction top-cited



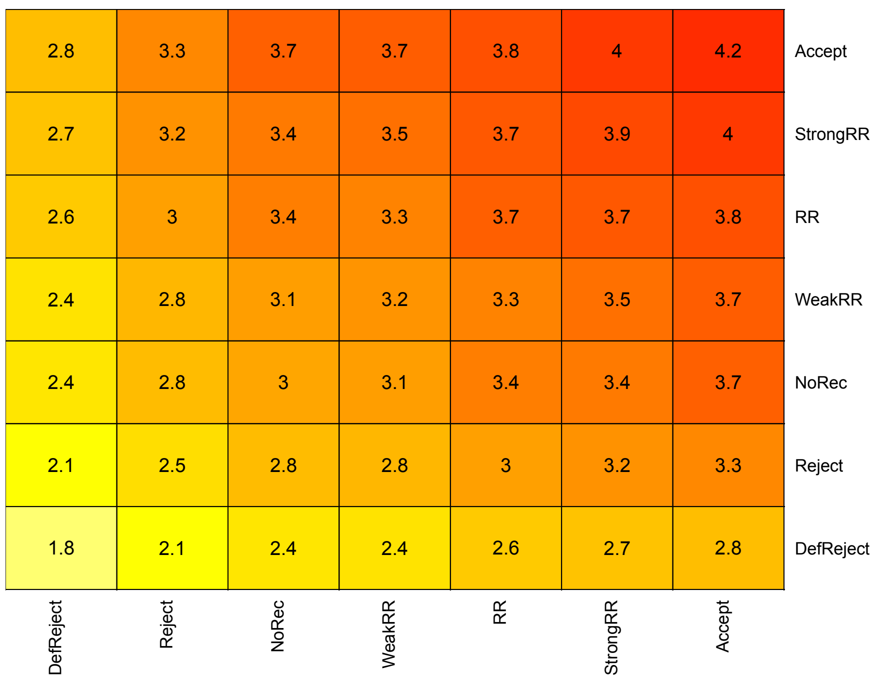
Online Appendix Figure 2b. SSCI Citations, 2006-10



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



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



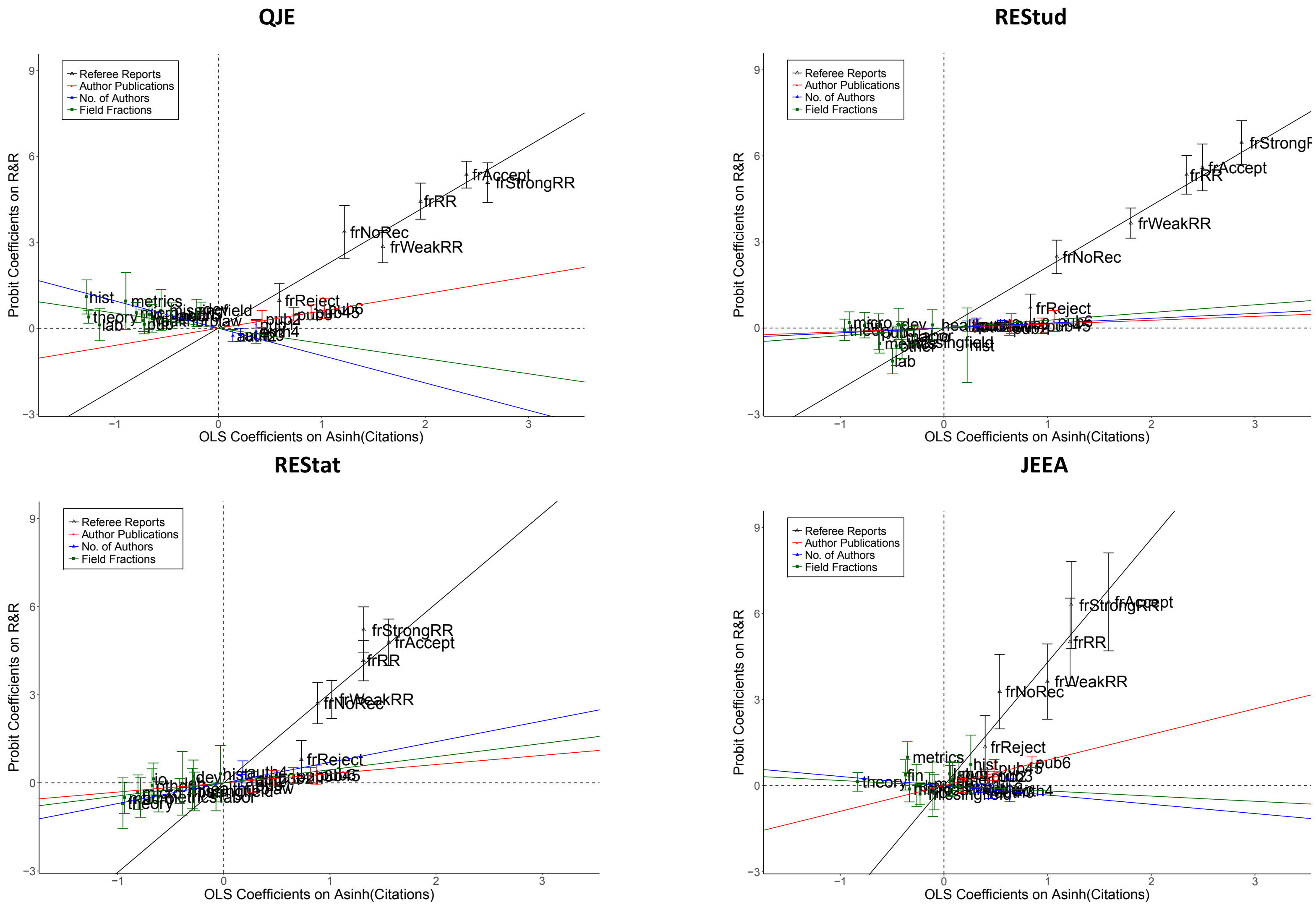
Notes: Online Appendix Figures 2a-d provide robustness checks of the correlation between referee reports and citations displayed visually in Figures 3a-d in the text. Panels a and b plot the weighted average citation measure for a paper receiving a given recommendation. The unit of observation is a referee 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.

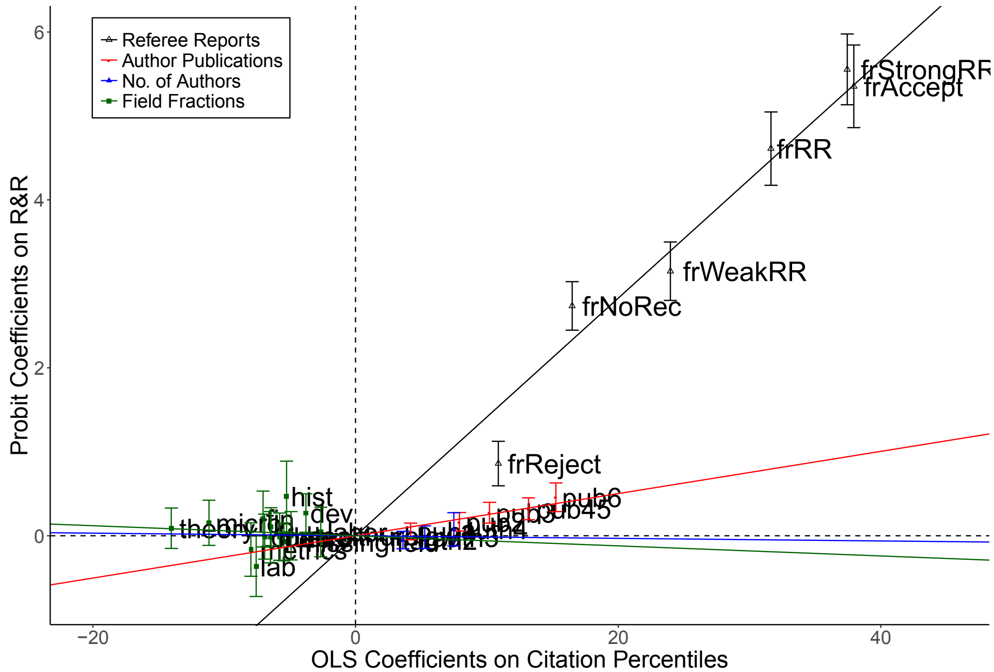
Online Appendix Figure 4. The Relative Effect of Referee Recommendations and Paper Characteristics on Citations and the Probability of Revise and Resubmit – By Journal



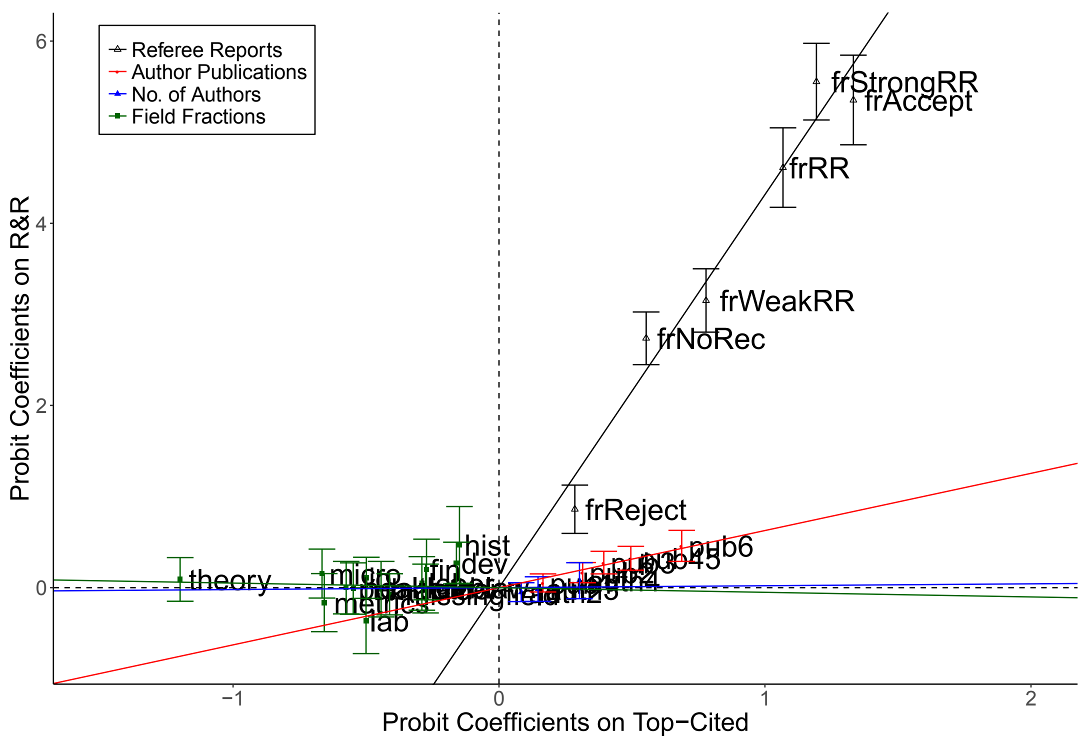
Notes: Online Appendix Figure 4 plots the coefficients from the main specifications of the citation and R&R regressions as in Figure 5b, separately by journal. Best fit lines through each group of coefficients are also shown (weighted by the inverse variance of the probit coefficient from the R&R regression).

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

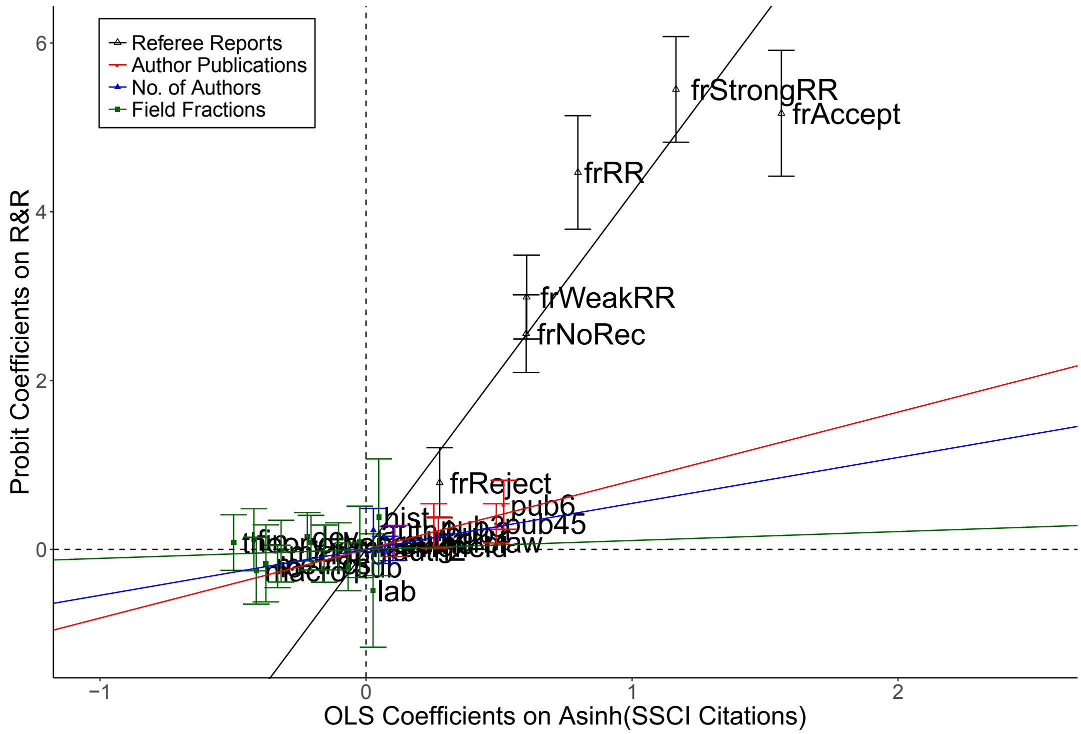
Online Appendix Figure 5a. Using Percentile Citations as Citation measure



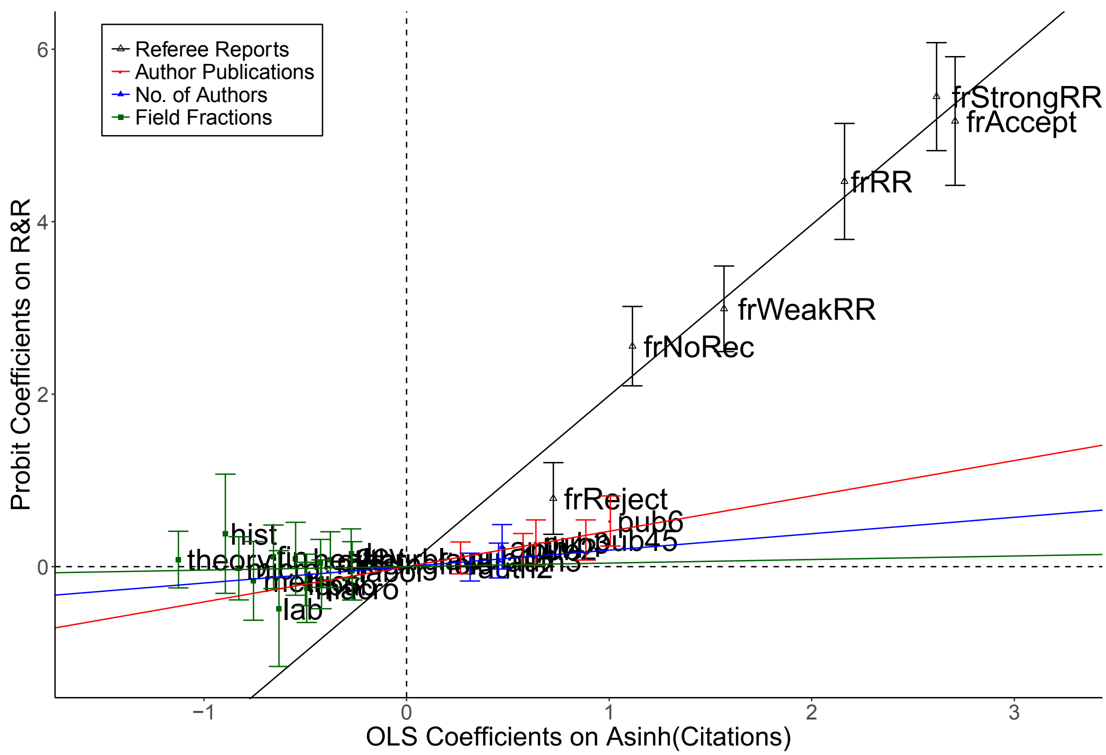
Online Appendix Figure 5b. Using top-cited as Citation measure



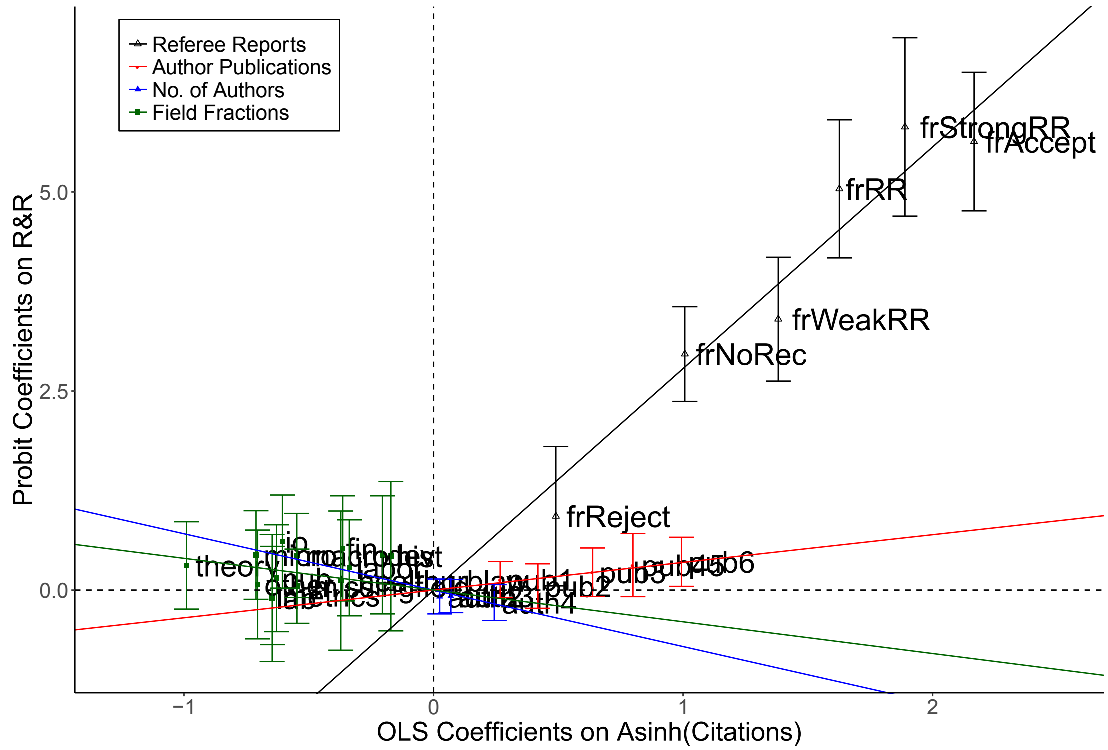
Online Appendix Figure 5c. Using Asinh(SSCI Citations), only years 2006-2010



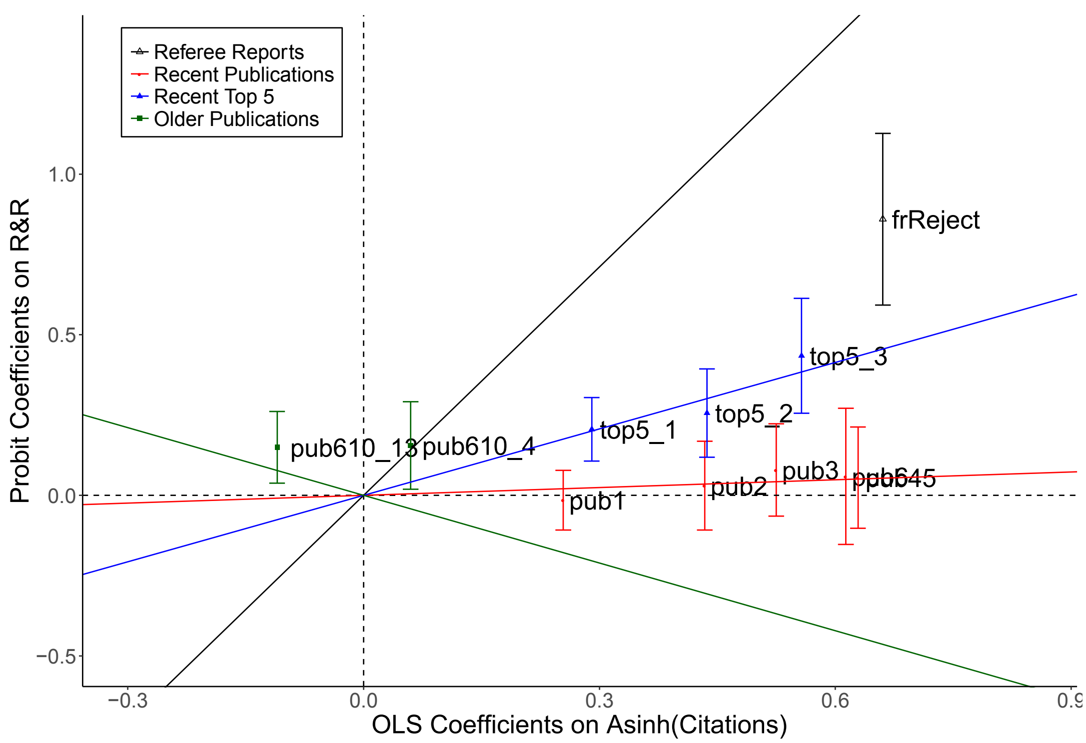
Online Appendix Figure 5d. Using Asinh(Google Scholar Citations), only years 2006-2010



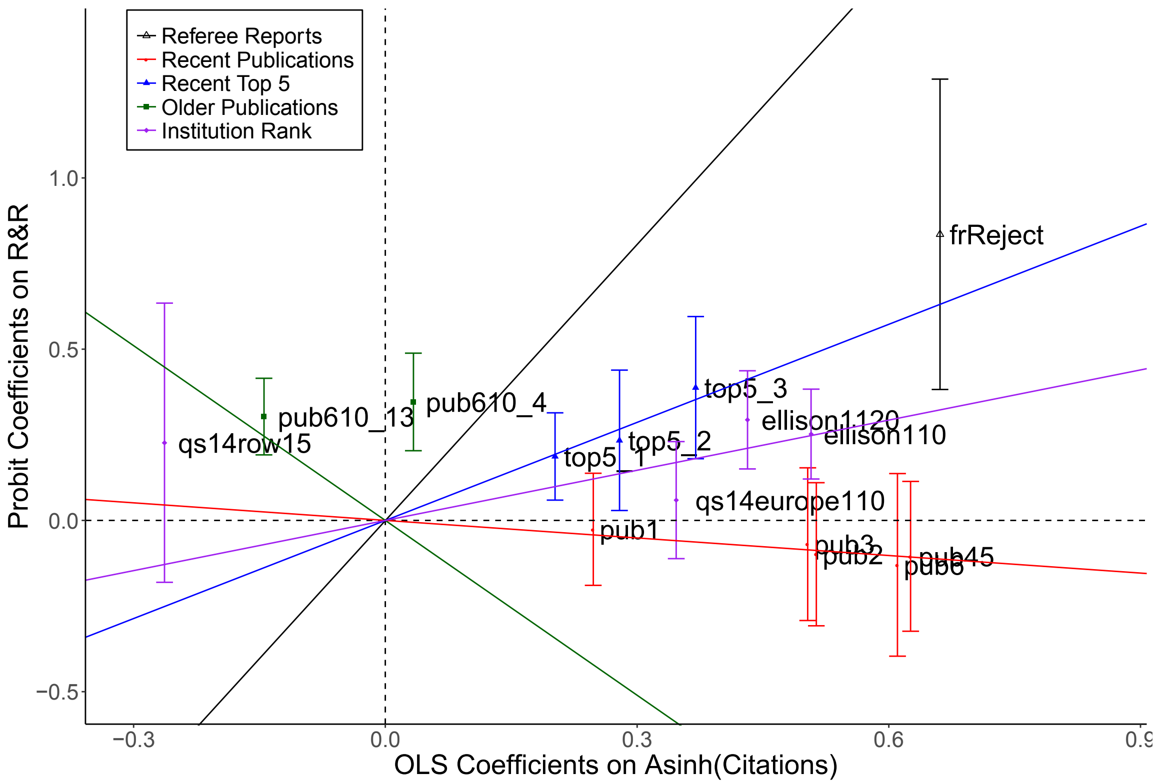
Online Appendix Figure 5e. Using Asinh(Google Scholar Citations), only years 2012-2013



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

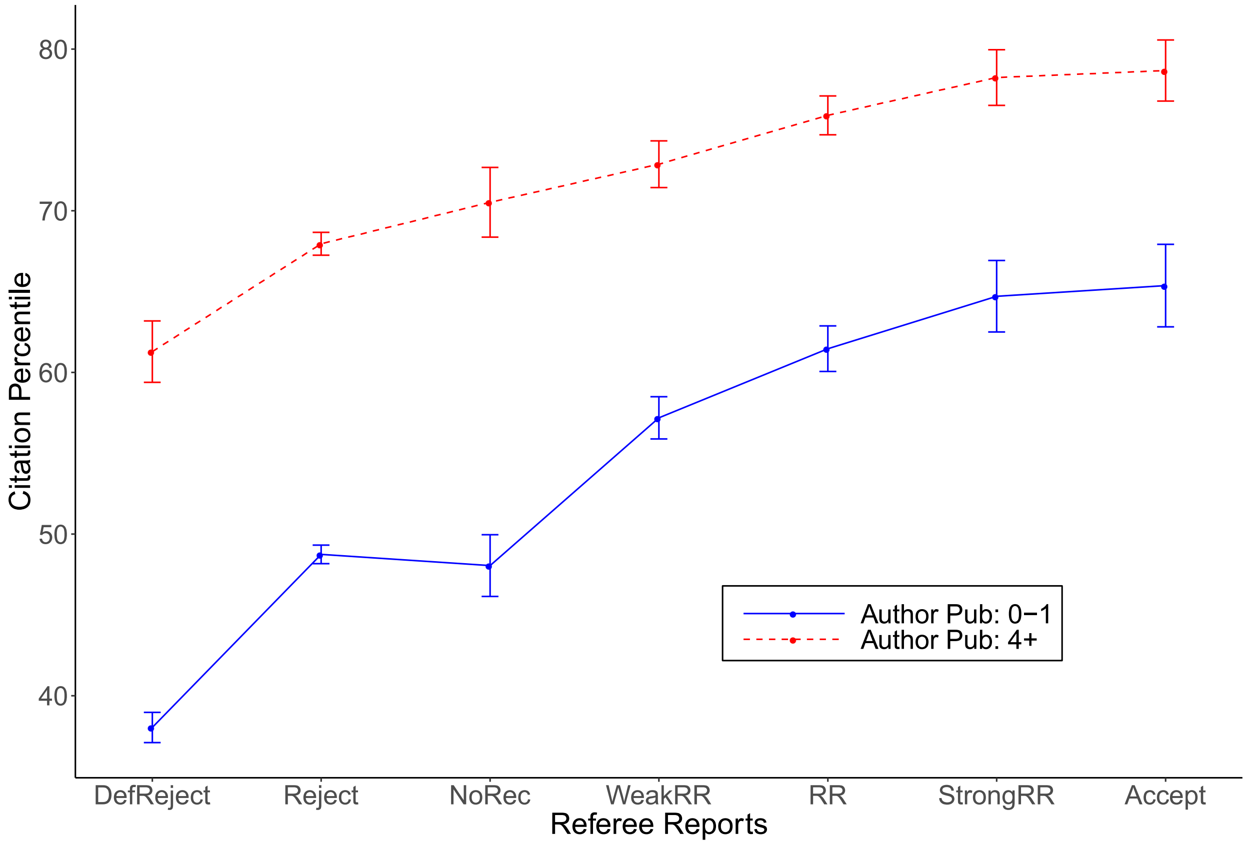


Online Appendix Figure 5g. Including top 5 publications, older publications, and institution rank in regression (JEEA and REStud papers only, not showing coefficients for number of authors and fields)



Notes: Online Appendix Figure 5a-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).

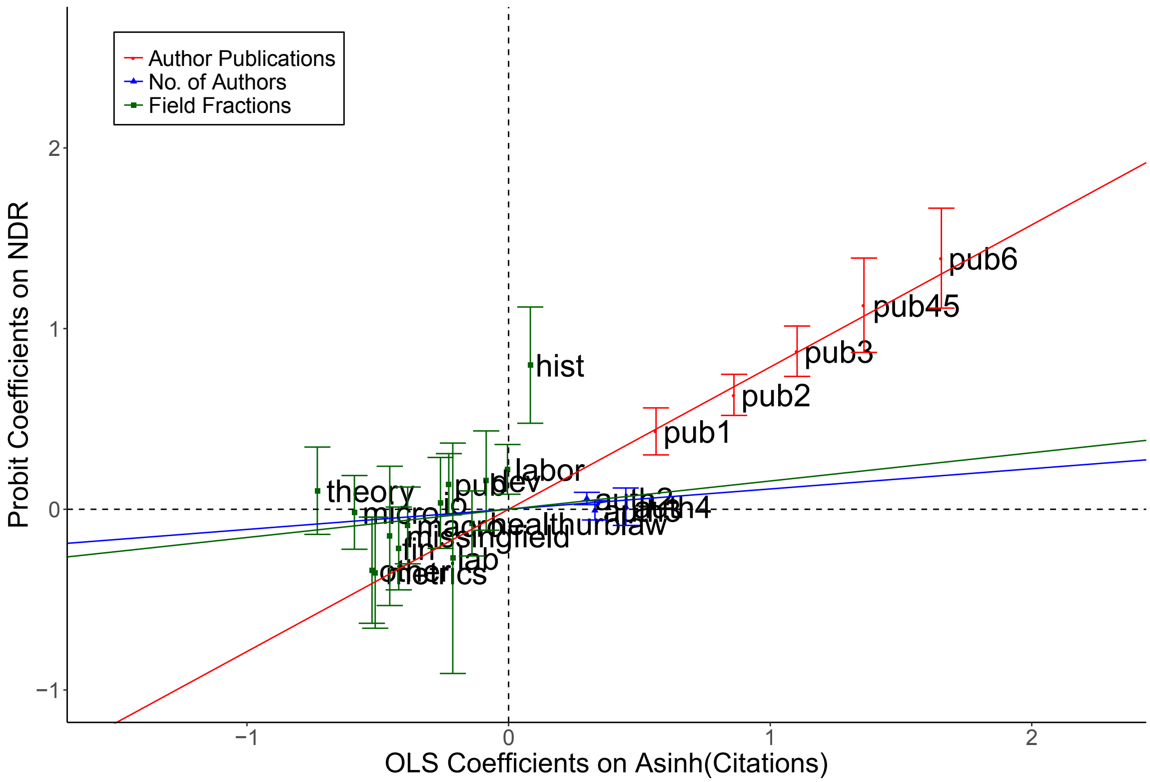
Online Appendix Figure 6. Discounting of Citations of Prolific Authors, Referees (Citation Percentiles)



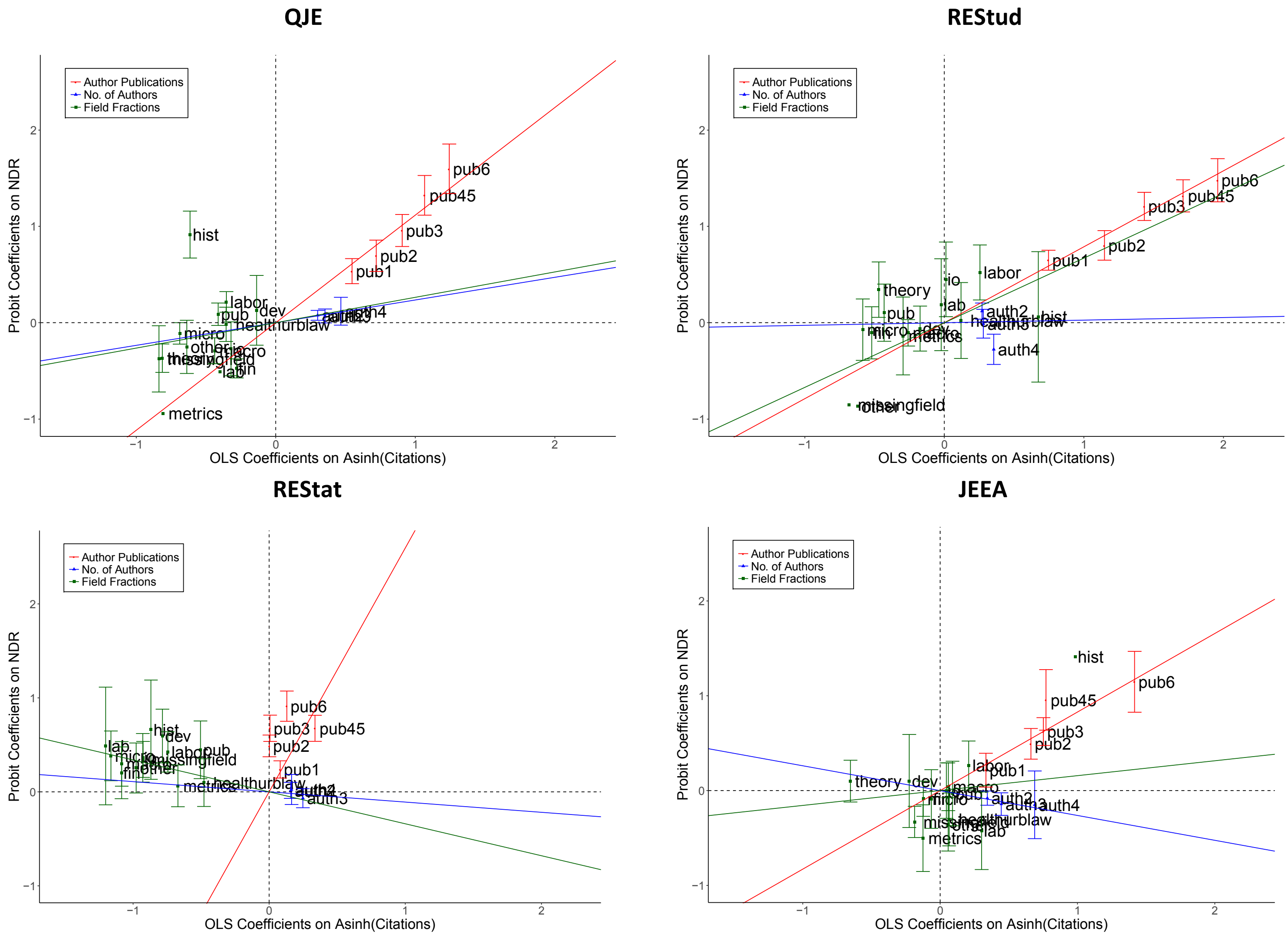
Notes: Online appendix figure 6 checks the robustness of the main figure 8a, using the citation percentile as the measure of citations instead of Asinh(Citations). The key result that conditional on referee recommendation, ex-post citations for prolific authors are higher on average, remains unchanged.

Online Appendix Figure 7. The Relative Effect of Referee Recommendations and Paper Characteristics on Citations and the Probability of Desk Rejection

Online Appendix Figure 7a. Pooled sample

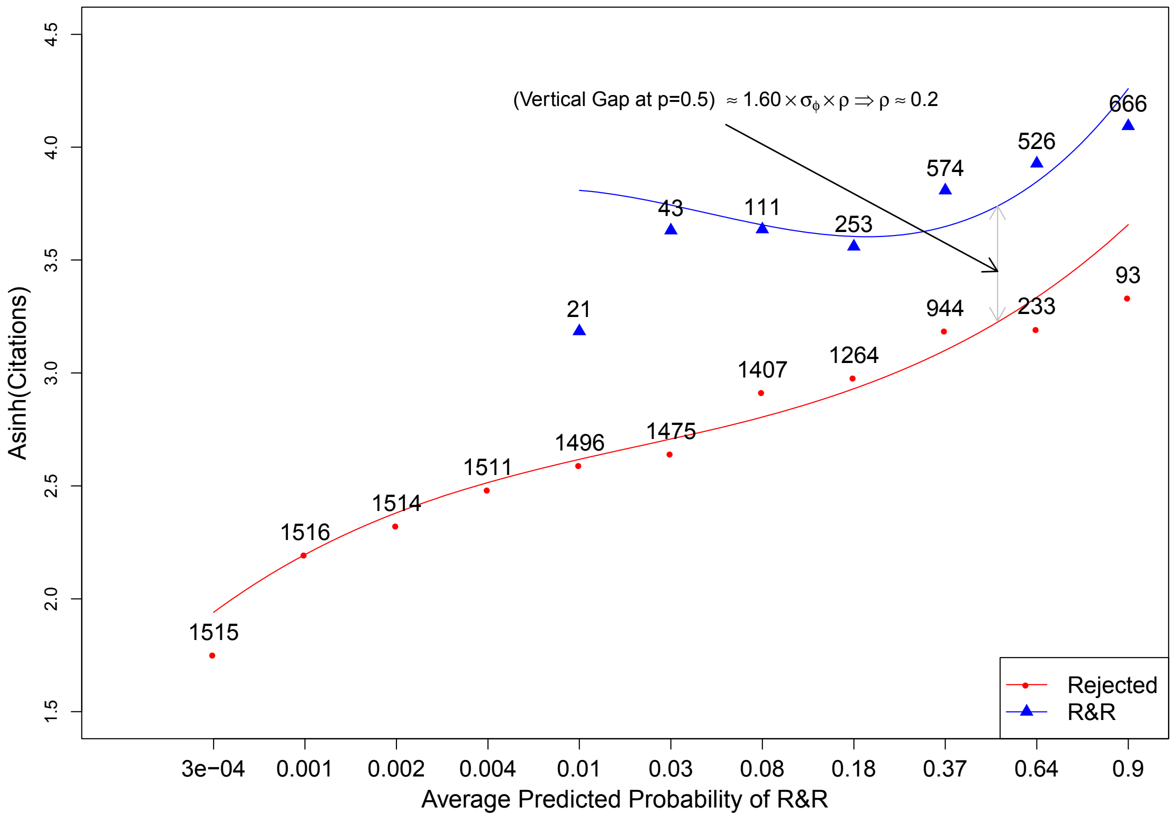


Online Appendix Figure 7b. By Journal



Notes: Online Appendix Figure 7a-b plot the coefficients from the main specifications of the citation and non-desk-rejection regressions (Online Appendix Table 7). 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 8. The Relationship Between the Editor's Revise and Resubmit Decision and Realized Citations – Model Fit



Notes: Online Appendix Figure 8 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 6 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 9. 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	O	O	O	O	O
Exposition (organization, clarity, detail, writing)	O	O	O	O	O

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 9 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 Growth
American Economic Journal: Macroeconomics	Journal of Economic Theory
American Economic Journal: Microeconomics	Journal of Finance
American Economic Journal: Economic Policy	Journal of Financial Economics
American Economic Review	Journal of Health Economics
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
International Journal of Industrial Organization	Journal of Urban Economics
Journal of the European Economic Association	Quarterly Journal of Economics
Journal of Accounting and Economics	The RAND Journal of Economics
Journal of American Statistical Association	Review of Economics and Statistics
Journal of Business and Economic Statistics	Review of Financial Studies
Journal of Development Economics	Review of Economic Studies
Journal of Econometrics	

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
	(0.14)	(0.11)	(0.10)	(0.12)
Lab/Experiments	0.91	0.42	-0.19	-0.45
	(0.26)	(0.19)	(0.12)	(0.20)
Labor	0.64	0.77	-0.18	-0.04
	(0.13)	(0.10)	(0.07)	(0.10)
Health, Urban, Law	0.58	0.59	-0.19	-0.09
	(0.20)	(0.17)	(0.10)	(0.12)
Development	0.71	0.75	0.04	0.18
	(0.17)	(0.16)	(0.12)	(0.12)
History	0.49	0.51	0.53	0.38
	(0.31)	(0.27)	(0.14)	(0.21)
Public	0.54	0.56	-0.04	-0.10
	(0.15)	(0.12)	(0.13)	(0.12)
Industrial Organization	0.46	0.53	-0.12	0.02
	(0.15)	(0.13)	(0.11)	(0.13)
Finance	0.51	0.52	-0.06	0.11
	(0.15)	(0.14)	(0.09)	(0.11)
Macro	0.59	0.56	-0.02	-0.11
	(0.13)	(0.10)	(0.09)	(0.09)
Field Missing	0.59	0.58	-0.10	-0.16
	(0.12)	(0.09)	(0.10)	(0.11)
Micro	0.30	0.23	0.00	0.06
	(0.11)	(0.11)	(0.08)	(0.11)
Unclassified	0.45	0.55	-0.25	-0.10
	(0.12)	(0.10)	(0.11)	(0.13)
Econometrics	0.45	0.41	-0.16	-0.25
	(0.11)	(0.08)	(0.08)	(0.12)
Control Function	No	Yes	-	-
R&R Indicator	No	Yes	-	-
Editor's leave-one-out R&R rate	-	-	No	Yes
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	0.27	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. The control function for selection in column 2 is calculated using predicted probabilities based on column 4. We code REStat as having a single editor because we lack information on editors for REStat. Standard errors clustered by editor 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	0.73	0.40	0.97	0.70	0.80	1.35
	(0.09)	(0.11)	(0.15)	(0.09)	(0.28)	(0.23)	(0.32)	(0.54)
No Recommendation	1.22	1.09	0.90	0.54	3.36	2.48	2.72	3.28
	(0.22)	(0.18)	(0.24)	(0.12)	(0.46)	(0.29)	(0.35)	(0.64)
Weak R&R	1.59	1.80	1.03	1.00	2.85	3.66	2.84	3.63
	(0.06)	(0.17)	(0.19)	(0.15)	(0.28)	(0.25)	(0.32)	(0.65)
R&R	1.96	2.34	1.35	1.22	4.43	5.34	4.16	5.01
	(0.15)	(0.22)	(0.18)	(0.13)	(0.31)	(0.32)	(0.33)	(0.75)
Strong R&R	2.60	2.87	1.36	1.23	5.09	6.46	5.21	6.30
	(0.31)	(0.28)	(0.23)	(0.32)	(0.32)	(0.37)	(0.38)	(0.75)
Accept	2.40	2.49	1.60	1.59	5.36	5.60	4.79	6.41
	(0.11)	(0.23)	(0.23)	(0.55)	(0.22)	(0.38)	(0.38)	(0.85)
Author Publications in 35 high-impact journals								
Publications: 1	0.37	0.30	0.27	0.17	0.14	0.09	-0.07	0.00
	(0.08)	(0.08)	(0.09)	(0.08)	(0.06)	(0.11)	(0.12)	(0.12)
Publications: 2	0.42	0.63	0.47	0.42	0.33	-0.01	0.18	0.32
	(0.05)	(0.06)	(0.11)	(0.07)	(0.14)	(0.13)	(0.12)	(0.12)
Publications: 3	0.73	0.65	0.66	0.48	0.49	0.16	0.25	0.34
	(0.05)	(0.04)	(0.12)	(0.09)	(0.11)	(0.17)	(0.14)	(0.08)
Publications: 4-5	0.90	0.95	0.82	0.48	0.56	0.08	0.26	0.61
	(0.12)	(0.08)	(0.12)	(0.08)	(0.12)	(0.14)	(0.13)	(0.15)
Publications: 6+	1.03	1.06	0.87	0.85	0.69	0.24	0.28	0.82
	(0.12)	(0.09)	(0.13)	(0.13)	(0.17)	(0.17)	(0.15)	(0.09)
R&R Indicator	0.09	-0.25	0.65	0.78				
(Mechanical Publ. Effect)	(0.24)	(0.15)	(0.10)	(0.29)				
Control Function for Selection	0.41	0.45		-0.14				
(Value Added of the Editor)	(0.13)	(0.07)		(0.17)				
Editor Leave-out-Mean R&R					1.69	4.34		1.07
Rate					(4.22)	(1.53)		(1.50)
Authors	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls 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	0.28	0.25	0.27	0.22				
Pseudo-R^2					0.45	0.52	0.48	0.51

Notes: The sample for each journal includes all non-desk-rejected papers with at least two referees assigned. The control functions for selection in column 1, 2, 3, and 4 are calculated using predicted probabilities from columns 5, 6, 7, and 8 respectively. The specification in column 7 excludes the editors' leave-out-mean R&R rate because we do not have information on editors for REStat. We code REStat as having a single editor because we lack information on editors for REStat. Standard errors clustered by editor in parentheses, except for REStat in columns 3 and 7, where we use robust standard errors.

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)	(4)	(5)	(6)
Fractions of Referee Recommendations						
Reject	0.66	0.69	0.54	0.50	1.15	1.58
	(0.08)	(0.12)	(0.22)	(0.19)	(0.27)	(0.42)
No Recommendation	0.74	0.95	1.84	2.31	3.18	4.11
	(0.13)	(0.15)	(0.38)	(0.25)	(0.29)	(0.64)
Weak R&R	1.30	1.39	1.37	2.83	3.61	4.21
	(0.13)	(0.14)	(0.33)	(0.24)	(0.30)	(0.52)
R&R	1.41	1.77	2.16	3.98	5.37	5.75
	(0.15)	(0.19)	(0.36)	(0.24)	(0.30)	(0.63)
Strong R&R	1.70	2.01	2.95	5.11	6.17	6.41
	(0.23)	(0.23)	(0.56)	(0.27)	(0.35)	(0.65)
Accept	1.88	1.92	2.88	4.94	5.89	6.64
	(0.26)	(0.26)	(0.60)	(0.30)	(0.28)	(0.76)
Author Publications in 35 high-impact journals						
Publications: 1	0.31	0.23	0.14	0.03	0.03	-0.02
	(0.06)	(0.06)	(0.12)	(0.08)	(0.06)	(0.15)
Publications: 2	0.51	0.44	0.35	0.17	0.16	-0.08
	(0.05)	(0.06)	(0.12)	(0.09)	(0.08)	(0.13)
Publications: 3	0.60	0.59	0.62	0.32	0.12	0.28
	(0.06)	(0.06)	(0.13)	(0.07)	(0.10)	(0.11)
Publications: 4-5	0.88	0.70	0.80	0.42	0.19	0.32
	(0.08)	(0.05)	(0.15)	(0.10)	(0.09)	(0.13)
Publications: 6+	0.96	0.95	0.82	0.61	0.29	0.32
	(0.09)	(0.07)	(0.12)	(0.15)	(0.11)	(0.17)
R&R Indicator	0.50	0.21	-0.30			
(Mechanical Publ. Effect)	(0.18)	(0.18)	(0.26)			
Control Function for Selection	0.13	0.17	0.49			
(Value Added of the Editor)	(0.12)	(0.11)	(0.13)			
Editor Leave-out-Mean R&R				3.15	3.48	1.19
Rate				(1.31)	(1.07)	(1.89)
Indicators for Number of Authors	Yes	Yes	Yes	Yes	Yes	Yes
Controls 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	1,795	7,238	5,414	1,795
R-squared	0.24	0.28	0.34			
Pseudo-R^2				0.57	0.48	0.37

Notes: The samples include e all non-desk-rejected paper with at least two referees assigned with that received 2, 3 and at least 4 referee reports respectively. The control functions for selection in columns 1, 2, and 3 are calculated using predicted probabilities from columns 4, 5, and 6 respectively. We code REStat as having a single editor because we lack information on editors for REStat. Standard errors clustered by editor in parentheses.

Online App. Table 5. Predictors of Citations and R&R, 2006-10 vs. 2012-13

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	0.93
	(0.08)	(0.08)	(0.21)	(0.42)
No Recommendation	1.11	1.01	2.56	2.96
	(0.11)	(0.24)	(0.23)	(0.29)
Weak R&R	1.57	1.38	2.99	3.40
	(0.11)	(0.17)	(0.25)	(0.38)
R&R	2.16	1.63	4.47	5.04
	(0.13)	(0.24)	(0.33)	(0.42)
Strong R&R	2.62	1.89	5.45	5.82
	(0.28)	(0.28)	(0.31)	(0.54)
Accept	2.71	2.17	5.17	5.63
	(0.16)	(0.36)	(0.36)	(0.43)
Author Publications in 35 high-impact journals				
Publications: 1	0.27	0.27	0.10	0.13
	(0.05)	(0.08)	(0.09)	(0.11)
Publications: 2	0.58	0.42	0.20	0.05
	(0.04)	(0.06)	(0.09)	(0.14)
Publications: 3	0.64	0.64	0.28	0.22
	(0.03)	(0.10)	(0.13)	(0.15)
Publications: 4-5	0.88	0.80	0.31	0.31
	(0.06)	(0.09)	(0.12)	(0.20)
Publications: 6+	1.01	0.99	0.53	0.36
	(0.08)	(0.09)	(0.14)	(0.15)
R&R Indicator	-0.09	-0.04		
(Mech. Publ. Effect)	(0.18)	(0.24)		
Control Function	0.51	0.22		
(Ed. Value Added)	(0.10)	(0.16)		
Editor Leave-out-Mean R&R			3.32	2.20
Rate			(1.68)	(1.15)
Indicators for Number of Authors	Yes	Yes	Yes	Yes
Controls 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	0.25	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. The control function for selection in columns 1 and 2 are calculated using predicted probabilites from column 3 and 4. We code REStat as having a single editor because we lack information on editors for REStat. Standard errors clustered by editor in parentheses.

Online Appendix Table 6. Predictors of Citations and Desk Rejection

Specification:	**OLS Regression**			**Probit**
Dependent Variable:	**Asinh of Citations**			**Indicator for Paper Not Desk Rejected**
	(1)	(2)	(3)	(4)
Author Publications in 35 high-impact journals				
Publications: 1	0.56	0.62	0.50	0.43
	(0.06)	(0.05)	(0.05)	(0.07)
Publications: 2	0.86	0.94	0.77	0.63
	(0.08)	(0.05)	(0.05)	(0.06)
Publications: 3	1.10	1.21	0.98	0.87
	(0.08)	(0.06)	(0.05)	(0.07)
Publications: 4-5	1.36	1.49	1.20	1.13
	(0.12)	(0.06)	(0.05)	(0.13)
Publications: 6+	1.65	1.81	1.47	1.39
	(0.14)	(0.07)	(0.06)	(0.14)
Number of Authors				
2 authors	0.30	0.30	0.29	0.06
	(0.03)	(0.03)	(0.03)	(0.02)
3 authors	0.33	0.33	0.33	-0.01
	(0.04)	(0.04)	(0.04)	(0.03)
4+ authors	0.45	0.45	0.44	0.01
	(0.07)	(0.07)	(0.07)	(0.05)
NDR Indicator	0.40		0.86	
(Mechanical Publ. Effect)	(0.33)		(0.07)	
Control Function for Selection into NDR	0.29	0.52		
(Value Added of the Editor)	(0.17)	(0.04)		
Editor Leave-out-Mean R&R				3.27
Rate				(0.31)
Controls for Field of Paper	Yes	Yes	Yes	Yes
Indicators for Journal-Year Cohort	Yes	Yes	Yes	Yes
Number of Observations	29,868	29,868	29,868	29,868
R-squared	0.27	0.27	0.27	
Pseudo-R^2				0.23

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. The control function for selection in columns 1 and 2 is calculated using predicted probabilities from column 4. We code REStat as having a single editor because we lack information on editors for REStat. Standard errors clustered by editor in parentheses.

Online Appendix Table 7. 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		0.78
		(0.05)		(0.13)
No Recommendation		0.99		2.65
		(0.11)		(0.15)
Weak R&R		1.30		3.07
		(0.11)		(0.18)
R&R		1.63		4.52
		(0.13)		(0.22)
Strong R&R		2.00		5.46
		(0.23)		(0.21)
Accept		1.93		5.23
		(0.18)		(0.25)
Author Publications in 35 high-impact journals				
Publications: 1	0.40	0.21	0.36	0.03
	(0.05)	(0.04)	(0.05)	(0.05)
Publications: 2	0.65	0.42	0.54	0.10
	(0.06)	(0.04)	(0.04)	(0.06)
Publications: 3	0.84	0.52	0.79	0.25
	(0.07)	(0.04)	(0.06)	(0.06)
Publications: 4-5	1.03	0.68	1.04	0.30
	(0.10)	(0.04)	(0.12)	(0.07)
Publications: 6+	1.37	0.92	1.29	0.41
	(0.12)	(0.05)	(0.13)	(0.08)
Control Function for Selection into NDR	0.20			
(Value Added of the Editor)	(0.17)			
Control Function for Selection into R&R		0.25		
(Value Added of the Editor)		(0.07)		
NDR Indicator	0.42			
(Mechanical Publ. Effect)	(0.33)			
R&R Indicator		0.16		
(Mechanical Publ. Effect)		(0.12)		
Editor Leave-out-Mean NDR			3.35	
Rate			(0.31)	
Editor Leave-out-Mean R&R				2.54
Rate				(0.91)
Indicators for Number of Authors	Yes	Yes	Yes	Yes
Controls 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	0.26	0.29		
Pseudo-R^2			0.22	0.49

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 citations. The control function for selection in columns 1 and 2 are calculated using predicted probabilities based on columns 3 and 4. We code REStat as having a single editor because we lack information on editors for REStat. Standard errors clustered by editor in parentheses.