Are Referees and Editors in Economics Gender Neutral?*

David Card
UC Berkeley and NBER

Stefano DellaVigna
UC Berkeley and NBER

Patricia Funk
Università della Svizzera Italiana

Nagore Iriberri
University of the Basque Country and Ikerbasque

October 2018

Abstract

We study the effects of gender on the evaluation of economic research using submissions to four leading journals. We begin by comparing the relative recommendations of female and male referees in assessing female- versus male-authored papers. The difference-in-differences is precisely estimated and close to zero, suggesting that the *differential* biases of male referees are negligible. To assess whether referees as a whole impose a different standard for female authors, however, we have to account for quality differences across papers. Building on a pre-specified design, we compare citations for female and male-authored papers, holding constant the referee evaluations and other characteristics, including prior publications of the authors. We find that female-authored papers receive about 25% more citations than observably similar male-authored papers. A survey of professional economists suggests that this result is unlikely to be driven by excess citations for female-authored papers. Rather, referees of both genders appear to set a higher bar for female-authored papers. Editors largely follow the referees, resulting in a 7 percentage points lower probability of a revise and resubmit verdict for female-authored papers, relative to a citation-maximizing benchmark. In their desk rejection decisions editors treat female authors more favorably, though they still impose a higher bar than would be expected under citation-maximization. The recommendations of male and female referees are equally informative about future citations, and editors weigh the recommendations of both groups equally. We also find no significant differences in the time that referees or editors take in reaching decisions about female-authored versus male-authored papers. We discuss the implications of our findings for the profession.

*We thank the editors and staff of the *Journal of the European Economics Association*, the *Quarterly Journal of Economics*, the *Review of Economics and Statistics*, and the *Review of Economic Studies* for their invaluable assistance and support. We also thank Lawrence Katz, Andrei Shleifer, Lise Versterlund, and Heidi Williams for helpful comments. We are also grateful to Luisa Cefala', Alden Cheng, Johannes Hermle, Andrew Tai, and a team of undergraduate research assistants for their extraordinary help. Nagore Iriberri acknowledges financial support from grants ECO2015-66027-P MINECO/FEDER and IT869-13. Patricia Funk acknowledges financial support from the Swiss National Science Foundation (grant 178887). The analysis plan is posted on the AEA registry under AEARCTR-0003048.

1 Introduction

Women are under-represented in the top ranks of many professions, including corporate management (Bertrand and Hallock, 2001), law (Azmat and Ferrer, 2017), and academic research (e.g., Ceci et al., 2014). While numerous explanations have been offered for this gap, including differences in competitiveness (Niederle and Vesterlund, 2011; Reuben et al., 2015) and different choices in the allocation of time between work and family (Goldin, 2014), an abiding concern is that stereotyping biases (Reuben et al., 2014; Bordalo et al., forthcoming; Bohren et al., 2018) or other forms of discrimination lead decision makers to *under-value* the contributions of women. This concern is particularly salient in economics, where the vast majority of gatekeepers – senior faculty, journal editors, and referees – are male (Ginther and Kahn, 2004; Bayer and Rouse, 2016; Lundberg, 2017).

Existing evidence on the presence of gender biases in the evaluation of economic research is mixed. Blank (1990) conducted a randomized experiment in which submissions to the *American Economic Review* were assigned to referees with or without masking the author's name and affiliation. She found no significant difference in acceptance rates of female-authored papers under the masked or unmasked conditions. Broder (1993) studied the reviews of National Science Foundation proposals, finding that female reviewers rated female-authored proposals lower than those of males. Abrevaya and Hamermesh (2012) find no significant differences between the female and male referees at an anonymous journal in their relative assessment of male- versus female-authored papers. Chari and Goldsmith-Pinkham (2017) find that acceptance rates of female-authored submissions to NBER conferences are similar to those of males. Hengel (2017), however, argues that female authors are held to a higher standard in the peer review process and are forced to make more revisions before their work is published. Focusing on the general climate in economics, Wu (2018) documents that online discussions of female economists often gravitate toward personal characteristics and away from professionally-oriented accomplishments. Nevertheless, Donald and Hamermesh (2006) conclude that the mostly male members of the American Economics Association exhibit a *positive* preference for female candidates to serve on the Association's executive board.

In this paper we assess the role of gender in the evaluation process, using data on nearly 30,000 submissions to four leading economics journals: the *Journal of the European Economics Association*, the *Quarterly Journal of Economics*, the *Review of Economics and Statistics*, and the *Review of Economic Studies*. We use a combination of name-based algorithms and individual look-ups to assign gender to the co-authors and referees of each paper.[1] We combine the observed characteristics of each submission – including the previous publication record of the authors – with the summary recommendations of the referees, the decision of the editor, and ultimate citations received by the paper, regardless of whether it was accepted or not.[2] We use these data to analyze gender differences in how papers are assigned to referees, how they are reviewed, and how editors use the inputs from referees to reach a revise and resubmit (R&R) verdict. We also consider whether desk-

[1] Nearly all the editors in our 2003-2013 sample period were male, making it very difficult to conduct an analysis of editor's gender without revealing individual identities, so we leave the study of this potentially important topic to future work. The evidence in Bransch and Kvasnick (2017) suggests that having female editors does not appear to increase the share of female authored papers published in top journals.

[2] We do not have access any textual information in the referee report, the editorial letter, or the paper itself.

rejection decisions and the delays imposed by referees and editors depend on the gender of the author team. Our analysis largely follows the analysis plan AEARCTR-0003048, which we drafted prior to the completion of our data collection to address concerns over data mining and p-hacking (see Christensen and Miguel, 2018).

We complement our database of journal submissions with a survey of 141 economists covering the main topics of our analysis. This allows us to compare, both qualitatively and quantitatively, our results with the expectations of the survey population, as in DellaVigna and Pope (2018). We also elicit quantitative beliefs about the link between citations and quality, which we use to interpret the relationships between referee recommendations and realized citations.

We begin in Section 2 with a brief summary of the procedures we have developed to assemble a data base of genders for over 50,000 economists. This procedure allows us to classify the gender of over 95% of authors and referees with an error rate of under 1 percent.[3]

In Section 3 we present a descriptive overview of the submissions data. Our database builds on the sample originally collected by Card and DellaVigna (forthcoming) – hereafter, CDV – adding three new pieces of information: the gender of each author; the publication record of each author in the years before submission[4]; and the gender of each referee.[5] Two-thirds of submissions were written by a single male or all-male team of coauthors, 8 percent were written by females, and 19 percent were written by mixed-gender teams. We use the publication records of each co-author to classify mixed-gender teams by whether the "senior" (i.e., most-published) co-author is female (3 percent of all submissions) or not (16 percent), thus yielding four gender-mix categories for authors. Similarly, we assign gender and previous publication records to referees.

As a first step in our analysis, in Section 4 we analyze the matching process by which papers are assigned to referees. A major factor is the field of the paper. Consistent with earlier studies (e.g., Dolado et al., 2012; Lundberg, 2017; Chari and Goldsmith-Pinkham, 2017), we find that the share of female authors is lower in some fields (e.g., macro) and higher in others (e.g., labor economics), with parallel shifts in the fraction of female referees assigned to papers in that field. Even controlling for subfield differences in the fraction of female authors, however, editors are 7 percentage points (50 percent) more likely to assign a female-authored paper to a female referee. This gap is only slightly attenuated for mixed-gender papers with a senior female author. Interestingly, our survey respondents are largely unaware of the degree of gender matching in the assignment of papers.

Editors appear to pay attention to gender in assigning referees. To what extent do the referee recommendations vary by gender? In Section 5 we begin with a simple audit-style analysis, comparing the assessments of female and male referees. Our most general models include paper fixed effects, allowing us to isolate the differential assessments of *the same paper* by referees of different genders. We find a precisely estimated null effect: the difference-in-differences of female versus male referees in assessing female- versus male-authored papers is zero, consistent with earlier findings of Abrevaya and Hamermesh (2012) for a single journal. In contrast, our survey respondents predict a

[3] We cannot assign gender at a high level of accuracy for about 3% of authors. We therefore include a category for papers with a coauthor of unknown gender. The rate of missing information on the gender of referees is lower (1%).

[4] CDV only collected the publication record of the co-author with the most previous publications.

[5] We also gathered more granular information on waiting times in the review process, information on the gender composition of the sub-field of the paper, and information on the complexity of the abstract.

2 percentage point difference-in-difference in the probability of giving an R&R recommendation.

While these simple comparisons rule out any large relative bias in the assessmentsof female-authored papers, they do *not* allow us to conclude that female authors face the same bar as males. It is possible that both male and female referees are biased for, or against, female-authored papers. To make further progress we need to be able to make *between-paper* comparisons, accounting for differences in the quality of female- versus male-authored papers. While there is no perfect measure of quality, we observe ex-post realized citations, which are arguably correlated with paper quality and are highly relevant to publishers and editors. We therefore test whether male- and female-authored papers that are similarly reviewed by the referees receive the same number of citations, holding constant other characteristics of the paper and its authors.

As noted in CDV, there are at least two important concerns with such a comparison. First, there is a potential "publication bias" arising from the fact that positively reviewed papers are more likely to be published, and published papers tend to receive more citations. This could lead us to conclude that referee opinions are validated by citations when in fact the referees are biased. As in CDV, we address this by including an indicator for R&R status and a control function based on the editor's R&R probit model that corrects for any endogeneity in the editor's decision. Second, it is possible that female authors receive fewer citations for a paper of given quality than male authors – i.e., there is a systematic gender bias in citations. In this case, a finding of equal citations for female- and male-authored papers with similar referee recommendations would imply that the referees in fact set a higher bar for female-authored papers.

In an attempt to asses the likely magnitude of any such gender bias, we elicited beliefs from our survey respondents about the relative gap between citations and quality for female- versus male-authored papers. On average, respondents believe that female authors receive about 6 log point (i.e., about 6 percent) *fewer* citations than male authors, holding constant the quality of their work. There is a modest divergence between female and male respondents in the average size of this predicted bias, with female respondents reporting a 10 log point gap and males reporting a 3 log point gap.

With this estimate at hand, we turn to our main specifications. We find that, controlling for referee recommendations and other characteristics of a paper, female-authored papers receive 22 log points (s.e.=0.05), that is 25 percent, *more* citations than male-authored papers. Assuming that female-authored papers get 6 log points fewer cites for given quality, this estimate suggests that female-authored papers need to be of 28 log points higher quality to receive the same referee evaluation. This gap is robust to controlling for the share of female authors in a particular subfield, to alternative measures of citations, and a variety of alternative specifications. The gap falls slightly to 17 log points (s.e.=0.05) if we also control for the institutional affiliation of authors.[6] The magnitude of the gap *does* depend on whether we control for the prior publications of authors, since females have fewer prior publications and prior publications strongly predict citations.

What about mixed-gender papers? We find no citation premium for mixed-gender papers with a senior male co-author, consistent with our survey respondents' view that such papers are treated about the same as male-authored papers. In contrast, for mixed gender papers with a senior female

[6]We use as our benchmark specification the one without institutional prominence since it was the one pre-specified in the analysis plan, but we consider a large number of alternative specifications in our robustness tables.

co-author we find a 6 log point (s.e.=0.07) citation premium. We cannot reject that this premium is one-half as large as the premium for papers written by all-female teams, again consistent with our survey respondents' views about how such papers are treated in the editorial process.

In Section 6 we turn to the R&R decision of editors. We estimate a series of models that include controls for the summary recommendations of the referees, the publication record of the authors, and the gender composition of the author team. We find that editors tend to follow the referees' recommendations, with small (and insignificant) effects of gender in their R&R decisions. This behavior is consistent with the expectations of our survey respondents, who do not predict any difference in editorial decision-making conditional on the referee opinions. The fact that editors ignore the gender of authors, however, means that they are *over-rejecting* female-authored papers relative to a citation-maximizing benchmark.

We can use our model to compute how the R&R rate (for non-desk-rejected papers) would change if editors and referees were to assign a weight on female authorship consistent with citation maximization.[7] We estimate that the observed R&R rate for female-authored papers would increase from the observed 12.2 percent to 19.1 percent, a 57 percent increase. The corresponding adjustment for mixed-author papers would be small (0.5 percentage points). Averaging across all papers with at least one female author, we estimate that the average R&R rate for female authors would rise from 14.9 to 16.5 percent, an 11 percent increase.

Next, we study the desk-rejection decision. At this stage the only information available to editors is their own reading of the paper (a "private signal") and the characteristics of the paper and its authors. Similar to what we find at the R&R stage, we estimate that female-authored submissons receive about 24 log points more citations than male-authored papers, conditional on other observable controls. Thus, an editor who sets a gender-neutral citation bar should be *less likely* to desk-reject female-authored papers, conditional on the controls. Consistent with this prediction, we find that editors desk reject fewer female-authored papers. Nevertheless, the desk-rejection gap is smaller than would be predicted by a citation-maximizing benchmark.

We then address a further issue in the editorial process: Are some subgroups of referees more or less reliable in judging quality (as revealed by citations)? Do editors tend to pay more attention to more reliable groups? CDV find that the recommendations of more and less prolific referees are equally predictive of future citations, yet editors tend to place more weight on recommendations from referees with more prior publications. In the case of gender, we find that male and female referees are about equally informative. Editors, in turn, follow the recommendations of the two groups of reviewers about equally. Thus, we do not find any gender differences in this domain.

Finally, in Section 7 we study the impact of gender on delay times in the review process, including the time that referees take to return a recommendation, the time that editors take to reach a decision, and the number of rounds and the total delay between submission and acceptance for papers that obtain an R&R. We find no gender differences in any of these variables. Moreover, we find that male and female referees take about the same time to return their reports and are also about equally likely to accept a referee invitation, with no differential gap depending on the gender mix of the

[7]These calculations assume that editors assign the citation-maximizing weight to the author gender variables, but do not correct other deviations from citation maximization, such as those associated with author previous publications.

authors of a paper. Thus, we conclude that female and male authors experience similar delays in the review process, in contrast to the conclusion of Hengel (2017) based on data on the time from submission to acceptance at one journal.

In light of all these results, in our concluding section we revisit, and partially reconcile, the conflicting findings in the literature. The divergent findings appear due, at least in part, to the four different strategies employed to identify discrimination. One strategy, used by Broder (1993) and Abrevaya and Hamermesh (2012), is an audit-style comparison of recommendations by different reviewers of the same paper. These studies, like us, find no evidence of differential biases between male and female referees. This suggests that the animus documented by Wu (2018) against female economists in online discussion boards (which is widely attributed to male commentators) is largely absent in the review process.[8] A second strategy is to compare the acceptance rate of male-authored and female-authored papers without explicit quality controls, e.g, the analysis of NBER submissions by Chari and Goldsmith-Pinkham (2017). In our analysis, we also find that female- and male-authored papers have similar R&R rates when we do not control for the prior publication record of the authors. This reinforces the importance of controlling for observable measures of paper quality in trying to assess the impact of gender. A third strategy, which we use, makes comparisons of outcomes conditional on quality controls. We are aware of only one prior analysis using this design – Donald and Hamermesh (2006) – which comes to opposite conclusion as us, albeit in a very different setting (the election of officers of the AEA). Finally, a fourth strategy is to compare outcomes when author gender is blinded. Again, we are aware of only one such study, by Blank (1990), which finds that the masking of authors' names has no differential effect on the referees' evaluations of female authored papers. We note that our results are not necessarily inconsistent with Blank (1990)'s findings, since she does not evaluate the ultimate citations received by different papers, only whether the referees evaluate them more or less favorably. One interpretation of our results is that female authored papers have certain attributes – for example, a different mix of substantive versus purely methodological contributions – that are under-valued by referees but ultimately lead to higher citations. We further discuss the implications of our findings in the conclusions.

2 Gender Coding and Data

Gender Assignment. Given that (to the best of our knowledge) no journal in economics collects information on the gender of authors or referees, we follow the widespread practice of assigning gender based on names. We developed a multi-step process that relies on a combination of (1) publicly available data on the fractions of first names that are male versus female; (2) lists of female economists' names; (3) hand-collected data for lists of the authors and referees at each journal.

We began by assembling a test data set from EconLit of the names of 48,000 authors with a published article between 1990 and mid-2017 in a set of 53 economics journals (listed in Online Appendix Table 1). We then assembled five data sets of names:

1. The R-package "gender," which uses U.S. Social Security data to calculate the fraction of

[8] We emphasize though that we do not have access to the text of the reports and thus cannot directly test for any such difference in language.

people with a given first name who are male, *p(Male).* As Online Appendix Figure 1b shows for names in our EconLit sample, the distribution of *p(Male)* is highly bimodal with most of the mass at 0 or close to 1.

2. A dataset of given names assembled by Jörg Michael and first published by the German computing magazine, *c't.* This has coverage across European and many Asian countries and lists the relative frequencies of males and females with each name.

3. The RePEc list of the top 10% of female economists.[9]

4. A list of female members of the European Economic Association compiled by the Committee on Women in Economics.[10]

5. A list of common Chinese given names.

As shown in Online Appendix Figure 1, the first step in our assignment process is to assign "unknown gender" to a small number of common Chinese first names. We triage these names on two grounds: (i) Chinese names are not easily gendered, leading to a higher error rate, and (ii) there are often several individuals with the same Chinese first name (i.e., Chen Li), biasing upward the count of publications. This exclusion affects less than one percent of the names.

In the second step, we classify an author name as **female** if *both* the US and German datasets assign *p(male)*$<$ 0.01 for the author's first name, or if the full name is present in either the RePEc or EEA lists of female economists.[11] An audit showed a false positive rate for being classified as female of less than 1% for names in the test data set.[12] Likewise, we classify a name as **male** if one of the US or German datasets assigns *p(Male)*$\geq$ 0.99 to the author's first name and the other assigns *p(Male)*$\geq$ 0.50. An audit showed an average false positive rate for being classified as male of less than 1% for names in the test data set (Online Appendix Figure 1c).

For all names in the test data set that could not be gendered, we then used a team of undergraduate research assistants to look up each name, using the same process as in our audits. Any name that was not found by an initially-assigned assistant was assigned to a second assistant. This process ended up with about 3% of names in the test data set being unassigned a gender.[13]

Using this procedure, we were able to "pre-code" the genders of authors and referees. Specifically, prior to our main data extraction at each journal, the editorial assistant provided us with a list of the names of all authors and referees in the Editorial Express system. We then followed the same steps as in our test data set, first assigning gender using the four sources above and the list of gender-coded names from our EconLit sample, and then hand-coding the remaining names.

[9] https://ideas.repec.org/top/top.women.html. Downloaded in December 2016.

[10] https://www.eeassoc.org/index.php?site=&page=208&trsz=206. Downloaded in July 2017.

[11] We initially tried to assign female gender to a name for which one of the US or German data sets assigned *p(Female)*$\geq$ 0.99 and the other assigns *p(Female)*$\geq$ 0.50. We found, however, that this leads to too many "false positives" given the low fraction of female economists.

[12] We instructed undergrad research assistants to search for the name on the internet, looking for a picture or a pronoun reference. Typically the search would find a personal web page or a profile on LinkedIn, Google Scholar, or ResearchGate. In some cases the assistants would also find a news release or other mention with a pronoun or picture.

[13] We checked the reliability of the hand-coding process by having a fraction of names double-coded. The coders agreed on gender 74% of the time; one of two coders found enough evidence to determine a gender 14% of the time; neither was able to determine a gender 11% of the time; and the coders disagree 1% of the time. The low rate of disagreement suggests that if an assistant was able to find a positive way to identify gender then it was likely correct.

Data Extraction. For our data extraction process we wrote a program that could run in the editorial office of each journal and access submission information stored in the Editorial Express system and the pre-coded list of author and referee names with gender attached. This program created an anonymized data set with gender information on authors and referees for each submission, as well as all the other variables in the data set except citation counts. We are grateful to the four journals for agreeing to allow us to access their data.[14]

Google Scholar has created new barriers to accessing its database since the creation of the original CDV data set. We therefore decided to match our new data base back to the CDV data set, providing GS citations as of mid 2015. We used a fuzzy match algorithm based on all the identifying variables stored in the (anonymized) CDV data base. This yields perfect matches for all non-desk rejected papers, but multiple matches for some desk-rejected papers (which lack relatively rich referee-based information). For desk-rejected papers with multiple matches, we calculate our primary measure of citations as a simple average of $asinh(citations)$ across all possible matches.[15]

Given that the citations were extracted in mid-2015, for our main analysis we focus on submissions up to 2013 (as in CDV) in order to leave enough time for the citations to be realized. We use the manuscripts submitted in years 2014-2017 in the online appendix to replicate the parts of our analysis that do not rely on citations, such as the analyses of R&R decisions and delays.

Analysis Plan. We posted an analysis plan on the AEA site under number AEARCTR-0003048 prior to the completion of our data collection. The plan describes the key steps in our analysis, which we follow in this paper, with the addition of a few robustness checks which we had not envisioned.

3 Descriptive Overview

3.1 Summary Statistics

Table 1 presents summary statistics on our database of 29,890 submissions during the period 2003-2013 (Columns 1-6). About half of the submitted papers (15,147) were not desk-rejected (NDR) and were assigned to at least two referees; we present statistics for these papers in Columns 7-12.[16]

We classify papers into five groups, based on the gender composition of the author team: 1) all male; 2) all female; 3) mixed gender with a senior female co-author (i.e., the co-author with most publications is female); 4) other mixed gender; and 5) gender undetermined. The last group is comprised of papers with at least one co-author with unassigned gender. In the overall sample (Column 6), 66% of papers are authored by an all-male team; 8% by an all-female team, 3% have mixed-gender teams with a senior female co-author; 16% are from mixed-gender teams where the senior author is male or teams with a "tie" - most often teams where all co-authors have no previous publications; and 7% are from a team with undetermined gender. The gender distribution for NDR papers (Column 12) is similar, with a smaller proportion with undetermined gender, at 4%.

[14] The data agreement with the journals has two conditions: (i) no analysis should present separate results by journals, and (ii) unlike the CDV data set, this supplemented data set will not be posted, even upon publication.

[15] Our results are essentially the same if we retain all possible matched pairs and estimate our models using the inverse of the number of matches for a given paper as a weight.

[16] Among the non-desk-rejected papers, we exclude papers that were assigned to only one referee, since this process (which is especially common at the *Review of Economic Studies*) appears to be a form of desk-rejection.

All-male papers (Column 1) have a higher probability of NDR and of receiving an R&R than all-female papers (Column 2). Mixed-gender papers of all types (Columns 3 and 4) have higher rates of NDR and more favorable R&R decisions than those of either single gender. As we discuss below, these gaps are largely explained by differences in previous publications and number of authors.

Figures 1a-b complement the data in Table 1 by showing the distribution of referee recommendations and editorial decisions for the non-desk-rejected papers. The summary recommendations submitted by referees fall in 7 categories, from "Definitely Reject" to "Accept".[17] The referees at our four journals are generally quite negative: 54% of recommendations are "Reject", and another 12% are "Definitely Reject". Ten percent are "Weak Revise and Resubmit", 10% are "Revise and Resubmit", and only 7% are "Strong Revise and Resubmit" or "Accept". Female-authored papers have the highest fraction of "Reject" recommendations (56%) while papers written by a mixed gender team with a male senior author (or no clearly senior author) have the lowest rate (51%).

Our benchmark measure of paper quality–$asinh(citations)$ based on GS citations collected in mid 2015[18]– is highest for mixed-gender papers, followed by all-male papers, with all-female papers at the bottom. Figure 1e shows the cumulative distributions of this variable by gender group. The rankings across gender groups are the same at all quantiles, suggesting only limited heteroskedasticity. In Figure 1f we show the same variable, but now residualized with respect to our key control variables, journal-year fixed effects, field, number of authors, and previous publications. Taking into account these controls reverses the ranking, with higher citations for female-authored papers.

The number of coauthors is an important characteristic of papers. Two-authored papers are the largest group in our overall sample (39% of all submissions, 42% of NDR papers), followed closely by single-authored papers (37% of all submissions, 31% of NDR's). Three-author papers represent around 20% of the sample, while only 5% have four or more authors. Author team size is quite different for all-female than all-male papers, reflecting the fact that in a field like economics with only 16% female authors, the likelihood of a large team of female co-authors is low.[19] This leads us to include controls for the number of co-authors in all our analysis below.

We measure the prior productivity of authors by the number of publications in the 5 years prior to the submission year in a set of 35 high-impact journals (Online Appendix Table 1). For papers with multiple authors, we use the publication record of the most prolific co-author. Authors of all-male papers tend to be better published than the authors of all-female papers (see also Figure 1c), possibly reflecting the fact that the share of women is higher in younger cohorts of economists. Again, this difference leads us to control for previous author publications in our models.

As might be expected, referees tend to have more publications than submitting authors: 45% of referees have three or more recent publications, compared to 27% of submitting authors. On average 15% of referees are female. Consistent with the results for authors, Figure 1d shows that female

[17] There are actually 8 categories with "Conditionally Accept" and "Accept" at the top. Since these two are very rare we collapse them into a single "Accept" category.

[18] We use the $asinh$ transformation to accommodate zero citations. For reference: $asinh(x) \equiv ln(x + (1 + x^2)^{1/2})$; $asinh(0) = 0$; $asinh(1) = 0.88$; $asinh(x) \approx ln(x) + ln(2) \, for \, x \geq 2$. Thus for more than 2 citations the $asinh$ function closely parallels the natural log function.

[19] Interestingly, the fraction of single-authors who are female is very close to the fraction of authors of 2-author papers who are female, suggesting that females are no more likely to work alone. There is, however, some evidence of assortative matching of co-authors by gender. For example, among 2-authored papers the fraction written by two females is 4.1%, higher than the 2.5% rate expected under random matching.

referees tend to have fewer prior publications than male referees.

We compare the gender distributions of authors and referees in our database to the distribution among *all authors* of papers published in a set of 53 journals over the 2008-2015 period, drawn from EconLit.[20] Figure 2a plots the share of female authors in our database for each of 13 broad fields (identified by the first letter of the JEL code) against the corresponding share in the EconLit data base. (Papers with multiple JEL codes are treated as being fractionally represented in each field. So a paper with 3 JEL's is treated as being one-third in each field). There is wide variation in the share of female authors across fields, with higher rates in labor and development and lower rates in macro, theory and econometrics. Interestingly, the share of female authors in a given field in our 4-journal sample (on the y-axis) tends to match the share among authors in EconLit (on the x-axis), confirming that the gender distribution in our sample is broadly representative of the distribution among actively publishing researchers. Figure 2b shows that the share of female *referees* assigned to a given paper also generally matches the share of female authors in the field.

Figure 2c shows the evolution over time of the female share for: (1) authors in our EconLit data base; (2) authors of papers in our submission database; (3) referees of papers in our submission database. The three series track each other relatively closely: females represent about 15% of authors and referees in 2006, with a slowly increasing trend, that rises to about 17% in 2013, the last year in our main sample. Thereafter the shares of female authors and referees remain fairly constant.[21]

The patterns in Figures 2a-b underscore the importance of controlling for field differences, since the presence of female authors is correlated with field, and both citations and the probability of an R&R verdict vary across fields. In our analysis we include indicators for each of 13 major fields. A plausible concern is that even within broad fields, some sub-fields have more female researchers than others. Our confidentiality agreements precluded us from retaining more granular subfield information. However, we were able to create two variables that serve as proxies for the gender composition of the subfields represented in a paper. The first is the share of female authors in the same narrow subfield (based on the 2-digit JEL code) published during a 5-year moving average around the year of submission in the 53 journals in our EconLit sample. For papers with 2 or more subfields we assign the average value of this share across all subfields. The second is the share of JEL subcodes assigned to a paper that are closely associated with gender-related topic, which we take to be JEL codes D1 (Household Behavior and Family Economics), J1 (Demographic Economics), K36 (Family and Personal Law), and K38 (Human Rights Law and Gender Law). This variable, which we call "gender-related subfield" has a mean of 0.04, but is twice as high for all-female papers.

3.2 Survey Evidence

To aid in the interpretation of our findings, we conducted a survey of editors and academic economists about their perception of gender differences in the publication process. The survey, which was approved under Berkeley IRB 2018-04-10955, was sent to three groups: (1) editors and co-editors at the 4 journals in our sample; (2) a stratified random sample of 200 economists (100 male and 100

[20] We use this period to roughly correspond to the period when the submissions in our data set would be published. We coded genders for these authors in constructing our test data set for evaluating our gender-assignment process.

[21] Online Appendix Figure 2 presents this evidence separately for each field.

female) with at least 4 publications in our top-35 journal set from 2013 to 2017; (3) all assistant professors of economics in the top 20 American schools and top 5 European schools with PhDs from 2015 to 2017. We selected these groups to capture potential heterogeneity in perceptions across subgroups of economists. The views of editors are obviously relevant given their role in the publication process. The views of the second group of "highly active" economists presumably reflect extensive recent experience with the editorial process. Finally, the views of the third group of recent PhD's represent the perceptions of promising researchers at the start of their careers.[22] As shown in Table 2, our response rates were reasonably high, especially among female economists (50 percent). The survey included 14 different questions, with the key ones reported in Table 2 focusing on:

- whether female-authored papers are more likely to be assigned to female referees
- the difference in how male and female referees evaluate male- and female-authored papers
- the likelihood that the editor gives an R&R to male- versus female-authored papers
- the extent to which citations vary with author gender, holding constant the quality of a paper
- the informativeness of male and female referees
- the degree to which editors follow the recommendations of male and female referees
- how papers by mixed-gender teams are treated in the review process

We use these answers in three main ways. First, following our analysis plan, we use survey respondents' beliefs about differences between the two types of mixed gender teams to inform our classification of these papers. Second, we use beliefs about the potential differences in citation rates for male-authored and female-authored papers to help interpret the gaps in citations we measure in our analysis. Finally, we use the answers as "priors" to help understand how consistent our findings are with the expectations of people in the field, as in DellaVigna and Pope (2018).

4 Assignment of Referees

We begin our analysis by focusing on the "matching" process used by editors to assign non-desk-rejected papers to referees. This analysis provides three pieces of information. First, it helps us to understand whether the assignment of referees appears to be "as good as random" (conditional on some paper characteristics) or not. Second, it provides revealed-preference evidence on the degree to which editors appear to be concerned about gender-related issues in the review process. Third, it also yields direct information on how editors treat different types of mixed gender teams.

Figure 3 shows the probability of assignment to female referees. Papers by all-female authors are assigned to female referees at nearly twice the rate (26 percent) as all-male authors (14 percent). The mixed-gender teams fall in between, with a higher fraction for mixed-gender papers with a senior female co-author (21 percent) than for the other mixed-gender papers (18 percent).

[22] Within each group, the survey was conducted anonymously: we did not keep track of individual respondents. Within the second and third group, however, we referred male and female respondents to different URL's to keep track of gender.

These simple comparisons do not take account of other paper characteristics that may be relevant for the assignment of referees – including the field of the paper. The differences in the share of female authors and referees across fields (Figures 2a and 2b) could lead to assortative matching of referees that is driven entirely by field-specific expertise. We thus turn to a regression-based analysis in Table 3, using a linear probability model for the likelihood that a referee assigned to a paper is female. We fit the models using a paper-referee data set and cluster standard errors at the paper level.

The specification in Column 1 of Table 3 with no controls reproduces the differences in Figure 3. The specification in Column 2 adds our full set of controls, including dummies for the number of authors, dummies for the number of publications of the most prolific co-author and for the referee, and indicators for broad field. We also include the share of female authors in the 2-digit JEL code(s) of a paper, and an indicator for gender-related subfields. As expected, both of these variables are highly significant predictors of the assignment of a female referee: a 10 percentage point increase in the share of female authors in a given subfield is associated with a 3 percentage point increase in the share of female referees assigned to papers in that subfield, while a paper with all gender-related subfields has a 20 percentage point higher probability of a female referee.

Even controlling for all these variables, we still find a large effect of the author gender mix on the assignment of referees. Other things equal, female-authored papers are 7 percentage points (s.e.=1 ppt) – or about 50 percent – more likely to be assigned to a female referee.[23] Mixed-gender papers with a senior female co-author are 5 percentage points more likely to be assigned to a female referee, while the other mixed-gender papers are only 3 percentage points more likely to be assigned to a female referee. The differences between the two mixed-gender groups motivate our choice to analyze these groups separately in the rest of the paper, along the lines we laid out in the analysis plan.[24]

As a point of comparison, Column 3 displays how much the author gender mix affects other dimensions of the referee assignment, and in particular the prominence of the referees assigned. Controlling for other variables, female-authored papers are less likely to be assigned to referees with 3+ publications, but the impact is quantitatively much smaller, a 5 percent decrease (2.5 percentage points out of a mean of 46 percent), compared to the gender-based assortative matching.

One interpretation of journal editors' tendency to assign female-authored papers to female referees is that they are concerned about possible biases in the assessment of these papers by male referees. Indeed, the editors in our survey on average expect that male referees are less likely than female referees to give a positive evaluation of female-authored papers (17.5% probability versus 20.7%) but are (very) slightly more likely than female referees to give a positive evaluation of male-authored papers (19.7% versus 19.3%). Editors thus think that referees of one gender are relatively biased in favor of papers by their own gender group, and in the interests of fairness they may want to get at least one female referee for a female paper. Another possible explanation (for which we have no direct evidence) is that female authors are more likely to cite other female authors, and editors tend to select referees whose works are listed in a paper's bibliography.

[23] Abrevaya and Hamermesh (2012) also find a similar assortative gender matching, in particular for the later years in their sample (2000-2008) which is closest to our sample period.

[24] In the analysis plan we wrote "*We intend to use this pattern of assignment to infer how editors classify mixed-gender papers. Suppose for example that the rate of assigning a female referee is different for mixed gender papers with a senior author who is female [...] than for mixed gender papers with a senior author who is male. This would suggest that it is important to analyze the two mixed-gender author groups separately.*"

Whatever the reasons for this assortative matching, it appears to be a surprise to our survey respondents: 77% of the respondents expected no such matching. Even among editors, only one-third anticipated a pattern of gender-matching, which suggests that the assignment on the side of the editors may be more of an unconscious or case-by-case decision than a mediated one.

5 Referee Recommendations

5.1 Simple Audit Comparison

Given that female referees are more likely to be assigned to female-authored papers, the next question is whether in fact referee gender affects the evaluations of female-authored versus male authored papers. Following our analysis plan, we use two main measures of referee support for a paper. The first summarizes the seven categorical referee recommendations into an index based on the predicted $asinh(citations)$ associated with that recommendation, using the coefficients from the main citation model in CDV (Table 2, Column 4). This measure has the advantage that it uses all the variation across the recommendations. As a second, simpler measure, we use the share of recommendations that are positive – that is, "Revise and Resubmit" or better.

Figures 4a and 4b show the mean assessments of female and male referees (for papers that are assigned to at least one referee of each gender) separately by author gender group. Observations are weighted by the inverse of the number of referee reports for the paper to ensure that each paper receives equal weight. The bars show ± 2 standard error intervals, constructed by clustering at the paper level. The two measures of referee support show very similar results. On average, female referees and male referees have very similar evaluations, tracking each other across author groups with different author-gender composition. There is no evidence of *a relative assessment gap* between male and female referees that depends on the authors' gender composition.[25]

Table 4 presents a series of OLS models for the two measures of referee support, adding controls for other paper characteristics. These models are fit to referee-paper observations, weighing each observation by the inverse number of referees for the paper and clustering standard errors by paper.

With our full set of controls (Columns 2 and 6), the coefficient for female-authored papers is essentially 0: we find no differences in how the referees assess all-female or mixed gender papers relative to all male papers (the omitted group). One highly significant estimate is the negative coefficient for papers with an unknown gender composition. This presumably reflects the fact that papers for which we could not find any online profile of at least one of the co-authors, and thus which could not be gender-coded, are likely to have attracted little attention.

The models in the remaining columns test for any *differential assessments* of male- and female-authored papers by different referees by including controls for the gender of the referee and interactions between the referee's gender and the authors' gender group. In Columns 3 and 7, we restrict the sample to papers with at least one male and one female referee. In Columns 4 and 8, we include paper-specific fixed effects and thus identify the gender effects from within-paper differences in assessments, removing all the between-paper variation in quality or other features like field or author

[25] Online Appendix Figure 3a-b shows that we find a similar pattern if we use the full set of submissions up to 2017.

publication record. The interaction effects in these models represent *differences-in-differences* in the relative evaluation of female referees versus male referees of papers in a particular author gender group relative to papers with all-male authors.

In these most complete specifications we find that (1) there are no differences in recommendations across male and female referees, and (2) there are no interactions between the referee gender and the author gender mix. Finding (1) implies that the more negative average assessments offered by female referees in Columns 3 and 7 are driven by the fact that female referees tend to be assigned weaker papers. Once we isolate within-paper differences in average assessments, female referees are neither more, nor less, positive than male referees. This contrasts with the expectations of the survey respondents, who largely expected female referees to be more positive than male referees.

Finding (2), the key result in Table 4, implies that there is no *relative* favoritism (or bias) by referees of one gender for, or against, papers by authors in different gender groups. This also differs from the average expectations of our survey respondents (including the editors) who expected that female referees would be a bit more positive toward female-authored papers. The absence of any large or significant interaction between the gender of referees and the gender compositions of the submitted papers is consistent with the results in Abrevaya and Hamermesh (2012). Our estimate on the interaction between all-female authors and female referees in Column 8, a point estimate of 0.00 (s.e.=0.02), compares to their point estimate of 0.03 (s.e.=0.04).[26]

5.2 Recommendations and Citations

Though the difference-in-differences specifications in Columns 4 and 8 of Table 4 rule out the presence of any large or statistically significant *relative bias* by referees of one gender compared to the other, they do *not* allow us to conclude that referees set the same standards for female authors as for male authors. It is possible that both groups of referees are biased for, or against, female-authored papers. If, for example, female-authored papers are of higher quality conditional on field and prior publications, the fact that referees rate them equally would indicate that they set a higher bar for female-authored papers. To make further progress we need to be able to make comparisons *across* papers by different gender groups that take into account differences in quality.

While we do not have a perfect measure of quality, we do observe the Google Scholar citations received by each paper, which are plausibly correlated with paper quality, and are an outcome that journals clearly care about. We can thus test whether male-authored and female-authored papers receive the same citations for a given set of referee recommendations, holding constant other features of the paper and its authors.

There are at least two important confounding factors that complicate the interpretation of a comparison that holds constant referee evaluations and looks for differences in citations between female- and male-authored papers. The first is that referee assessments affect the probability that a paper is ultimately published, and publication arguably raises citation rates. In essence the referee

[26]Since submissions are double-blind in the journal they consider (unlike in the journals we consider), the referees may not be aware of the gender of the authors. Thus, Abrevaya and Hamermesh (2012) compare later submissions, where it would have been easier to infer the identity of the authors, to earlier submissions. This comparison (a triple interaction term) has a point estimate of -0.01 (s.e.=0.08).

evaluations are correlated with an omitted variable (publication status). Building on CDV, we address this by controlling for the editor's R&R decision, while including the generalized residual from the editor's R&R decision model to deal with potential endogeneity of the editor's decision.

Second, and more importantly given our focus, it is possible that female authors receive fewer citations than male authors, holding constant paper quality. One plausible channel for this "gender bias" in citations is networking: female economists may be less likely to get invited to conferences, or less likely to attend if invited. To the extent that female authors get fewer citations, a finding of no difference in citations would imply that referees set a higher bar for female-authored papers.

To get a sense of the potential gender gap in citation rates, we asked the respondents in our survey to quantify their belief about the gap (in log points) between citations and quality for female- versus male-authored papers: "*Q7. Now consider two different papers in the same field of comparable quality, one written by female authors, the other written by male authors. Do you think the female-authored paper will get more, about the same, or fewer citations? Q8. If you answered more or fewer, how large do you think the citation difference will be in log points? For example, if you think that female-authored papers will have X log points (X percent) higher citations (conditional on quality), write X. If you think that female-authored papers will have X log points (X percent) fewer citations (conditional on quality), write -X.*"

Figure 5 displays the distribution of this citation penalty for editors, female respondents, and male respondents. While we acknowledge that the question is a difficult one to answer, we are encouraged by the wide agreement among the respondents on two points. First, the modal response across all groups is that there is no differential citation bias. Second, all but a handful of respondents believe that, if there is a gender bias in citations relative to quality, it tends to lower citations of female authors. Overall, the mean elicited citation bias is 6 log points: among male respondents the mean bias is 3-4 log points, whereas among females it is 10-11 log points. Below we use a 6 log point citation discount gap as a benchmark, but alternative estimates between 0 and 10 log points do not qualitatively affect our conclusions.

With this estimate at hand, we turn to our key specifications in Columns 1-4 of Table 5, which relate the inverse hyperbolic sine of citations to the referees' recommendations and other paper characteristics.[27] We summarize the opinions of the referees by the fractions of recommendations in each of the 7 categories. For example, if a paper was reviewed by 3 referees, with 2 recommending "Reject" and 1 recommending "Revise and Resubmit" we set the fraction of "Reject" recommendations at 2/3, the fraction of "Revise and Resubmit" recommendations at 1/3, and the fractions of all other categories at 0. CDV documents that this simple procedure provides a relatively accurate representation of the effect of the recommendations on both citations and the editor's R&R decision.

An obvious problem in comparing citations for papers that have been circulating for a longer or shorter time is that cites take time to accumulate. Average citations may also be different for papers submitted to different journals in our database. To address these issues we include an unrestricted set of journal $\times$ submission-year effects in all our models.

In a specification that controls only for the referee recommendations and journal-year fixed effects (Column 1), female-authored papers receive 7 log points (s.e.=0.05) fewer citations than male-

[27]The specifications in this table follow exactly the format laid out in the analysis plan.

authored papers, while mixed-gender papers receive 26-37 log points more citations, and papers with undetermined gender teams receive 36 log points fewer citations. At face value, this would suggest that referees evaluate all-female and all-male paper about the same, but that they tend to give lower recommendations to papers by mixed gender teams than would be justified by their quality.

The picture changes substantially when we add controls for author's prior publications, the number of co-authors, field dummies, and our two measures of the gender-related field focus of the paper in Column 2. Controlling for all these observables variables but not the referee recommendations, female-authored papers attain higher citations by 24 log points. Thus, among the non-desk-rejected submissions, the ones with all-female authors appear to have higher quality, once one controls for other features of the papers (like the field) and of the authors (like previous publications). The submissions by mixed-gender authors, instead, do not have significantly different citations than the ones by all-male authors (the omitted category), once one adds the controls (especially previous publications which are higher for the mixed-gender group).

How do these results change once we also control for referee recommendations? As Column 3 shows, all-female papers still have a significant positive citation premium of 22 log points, and there is a small and statistically insignificant premiums for mixed gender papers. Thus, the referees do not appear to take into account the higher citations for female-authored papers documented in Column 2 and they appear to set a higher bar for all-female-authored papers.

As we mentioned above, a confounding factor in the relationship between referee recommendations and citations is the fact that the referee ratings are correlated with the probability of ultimate publication. To the extent that published papers get more citations, this leads to an upward bias in the effect of the referee opinions on citations. It is not obvious whether such a "publication bias" will affect the citation gap between male- and female-authored papers, but it is nonetheless important to address. We therefore add an indicator for a paper's R&R status to our citation models, along with a control function term to deal with the endogeneity of R&R status.

Following CDV, we model the probability of an R&R verdict as depending on the referee reports, the characteristics of the authors and the paper, and a variable meant to capture the relative leniency of the particular co-editor in change of the paper – his or her mean R&R rate (excluding the current paper). This model is shown in Column 7 of Table 5 and is discussed more extensively below. We take the generalized residual from the R&R model and add it to the citation model. The coefficient on this residual can be interpreted as a measure of the correlation between the editor's private information about the quality of the paper and the residual component of citations.

Our benchmark specification in Column 4 of Table 5 includes the referee recommendations, our full set of controls for characteristics of the paper, an R&R dummy, and the control function term. In this specification, female-authored papers receive 22 log points more citations (s.e.=0.05) than male-authored papers, very similar to the citation gap in the previous specifications.

One way to interpret this gap is to ask how much better a paper authored by an all-female team would have to be (relative to a similar-quality paper authored by an all-male team) to receive comparable recommendations from the referees. The answer, measured in units of citations, is 25% ($exp(0.22) - 1 = 0.25$). This gap is equivalent to the gap in citations between a paper that receives two "Weak R&R" recommendations and one that receives one "Weak R&R" and one "R&R"

recommendation. We return to the magnitudes below.

If one believes that female-authored papers tend to receive fewer citations than male-authored papers, the quality gap is even larger. Taking our estimate from the survey of a 6 log point gender bias in citations, female-authored papers would have to be of 28 log points (32%) higher quality than male-authored papers to receive the same referee assessment.

So far we have emphasized the citation gap between papers written by all-female teams and all-male teams. What about papers written by mixed-gender teams? We find a relatively precise 0 citation gap for mixed-gender papers in which the senior author is not female, consistent with the survey responses that indicate that such papers are considered similar to male-authored papers. For mixed gender papers with a senior female author, however, we find a citation gap of 6 log point (s.e.=0.07). We cannot reject the hypothesis that these papers are treated "half-way" between female-authored and male-authored papers, the modal answer given by our survey respondents about how such papers are treated in the review process.

We provide a graphical illustration in Figure 6a, where we plot the average asinh of citations for each recommendation category, computed separately for papers with different gender compositions. The citation variable is residualized with respect to all the control variables in Column 4 of Table 5, since these variables are clearly correlated with the author gender.[28] At nearly each level of referee recommendations, female-authored papers have about higher citations than male-authored papers, with an average difference of 20 log points, consistent with the regression results.

In Figure 6b we present a similar exercise, splitting the series by the referee gender. As documented also in Table 4, there are no obvious differences by gender of the referee: both male and female referees appear to hold female-authored papers to a higher bar in the citation metric.

5.3 Robustness and Heterogeneity

Robustness. Possible concerns with our results is that they may reflect the effect of some unobservable variable that is correlated with gender, or that they may depend on the particular functional form assumed for measuring citations, or that they may be due to a peculiar subset of the data. We thus consider a broad spectrum of robustness checks, summarized in Table 6. For each specification (in the rows), in Columns 1-3 we display the coefficient on the three author-gender variables in the citation regression, including the full set of controls as in Columns 4 of Table 5. We discuss below the robustness result for the R&R specifications, reported in Columns 4-6. For several of the robustness specifications, we report additional information in online appendix tables.

As far as unobserved factors, we saw in Table 5 that adding controls significantly increased the point estimate of the citation premium for all-female papers. If other unobserved variables tend to have the same pattern (as formalized in Altonji, Elder, and Taber, 2005), then we would expect their omission to lead to a downward-biased estimate of the all-female citation premium.

Of particular interest are the two controls for the gender-related sub-fields of each paper. A possible explanation for the citation premium for all-female authors is that their papers tend to be in fields with more gender-related content, and papers in these areas get more citations. Contrary to

[28] A graph with the raw citation variable is shown in Online Appendix Figure 4a.

this story, however, Table 5 shows that both the average share of female authors in a sub-field and the share of sub-fields in gender-related areas have small, insignificant effects on citations. Moreover, the effect of the female author-share variable on citations is actually slightly negative. Taken as a whole, there is no evidence of an upward bias due to gender-related subject matter.

To provide additional checks on the effects of the observed controls, in Online Appendix Table 2 we show alternative citation models in which we selectively add each subset of controls; all specifications control for the referee recommendation variables and year-journal fixed effects. When we add only the field controls (Column 1) we obtain results similar to the specification with no controls. Adding the author publication variables (Column 2) shifts the coefficient on the all-female papers to 0.14 log points (s.e.=0.05), indicating that the author prior publications controls is the key control. Further adding the controls for number of authors yields our benchmark specification (Column 4 in Table 5), further raising the all-female citation coefficient to 0.22 log points.

Another possibility is that our controls for the publication record are too crude. In our main specification, the publication record is the maximum number of publications in the 5 years prior to submission across the coauthors. In Column 3 we add controls for the *average* number of publications among all the coauthors, and in Column 4 (and also reported in Table 6), we further add controls for having publications in top-5 journals (as opposed to 35 high-quality journals), as well as controls for publications 6-10 years prior to submission. The point estimate of the citation premium for all-female papers is unaffected by either addition. Finally, in Column 5 (and also reported in Table 6) we add measures of the quality of the institution of the co-authors. Interestingly, female authors are located at slightly more prestigious institutions, and papers by authors at higher-ranked institutions get more citations, so the addition of controls for author institution lowers the estimated female-author premium slightly, to 17 log points (s.e.=0.05).

We then consider alternative specifications for the citation variable. In Online Appendix Table 4 we present estimates using $ln(1 + citations)$; using the percentile of *citations* in a journal-year cell (also summarized in Table 6); using an indicator variable for having a paper in the top x percent of citations, where x corresponds to the R&R rate in that journal-year cell (also summarized in Table 6); using an indicator for having a 'superstar' paper in the top 2 percent of citations for the journal-year-cell. Across all these specifications, we find similar results. While one cannot directly compare the size of the gender coefficients, by comparing their size to the magnitude of the referee coefficients, we can see that the key findings on the gender author mix are stable.

In Online Appendix Table 5 we consider the concern about the left-censoring of citations, given that 19% of papers have zero GS citations. A tobit specification with left-censoring (Column 2) yields the same qualitative insights, with, as expected, somewhat larger point estimates for all the variables. The left-censoring is a serious concern when using *asinh* of Social Science Citations (SSCI) citations instead of Google Scholar citations. The SSCI citations accrue only to published papers and thus take time to accumulate, with 61% of papers with zero SSCI cites even restricting to submissions in years 2006-10. We present the results for the years 2006-10 (Column 3) and 2006-08 (Column 5, also shown in Table 6), with for comparison estimates of our benchmark specifications for those same years in Columns 4 and 6. We find similar, if noisier, results, with attenuated coefficients for the all-female variable in years 2006-08 and similarly large coefficients for years 2008-10. Overall, the results are robust to taking into account censoring and the alternative citation measure.

Heterogeneity. Next, in Table 6 we consider the heterogeneity of the results in several sub-sets. We estimate the results for the earlier years in our sample (2003-2009) versus the later years (2010-2013); we find a stronger impact in the earlier years, but similar qualitative patterns. Citations to these older papers are presumably less affected by factors such as conference presentations and prior circulation of working papers, so the finding of a larger female premium suggests that such "short term" determinants of citation are not the primary driver of our main results.

Next, given that the controls for the number of authors play some role in the estimates, we present the results separately for papers with 1 author, 2 authors, and 3+ authors (with the full results in Online Appendix Table 3). We estimate a fairly similar citation gap for all-female papers for both papers with 1 author (0.17 log points, Column 1) and for papers with 2 authors (0.34 log points, Column 2). For papers with 3+ authors, we cannot reliably estimate the impact of papers with all-female authors given how rare such papers are, but we estimate that mixed-gender papers with a female senior author have 0.23 log points (s.e.=0.05) higher citations.

We then split the results by another important determinant of citations, the number of author publications, 0-3 versus 4+. We find higher citations for all-female-author papers in both categories, but the result is particularly large for all-female-authored papers with more prolific authors, with 4+ publications in the previous 5 years: 0.49 log point (s.e.=0.10). Thus, we do not find a reversal of the pattern for high-prominence female authors, as in Bohren et al. (2018).

We then compare the impact in fields with a lower share of female economists, like theory, versus fields with a higher share, like labor economics, doing a median split using the share of female authors in the JEL codes of that paper. The results are essentially the same in the two sub-samples.

For the last two splits, we consider characteristics of the referees. First, we split by whether the paper has 1-2 referees, versus 3 or more referees. The citations coefficient for all-female authors is larger for papers with 3+ referees, but the qualitative patterns are similar. Finally, we split by whether the referees are all male, or whether at least one of the reviewers is female. We find the pattern in both subsamples, but the citation differential for all-female papers is in fact larger for papers with some female referees, consistent with the findings above that female referees are not more positive than male referees in their evaluation of female-authored papers.

In Online Appendix Table 6, we present some of these same heterogeneity results by interacting the key variables with the dimension of heterogeneity. We find parallel results.[29]

6 Editorial Decisions

So far, we have focused on how referees treat teams of authors with different gender compositions. Using a citation-based benchmark we find that both male and female referees appear to set a higher bar for papers by all-female authors than for those by all-male authors. But of course the referees' opinions are only part of the editorial process: editors make the ultimate decision of whether to reject a paper or invite a revision. Moreover, editors make the initial screening decision on whether

[29] In the pre-analysis plan we pre-specified the heterogeneity analysis in Online Appendix Table 6. We present additional heterogeneity splits in Table 6 in response to comments we received and to further probe the results.

to desk-reject a paper. How do the editors treat papers with different gender composition? We turn to this analysis in this section, once again following our analysis plan.

6.1 R&R decision

We consider the editor's R&R decision using the framework developed in CDV. Specifically, we fit a series of probit models for the R&R decision using the same basic variables as are included in our citation models. Our main results are presented in Columns 5-7 of Table 5.

To aid in interpreting the R&R model, it is useful to start from the simplest benchmark in which citations are an unbiased (but potentially noisy) measure of quality, the citation-quality link does not differ across types of papers, and editors set the same quality bar for all types of papers (see CDV for details). Under this benchmark, any variable that predicts citations should similarly be a predictor of the R&R decision for the editor. Further, the coefficients in the R&R probit model should be proportional to the coefficients in the model for citations: the more a variable predicts citations, the higher its impact should be in the R&R decision. Online Appendix Figure 5 presents an example with simulated data, plotting on the x axis the coefficients from the citation regression and on the y axis the coefficients from the R&R probit model. Under the benchmark model, the coefficients should be lined up on a line, like coefficients β_1 through β_6 are in the simulated example.

If, conversely, the editor sets a higher bar for a type of paper, the characteristic for that paper will lie below the line: characteristics that predict citations are not used proportionately in the editorial R&R decision. In Online Appendix Figure 5, this is the case for coefficients β_7 and β_8.

Turning to Table 5, the benchmark R&R specification in Column 7 controls for the referee recommendations, the full set of paper characteristics, as well as the mean R&R rate of the editor assigned to the paper (excluding the paper under consideration). This variable is meant to represent a co-editor-specific taste shifter that affects the probability of R&R but does not directly affect citations. It is therefore excluded from the citation model, and plays the role of an instrumental variable in identifying the effect of the control function in the citation model.

As discussed in CDV, a comparison of our baseline R&R model (Column 7) and our baseline citation model (Column 4) shows that the referee recommendation variables enter nearly proportionately, as would be expected if editors take the measures of referee support as an index of paper quality, and citations depend on the same index. Specifically, a plot of the R&R model coefficients for the 7 referee recommendation variables (the 6 reported in the table plus a 0 for the omitted category) against the citation model coefficients is approximately linear with a slope of about 2.5.[30] If editors are trying to maximize expected log citations, then all the variables in the R&R model should have coefficients that are 2.5 times larger than their coefficients in the model for citations.

Given that papers written by an all-female team receive 0.22 log points more citations (Column 4), under a proportional decision model we would expect a coefficient of $0.55 = 0.22 \times 2.5$ in the R&R probit model. A coefficient of this size would be just large enough to offset the bias in the referee recommendations and ensure that female authored papers are evaluated in accord with their "quality", as revealed by citations. As Column 7 shows, the coefficient on the all-female paper in the

[30] The correlation is 0.98; the slope is 2.47.

R&R decision is instead 0.01 (s.e. 0.06): editors do not seem to *undo* the referee's apparent biases at all. The coefficient is precisely estimated, such that we can confidently reject the hypothesis of a value of 0.55 under the citation maximizing benchmark.

Note that the proportionality test is derived under the assumption that the citation-quality relationship is the same for male and female authors. The survey responses discussed earlier suggest that, if anything, female-authored papers receive about 6 log points fewer citations, given quality. If we take this into account, the editor coefficient should have been even larger, at $0.69 = (0.22 + 0.06) \times 2.5$. Thus, the violation of proportionality does not appear to be due to a difference in this citation-quality relationship.

It is useful to draw a parallel to the case of the author publication variables, which CDV considers in detail. Similar to the case of all-female authored papers, the impact of having a well-published coauthor on the editor's R&R decision is substantially smaller than one would expect under citation-maximizing behavior. For example, papers by authors with 6+ publications have an R&R probit weight of 0.41, compared to a predicted weight of $2.48 = 0.99 \times 2.5$. That is, author publications are underweighted by a factor of about 5; we cannot reject that all-female authorship is underweighted by a factor of 5 as well (in which case the expected coefficient in the R&R model would be 0.11).

A key difference, though, between the two coefficients is that a citation-quality gap may plausibly account for some of the deviation from proportionality for the author prominence variables, given that more-published authors have better access to working paper series and networks that accelerate the spread of citations. Conversely, as we discussed, this is unlikely to be the case for female authors.

Figure 6c shows the corresponding graphical evidence, displaying the R&R rate as a function of the referee recommendation, and as a function of the gender composition. The R&R rate as a function of recommendation is the same across the author-gender groups, even though the female-authored papers have higher citations, controlling on the other features of the paper (Figures 6a-b).

Published Papers. To further probe this result, we check a prediction of the model, which is that the estimated citation discounting for all-female papers should remain similar, if less precisely estimated, among papers which are ultimately published, given that the editors do not undo the referee bar. In Table 7 we re-estimate our benchmark citation specification (reproduced in Column 1) for the subset of papers that receive an R&R. We estimate a similar, if less precisely estimated, coefficient of of 0.26 log points (s.e.=0.13, Column 2). Finally, we obtain a similar result in the subset of papers that are accepted for publication within the time frame of our data (Column 3).

We can take this test one step further by considering, like CDV also did, the citations for published articles in the four journals in our sample for the years 2008-15, broadly corresponding to our submissions under the assumption of a 2-year delay between submissions and publication. In this sample of 1,530 published papers obtained from EconLit, we code the author gender, the field, the number of authors, and the author publications as in our sample, even if we of course cannot control for reviewer recommendations. Column 4 shows that we find a similar impact, at 0.39 (s.e.=0.15), as for the R&R papers in our Editorial Express data. In this sample, we find a sizable log point penalty for all-female papers, at 0.30 (s.e.=0.15) even if we do not control for author publications.

Implications. How large of an impact does this deviation from proportionality have for female authors? The baseline R&R rate for all-female papers is 12.2 percent in the sample of non-desk-rejected papers (Table 1, Column 8), also shown in the first bar of Figure 7. We can simulate the counterfactual R&R rate under the assumption that the editor in the R&R probit assigns the citation-maximizing weight on the gender variable.

We stress an important assumption. Under this counterfactual, the editor corrects the deviation from citation maximization with respect to the gender variables, but *not* with respect to other variables, such as the under-weighting of the author publication variables. This is the relevant counterfactual under two scenarios. First, the deviation from citation-maximization for the other variables may reflect biases in the citation-quality relationship, e.g., papers by more-published authors accrue more citations for given quality; the editor in this case is already quality-maximizing with respect to these variables. (As we discussed above, gaps between citation and quality likely run the other way for the author gender variables.) Second, the editor may follow a policy to support certain papers, e.g., papers by less published authors so as to facilitate the paths of these researchers. In this case again, the editor does not want to alter this policy with respect to these other variables. (In contrast, the editor presumably does not intend to set a higher bar for underrepresented groups.)

To compute the counterfactual, we start from the predicted R&R probability for the model in Column (7) of Table 5, $\Phi(X\hat{\beta})$, which matches, by design, the R&R rate of 12.1% in the subsample of female-authored papers. We then add the citation-based correction, $\hat{C} = 0.22 * 2.5 - 0.01 = 0.54$, just for the all-female papers. We thus compute $\Phi(X\hat{\beta}+\hat{C}d_F)$, where d_F is an indicator for all-female papers. The R&R rate for all-female papers would increase to 19.1%, as the second bar in Figure 7 shows. This is a large impact, a 7 percentage point and a 56 percent increase. If we also include the estimated citation-quality bias, the alternative correction is $\hat{C}' = (0.22 + 0.06) * 2.5 - 0.01 = 0.69$, leading to a counterfactual R&R rate of 21%, a 72 percent increase.

We can similarly compute a counterfactual for the mixed-author papers, leading to an increase in the R&R probability of only about half a percentage point. In the last bars in Figure 7, we average across the groups of papers and ask: for a female economist, taking into account that some of her papers with be in the mixed-gender category and some in the all-female category, how much is the R&R rate affected by the mechanism we point to? On average, the R&R rate would increase from 14.9% to 16.5%, an 11 percent increase. Thus, the mechanism which we point to in this paper could be a sizable determinant of the quality of publications for female economists.

A caveat is that these counterfactual R&R rates would lead to a slight increase in the overall number of R&Rs, given that we are holding the R&R rates for all-male authored papers (the omitted category) constant. While most journals are able to scale up, or down, the number of articles published, other journals are constrained by space. We can compute the counterfactual R&R rates, keeping the overall R&R rate constant, by adjusting the bar in the R&R probit regressions for all papers. The white lines in Figure 7 display the adjusted counterfactuals. The R&R rate for all-female authored papers would increase from 12.2% in the data to 18.3%, instead of 19.1%. The overall R&R rate for female authors would increase from 14.9% to 15.9%, instead of 16.5 %.

Robustness and Heterogeneity. We also consider the robustness of these results to different controls, as well as the heterogeneity, just as we did for the citation specifications. In particular, we examine (i) the impact of different sets of control variables, as well different specifications for the author prominence (Online Appendix Table 2); (ii) the results splitting by the number of authors of a paper (Online Appendix Table 3); (iii) the results estimating separately by submission years, by author publications, by share of women in the field, by number of referees, and by share of female referees (Table 6). Across the large majority of specifications, we replicate the key finding that the editors do not put a statistically significant weight on the author-gender mix, despite the fact that it is a variable that predicts citations. To illustrate this, in Column 8 of Table 6 we present for each specification the implied counterfactual R&R rate if the R&R decision gave the citation-maximizing weight to the all-female papers, comparing it to the empirical R&R rate (Column 7).[31] Across all specifications, the counterfactual would increase the R&R rate for all-female-authored papers, though the magnitude of the increase is larger for some groups and smaller for others.

6.2 Desk-Rejection

In the previous Section we show that editors take the recommendations as given and do not adjust the R&R bars for the higher citations of female-authored papers. An interpretation of these results is that the editors defer, perhaps too much, to the referees, as they do also in other dimensions. To provide evidence on editorial decisions with no input from referees, we turn to the desk-rejection decisions. At the initial submission stage, the editors have to rely just on their own private signal, as well as on observable features of the paper. Thus, this allows us to get a more direct glimpse into the editor beliefs and preferences.

Using the full sample of 29,890 submissions, we compare predictors of citations with predictors of the decision to not desk reject (NDR) in Table 8. As Column 1 shows, at the submission stage female-authored papers have 24 log points higher citations than submissions by all-male authors, holding constant other paper and author characteristics. This mirrors the citation result in the R&R regression in Table 5, though this specification uses all the 29,890 submissions.

Column 2 reports the estimates of a probit for the NDR decision. Interestingly, editors *do* take into account the author gender and are more likely to not desk-reject papers by female authors, holding all else constant. In this specification, it is not obvious how to estimate whether this is the optimal weight, given that we do not observe a variable like the referee recommendations.

In order to estimate the optimal weight, we build on a result in CDV: if the editors are putting the optimal weight on a variable X in their desk-reject decision, the variable X should not predict citations any more, once one controls for (a function of) the probability of desk-rejection. Thus, we create the predicted $P(NDR)$ based on Column 2, and we re-estimate a citation specification including a cubic polynomial in $P(NDR)$. In this specification we cannot include all the control variables, otherwise there will be essentially no identification left in the $P(NDR)$ cubic, but we do include at least the author publication variable since CDV show that it is a strong predictor of citations, even controlling for the $P(NDR)$ polynomial. In Column 3 we include just this variable, while in Column 4 we also include journal-year fixed effects and controls for the female share in the sub-field of the paper. Under either specification, we estimate a smaller, but still sizable, estimate for the all-female-authored papers, 0.17 (s.e.=0.05) in Column 3 and 0.15 (s.e.=0.04) in Column 4, compared to 0.24 in Column 1. Thus, the editorial desk-reject decision reduces the difference in

[31] These counterfactuals do not include the citation-quality gap of 6 log points.

citations by about a third, implying that the editors are only partially responding to the quality difference at initial submission between female-authored papers and male-authored papers.

6.3 Weight Placed on Referee Recommendations

As a third editorial decision, we switch from considering the impact of author gender and consider instead the referee gender. We study how much editors follow the referee recommendations by different groups of referees, following on gender differences. CDV find that referees with more recent publications appear to be equally informative about the quality of papers (measured with future citations) as referee with fewer recent publications; however, the editors follow the recommendation of the more-published referees 20 percent more. Is there a similar difference by gender?

Figure 8a shows the informativeness (i.e. relationship between referee recommendations and citations) of male and female referees, paired with information on their prominence (3 publications or more are classified as "prominent"). The crucial element is the *slope* of these lines. Male and female referees do not seem to differ in their informativeness and, consistent with results from CDV, more and less prominent referees do not differ in their informativeness either. Figure 8b shows how much weight the editor gives to recommendations from referees, who differ in prominence and gender. Consistent with the results in CDV, more prominent referees are valued relatively more, but gender does not seem to have much of an effect. Only for prominent referees, recommendations from male referees seem to be slightly more valued than recommendations from female referees.

To estimate these patterns with controls, we consider a nonlinear model:

$$\begin{aligned} Outcome_i = \sum_{j=1}^{N_{referees,i}} & (\alpha_0 Female_{ij} + (1 + \alpha_1 Female_{ij}) \times \\ & (\beta_{DefReject} DefReject_{ij} + \cdots + \beta_{Accept} Accept_{ij})/N_{referees,i}) + \boldsymbol{\gamma} \boldsymbol{X_i} + \varepsilon_i \end{aligned}$$

where $Outcome_i$ denotes paper i's outcome (receiving an R&R decision or $asinh(citations)$). This specification takes the specific recommendation of referee j for paper i, and assigns it an index value

$$R_{ij} = \beta_{DefReject} DefReject_{ij} + \cdots + \beta_{Accept} Accept_{ij}$$

with the same coefficients β_c for each recommendation type, regardless of the gender of the referee (or of the author team). It assumes that the outcome is affected by an average of "adjusted indexes":

$$\sum_j (\alpha_0 Female_{ij} + (1 + \alpha_1 Female_{ij}) R_{ij})/N_{referees,i}$$

where $Female_{ij}$ is an indicator for the gender of referee j of paper i and $N_{referees,i}$ is the number of referees who evaluate paper i. Such a specification effectively adjusts the index R_{ij} for both an intercept difference and a slope factor. If for example, female referees are more positive, we would expect to estimate a negative value for α_0 in the citation regression. Likewise, if female referees compress their recommendations in a narrower range then male referees, then we expect $\alpha_1 > 1$.

Table 9 shows the results. Columns 1-3 have informativeness as measured by citations as a dependent variable, and Columns 4-6 the editor's R&R probability. Female and male referees do not differ in their informativeness, and editors do not put different value on them either. Consistent with CDV, more prominent referees do not provide more informed recommendations (compared to less prominent referees), but nevertheless, the editors do value them more.

What do economists expect about the informativeness of male and female referees and how editors value them? In the survey, the large majority of economists expect male and female referees to be equally informative, and expect them to be followed equally, consistently with the data.

7 Delays and Other Outcomes

So far we have focused on the referee recommendations and editorial decisions, but another relevant dimension is the speed of decision-making. If there was discrimination against female authors, it may appear in the form of slower decisions (see Hengel, 2017 for papers published in *Econometrica*). We thus consider various outcome variables related to referee's and editor's speed.

Figures 9a-b show referee response time (Figure 9a) and editorial response time, from receipt of the last report to the editorial decision (Figure 9b). The sample is 4,341 non-desk rejected papers with both male and female referees.[32] If anything, all-female papers get quicker responses than all-male papers from both referees and editors. As far as the gender of the referee, female and male referees show no differences in their response time. Finally, Figure 9c shows the number of rounds by different gender compositions for those papers that got a R&R decision. Here, female papers do show slightly longer number of rounds, but the differences are tiny and statistically insignificant. In any case, it will be important to control for the referee recommendations, authors' and referees' characteristics and the final outcome of the paper in regression analysis.

Before we turn to the referee and editorial delays, in Columns 1-3 of Table 10 we consider the propensity of referees to accept a referee invitation. Each observation is a referee request for papers that were not desk-rejected. Once we include controls (Column 2), we estimate no difference in the probability of accepting a referee invitation for all-female-authored papers, and no evidence that female referees are more, or less, willing to provide referee reports. With paper fixed effects (Column 3) we find no difference by reviewer gender, and no relative difference depending on whether the reviewer gender matches the author gender.

In Columns 4-6 we consider the number of days from paper submission to the reception of a referee report for reviewers who return a report. With controls for paper and author characteristics (Column 5), the gender composition of the submitted papers does not affect the response time of referee reports. Also, female referees do not seem to be faster or slower than male referees, irrespective of the gender composition of the authors (see Column 6).

Building on the results in Table 10, in Table 11 we present a range of decision time measures for non-desk rejected papers (Columns 1-3) and papers who received an R&R (Columns 4-6). Each observation is a paper, and all regressions control for the full set of controls, including editor fixed

[32] Referees reports are weighted at the paper level to keep constant the share of female referees across papers.

effects. We detect no difference based on the gender mix of the authors on the number of days from the paper submission to the arrival of the last report (Column 1), from the arrival of the last report to the editor's decisions (Column 2), and on the overall number of days (Column 3). For the papers with an R&R, we similarly find no impact of the author gender mix on the number of rounds of revision (Column 4), on the time that the authors take to submit the first revision (Column 5), and the number of days from the resubmission until the final acceptance (Column 6).

Finally, motivated by Hengel (2017)'s measures of abstract complexity for papers, we also test in Online Appendix Table 7 if the gender composition of the submitted papers affects the complexity level of the abstracts. A caveat is that we only observe the abstract of the most recent version of the paper, and thus we cannot examine, as in Hengel (2017), the change in complexity from the submission to the published version of the paper. Instead, we look at two different samples: the papers that were desk rejected or rejected after being considered for publication (Columns 1-2), and the papers that received a R&R decision (Columns 3-4). Measuring the complexity of the abstract with the Gunning Fog (Columns 1 and 3) and Coleman-Liau (Columns 2 and 4) measures, we find no impact of the author gender mix on the readability of the abstract in either sample.

8 Discussion and Conclusion

Are the referees and editors in economics gender neutral? The answer is "Yes" or "No", depending on which outcomes we consider in the editorial process.

If we focus on key comparisons highlighted in previous research, we obtain relatively precise zero differences between gender groups. Considering delays in publication, as Hengel (2017) does, we find no differences by author gender. Considering referee recommendations, we replicate the findings of Abrevaya and Hamermesh (2012) that there are no differences in the average evaluations of papers by male and female authors, and no difference in how referees of different genders assess papers by female and male authors. Turning to editorial decisions, which have not been previously studied, we find that editors are gender-blind in the sense that they treat female- and male-authored papers the same, conditional on the referee recommendations. Further, editors give about the same weight to recommendations of male referees and female referees, which is appropriate given that the two groups are equally informative.

Yet, the editorial process does not appear to be gender-neutral once we take into account underlying differences in paper quality, as revealed by future citations. Female-authored papers get 25 percent *more* citations than male authored papers, controlling for other paper features including field and the authors' previous publication record. This estimate is relatively precisely estimated ($t > 4$) and is robust to a number of alternative specification choices. Given that any bias in citations as a measure of quality is likely to work against female authors, we interpret this finding as evidence that female researchers are held to a higher bar by referees (both male and female). Since editors do not adjust their thresholds for this higher bar, they effectively reject too many female-authored papers relative to a citation-maximizing benchmark. Interestingly, the citation gap of 25 percent in reviewer evaluation matches a quality difference at submission: controlling for features of the paper and of the authors, papers by all-female author teams have 27 percent higher citations *at submission.*

What accounts for these patterns? While we do not have direct evidence, we envision three main explanations. First, our findings are consistent with the presence of some discrimination towards female economists, which would explain not only the high bar imposed by referees, but also the higher quality of initial submissions by female economists, since female economists would need to be of higher quality to reach a given level of previous publications.

Second, it could be that female economists submit papers with somewhat different characteristics than those of male authors, such as a different combination of substantive and methodological contributions, that are under-valued by referees relative to their impact on longer-run citations. Under this interpretation, referees are not discriminating on the base of gender, but with respect to characteristics associated with the author gender mix.

A third possibility is that female economists wait longer for submission, leading to their paper accumulating more citations initially, conditional on quality. This explanation, unlike the previous two, holds that the observed patterns are due to a bias in observed citations that *favors* female economists for given paper quality. This explanation appears implausible given that we find the largest citation gaps for submissions from the earliest years of our sample, for which we can measure citations at least 5 years after submission.

Where does that leave us in terms of implications for our profession? Our finding that female authors are held to a higher standard is concerning. We estimate that as a result the R&R rate for female-authored papers is about 7 percentage points lower than the rate consistent with a gender-neutral citation-maximizing rule. This gap suggests an important hurdle, aside from the assignment of credit in coauthored work stressed by Sarsons (2018), for junior female economists, as well as a continuing obstacle to career progression for more senior female researchers.

One potential remedy to help female economists – using more female referees – is unlikely to help, given that female referees appear to hold female-authored papers to the same higher bar as male referees. Recruiting more female referees may only have the unintended consequence of requiring more public good provision by female economists (Babcock et al., 2017).

It appears to us that a simpler path is to increase the awareness of the higher bar for female-authored papers. The referees and editors can then take it into account in their recommendations and decisions. This would address the bias, whether its source is gender discrimination (the first explanation) or undervalued paper characteristics (the second explanation). In contrast, a policy of double-blind evaluation, setting aside implementation difficulties, would address only the first source of bias.

It would be great to revisit our analysis in 3 to 5 years to test whether the gender difference has been corrected. Perhaps, as for NBA referee bias (Price and Wolfers, 2010), publicizing the findings may be enough to correct the pattern (Pope, Price and Wolfers, forthcoming).

This is just an example of the importance of data transparency in the editorial process, as CDV also stress. Indeed, we are grateful to the four journals in economics who agreed to such data transparency, something with very few parallels outside economics. Keeping track of such data for the future should make it relatively straightforward not only to check for progress on any form of gender bias, but also more generally to help make the editorial process fairer and more efficient.

References

Abrevaya, Jason and Daniel S. Hamermesh. 2012. “Charity and Favoritism in the Field: Are Female Economists Nicer (To Each Other)?.” *Review of Economics and Statistics*, 94(1): 202-207.

Altonji, Joseph G., Todd E. Elder, and Christopher R. Taber. 2005. “Selection on Observed and Unobserved Variables: Assessing the Effectiveness of Catholic Schools.” *Journal of Political Economy*, 113(1), 151-184.

Azmat, Ghazala, and Rosa Ferrer. ”Gender gaps in performance: Evidence from young lawyers.” *Journal of Political Economy* 125.5 (2017): 1306-1355.

Babcock, Linda, Maria Recalde, Lise Vesterlund, and Laurie Weingart. 2017. “Gender Differences in Accepting and Receiving Requests for Tasks with Low Promotability,” *American Economic Review*, 107(3): 714-747.

Bayer, Amanda, and Cecilia Elena Rouse. 2016. “Diversity in the Economics Profession: A New Attack on an Old Problem.” *Journal of Economic Perspectives*, 30(4): 221-42.

Bertrand, Marianne, and Kevin F. Hallock. 2001. ”The gender gap in top corporate jobs.” *Industrial and Labor Relations Review* 55.1: 3-21.

Blank, R.M., 1991. ”The effects of double-blind versus single-blind reviewing: Experimental evidence from the American Economic Review”. *The American Economic Review*, pp.1041-1067.

Bohren, J. Aislinn, Alex Imas, and Michael Rosenberg. 2018. “The Dynamics of Discrimination: Theory and Evidence,” Working paper.

Bordalo, P., Katherine Coffman, and Nicola Genaioli. Forthcoming. “Beliefs about gender”. *American Economic Review.*

Bransch, F., and M. Kvasnicka, 2017, “Male Gatekeepers Gender Bias in the Publishing Process?”. IZA DP No. 11089.

Broder, I.E., 1993. “Review of NSF economics proposals: Gender and institutional patterns” *The American Economic Review*, 83(4), pp.964-970.

Card, David and Stefano DellaVigna. Forthcoming. “What do Editors Maximize? Evidence from Four Economics Journals.” *Review of Economics and Statistics.*

Ceci, Stephen J., Donna K. Ginther, Shulamit Kahn, and Wendy M. Williams. “Women in Academic Science: A Changing Landscape.” *Psychological Science in the Public Interest*, 15(3): 75-141.

Chari, Anusha and Paul Goldsmith-Pinkham. 2017. “Gender Representation in Economics Across Topics and Time: Evidence from the NBER Summer Institute.” NBER Working Paper No. 23953.

Christensen, Garret S., and Edward Miguel. 2018. “Transparency, Reproducibility, and the Credibility of Economics Research.” *Journal of Economic Literature,* Vol. 56, 920-980.

DellaVigna Stefano and Devin Pope. 2018. “What Motivates Effort? Evidence and Expert Forecasts.” *Review of Economic Studies*, Vol. 85, 1029–1069.

Dolado, Juan J., Florentino Felgueroso, and Miguel Almunia. ”Are men and women-economists evenly distributed across research fields? Some new empirical evidence.” SERIEs 3.3 (2012): 367-393.

Donald, Stephen, G., and Daniel S. Hamermesh. 2006. ”What is Discrimination? Gender in the *American Economic Association*, 1935-2004.” *American Economic Review*, 96 (4): 1283-1292.

Ginther, Donna, K., and Shulamit Kahn. 2004. ”Women in Economics: Moving Up or Falling Off the Academic Career Ladder?” *Journal of Economic Perspectives*, 18 (3): 193-214.

Goldin, Claudia. “A grand gender convergence: Its last chapter.” *American Economic Review* 104.4 (2014): 1091-1119.

Hengel, Erin. 2017. “Publishing while Female. Are Women Held to Higher Standards? Evidence from Peer Review.”

Lundberg, Shelly. ”Report: Committee on the Status of Women in the Economics Profession (CSWEP).” *American Economic Review* 107.5 (2017): 759-76.

Niederle, Muriel, and Lise Vesterlund. 2011. “Gender and competition.” *Annual Review of Economics*, 3.1: 601-630.

Pope, Devin, Joseph Price, and Justin Wolfers. Forthcoming. “Awareness Reduces Racial Bias”, *Management Science*.

Price, Joseph and Justin Wolfers. 2010. “Racial Discrimination Among NBA Referees,” *Quarterly Journal of Economics*, 125(4):1859-1887.

Reuben, Ernesto, Paola Sapienza, and Luigi Zingales. 2014. “How stereotypes impair women’s careers in science.” *Proceedings of the National Academy of Sciences*: 201314788.

Reuben, Ernesto, Paola Sapienza, and Luigi Zingales. 2015. “Taste for competition and the gender gap among young business professionals.” No. w21695. National Bureau of Economic Research.

Sarsons, Heather. 2018. ”Gender Differences in Recognition for Group Work.” Working Paper.

Wu, Alice. 2018. “Gender Stereotyping in Academia: Evidence from Economics Job Market Rumors Forum.”

Figure 1. Summary Statistics by Gender

Figure 1a. Distribution of Editorial Decisions

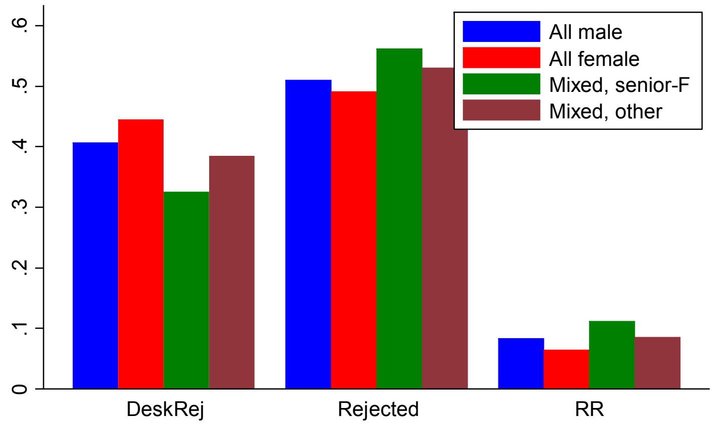


Figure 1b. Distribution of Ref. Recommendations

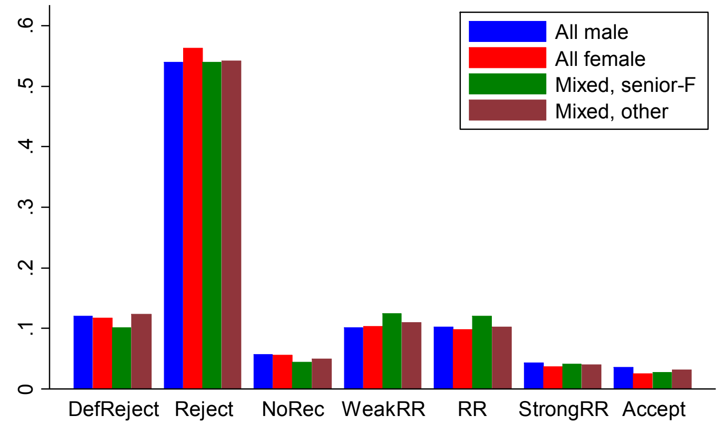


Figure 1c. Distribution of Author Publications

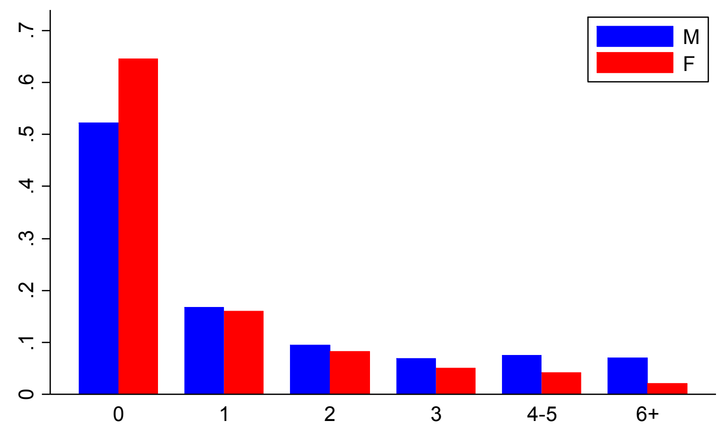


Figure 1d. Distribution of Referee Publications

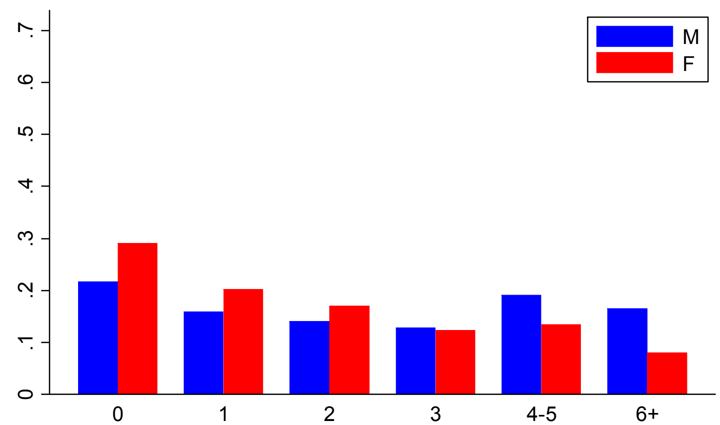


Figure 1e. Paper Citations by Gender

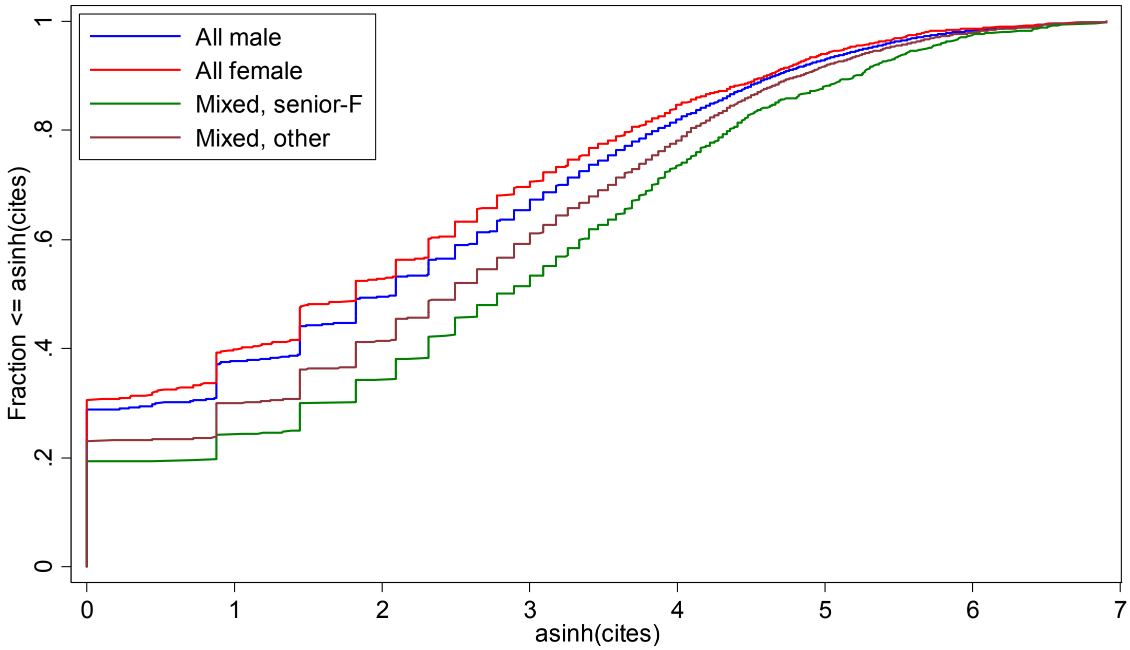
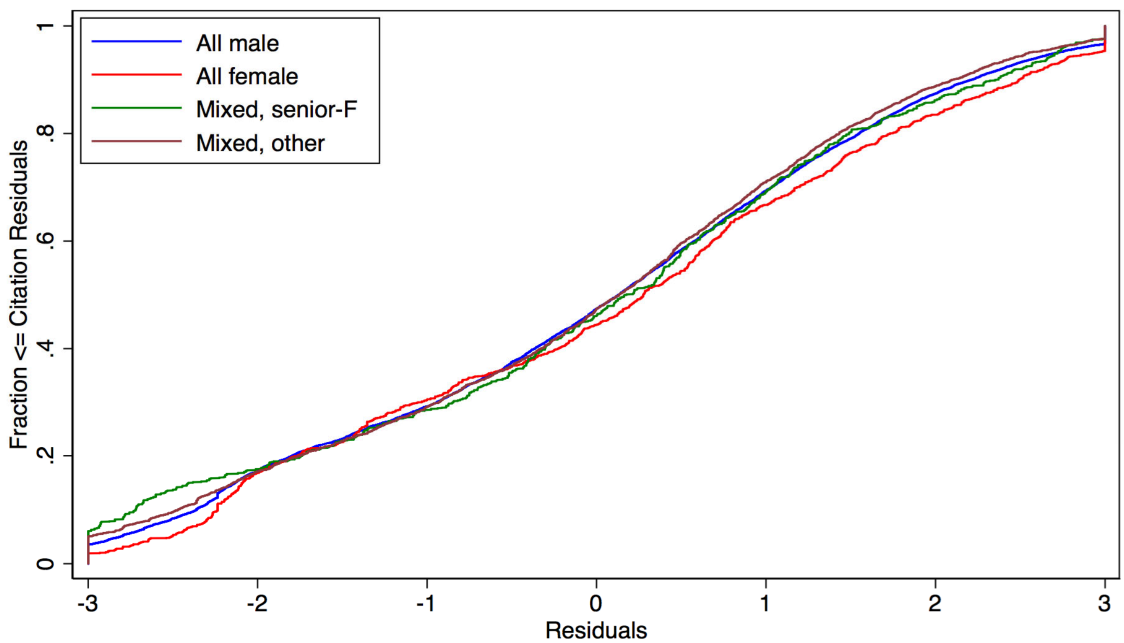


Figure 1f. Paper Citations by Gender, Citations Residualized



Notes: Figure 1 displays a few key summary statistics by gender. Figure 1a plots the distribution of the editor's decision and Figure 1b shows the distribution of referee recommendations. Figure 1c 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. Figure 1d reports the distribution of publications among referees by gender. Figure 1e displays the CDF of the (asinh of) paper citations. Figure 1f displays the same citation variable, but after partialling out the key controls for journal-year fixed effects, fields, number of authors, and number of author publications (as in Table 5, Column 3).

Figure 2. Share of Female Authors and Referees, by Field

Figure 2a. Author Gender by Field

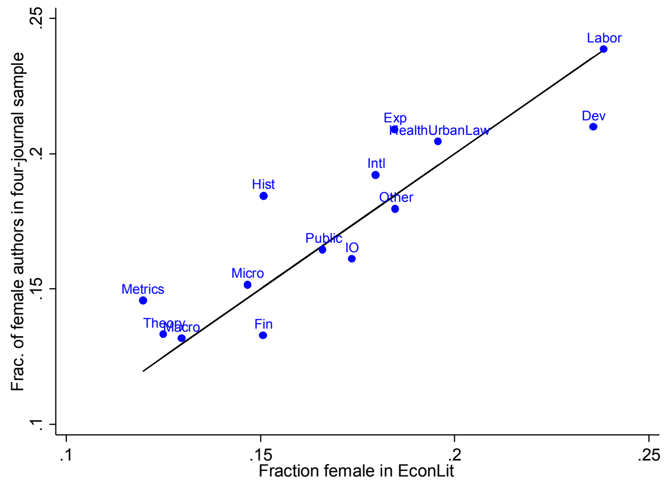


Figure 2b. Referee Gender by Field

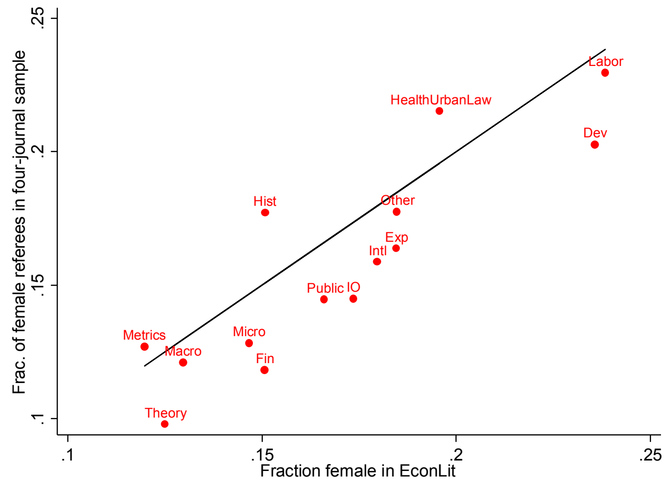
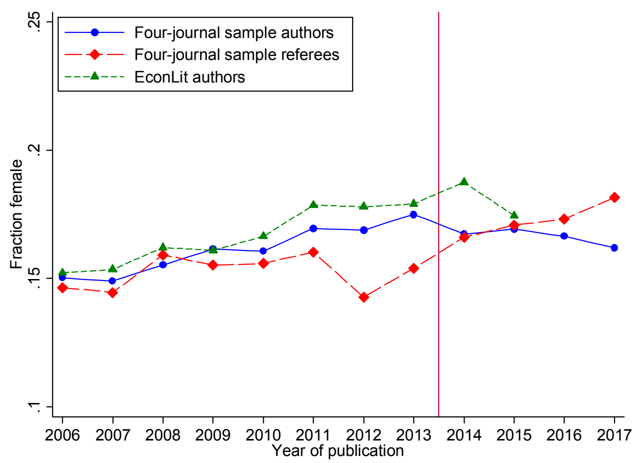
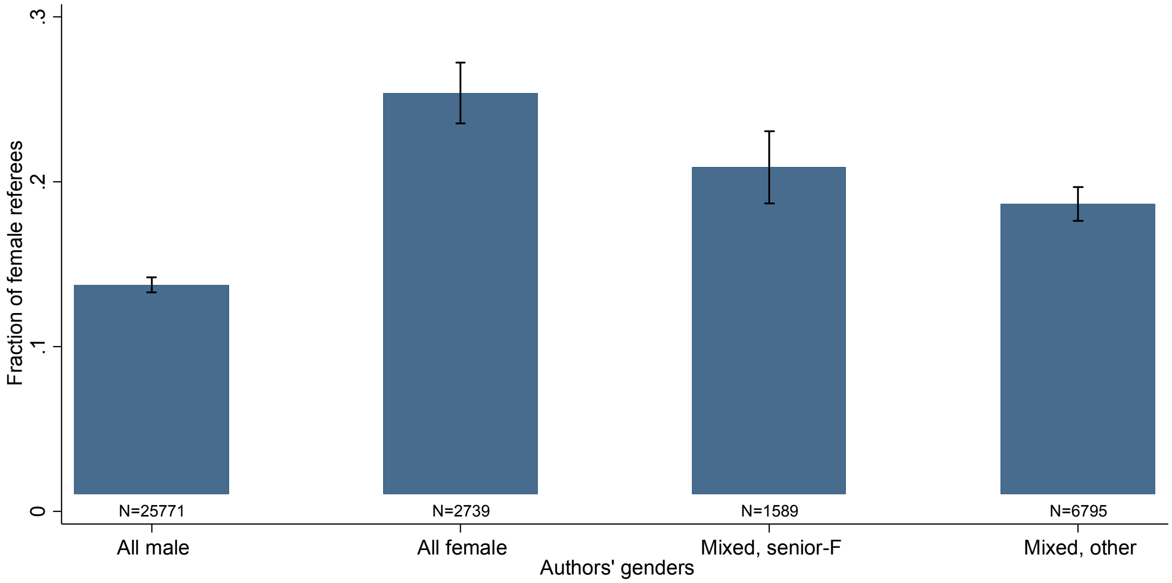


Figure 2c. Author and Referee Gender Over Time



Notes: Figures 2a and 2b show the average fraction of female authors and referees in the years 2006-13 in the four-journal sample and the years 2008-15 in the EconLit sample; the 2-year offset makes the timing in the two samples more comparable. In Figure 2c, EconLit observations are shown lagged two years to match the submission to publication delay. Observations are at the author/referee-paper-field level and weighted to the paper level, i.e. by the inverse of number of authors/referees times number of fields. Fraction of female calculated after excluding unknown gendered individuals.

Figure 3. Referee Assignment: Referee Gender as Function of Author Gender



Note: Observations are at the referee-paper level.

Figure 4. Referee Evaluation by Author Gender and Referee Gender

Figure 4a. Index of Referee Recommendations

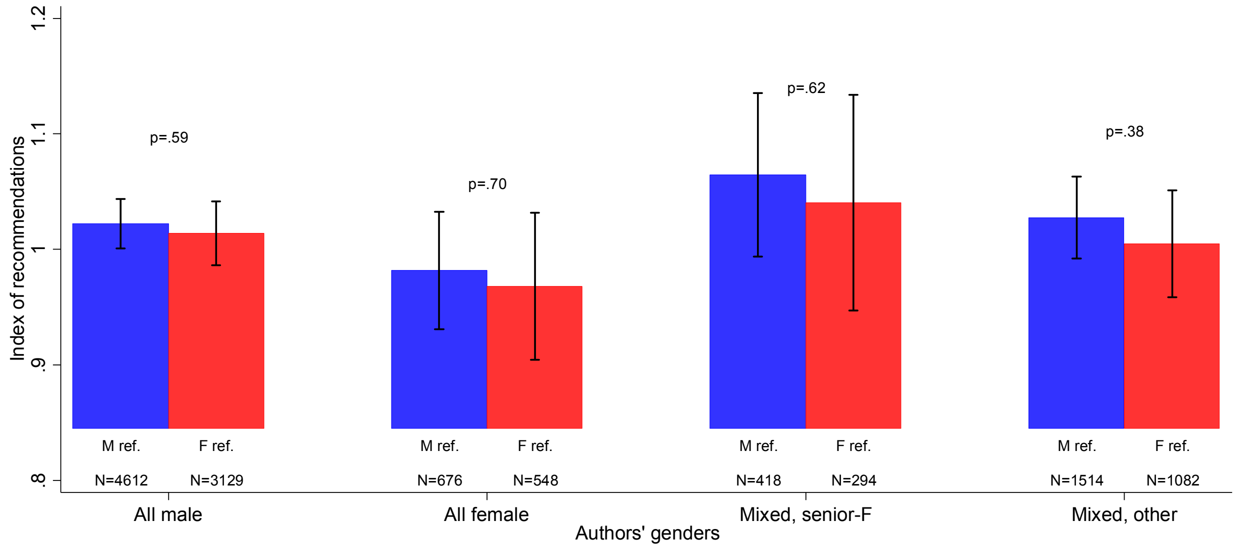
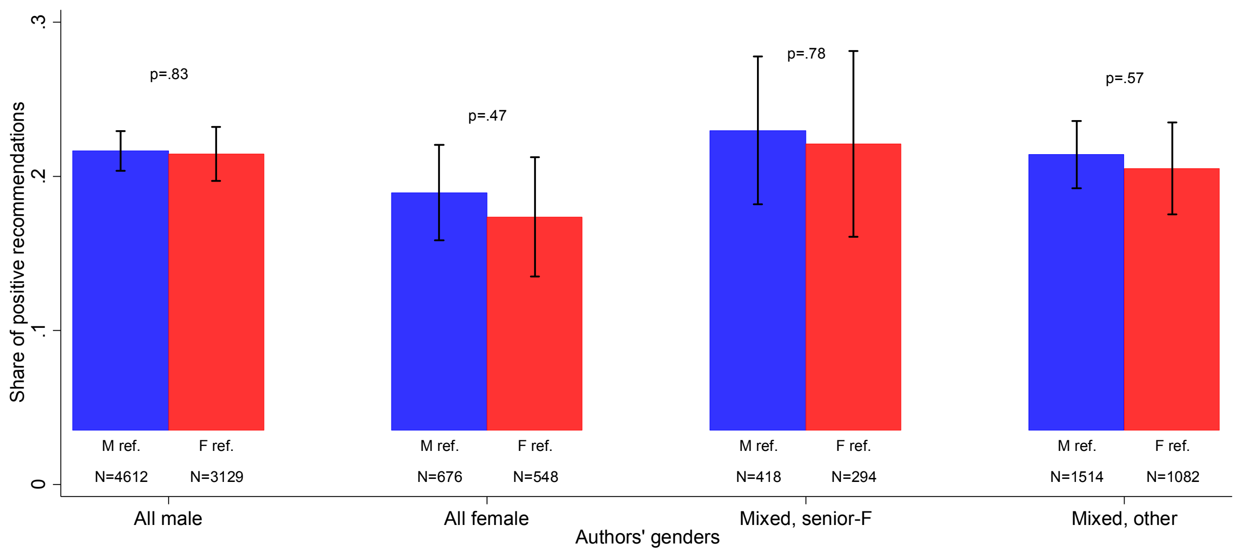
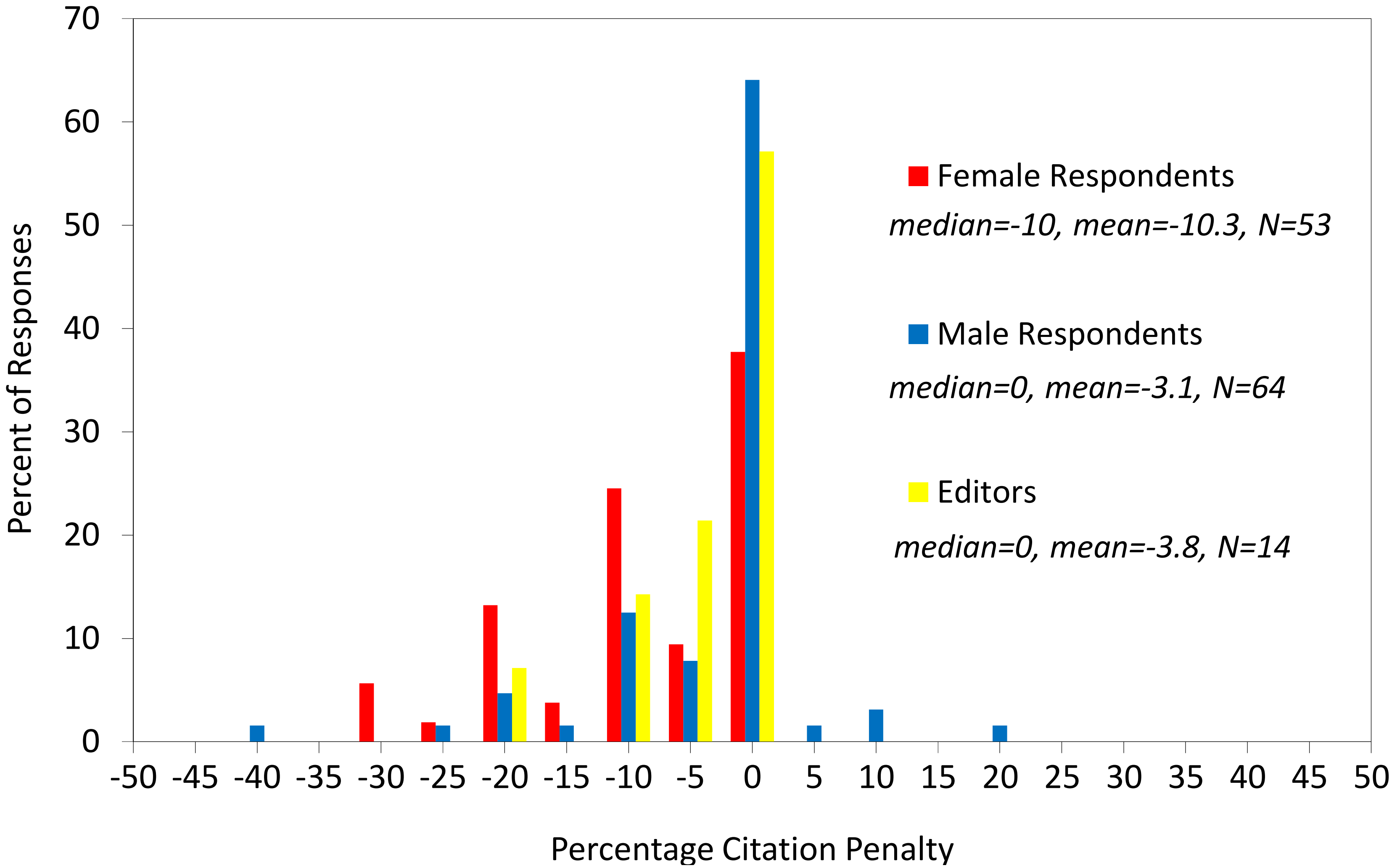


Figure 4b. Share of Positive Referee Recommendations



Notes: Figure 4a displays the mean recommendation given by referees based on gender. The index of referee recommendations is constructed using the coefficients in the cites model in Card and DellaVigna (2017). From Definitely Reject to Accept, the values are 0, 0.67, 1.01, 1.47, 1.92, 2.27, 2.33. The bands show 2 standard error intervals, clustered at the paper level. Includes only 4,341 papers with both male and female referees. Figure 4b shows the share of positive recommendations, defined as RR-Accept. In both panels, female referees are weighted at the paper level by N_{male} / N_{female}.

Figure 5. Citation Penalty for Female Authors: Survey Responses by Gender



Note: Tabulation of the response to question Q8 in the survey (Table 2). The number of observations differs from the one in Table 2 because some of the survey respondents did not answer question Q8.

Figure 6. Differences in (Residualized) Citations and R&R Rate, by Author Gender

Figure 6a. Referee Recommendations and Citations

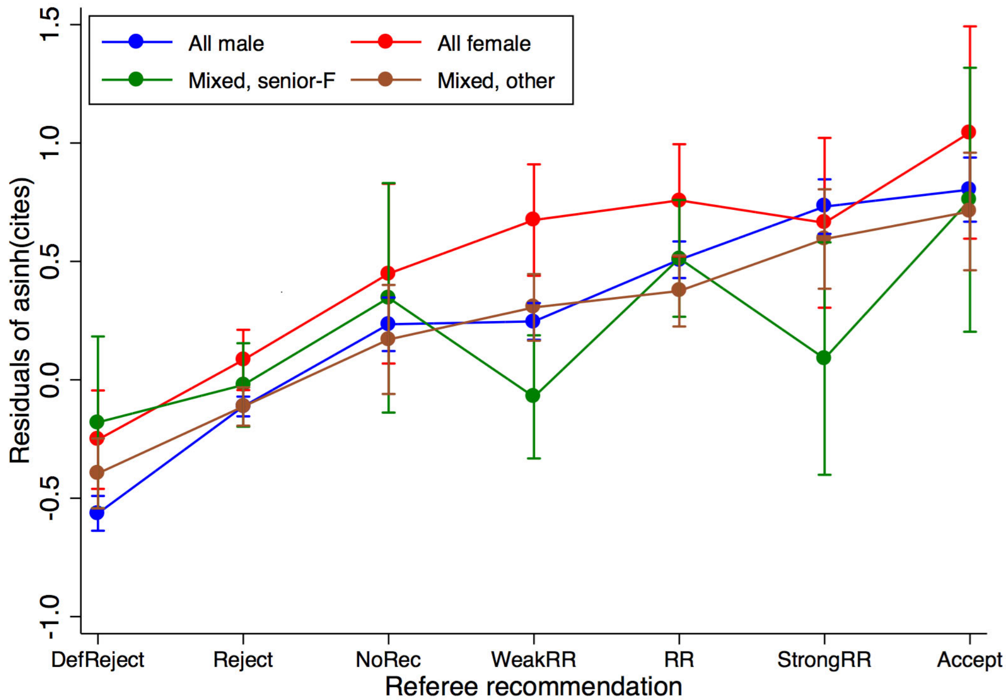


Figure 6b. Referee Recommendations and Citations, by Author Gender and Referee Gender

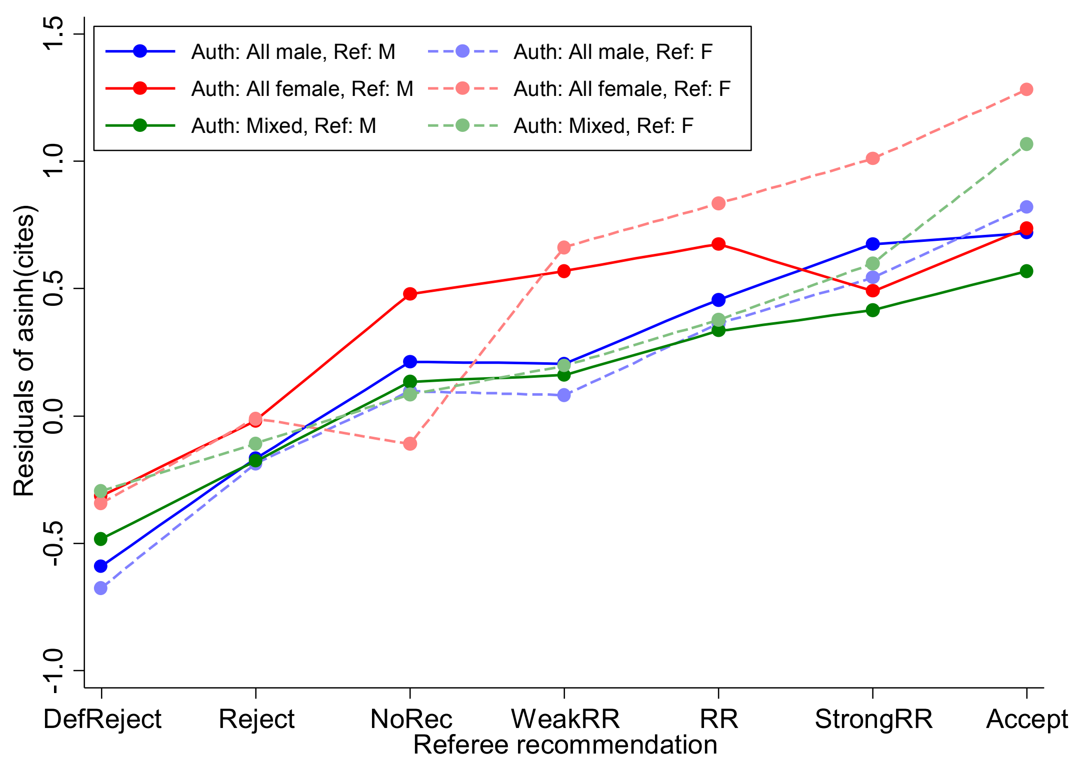
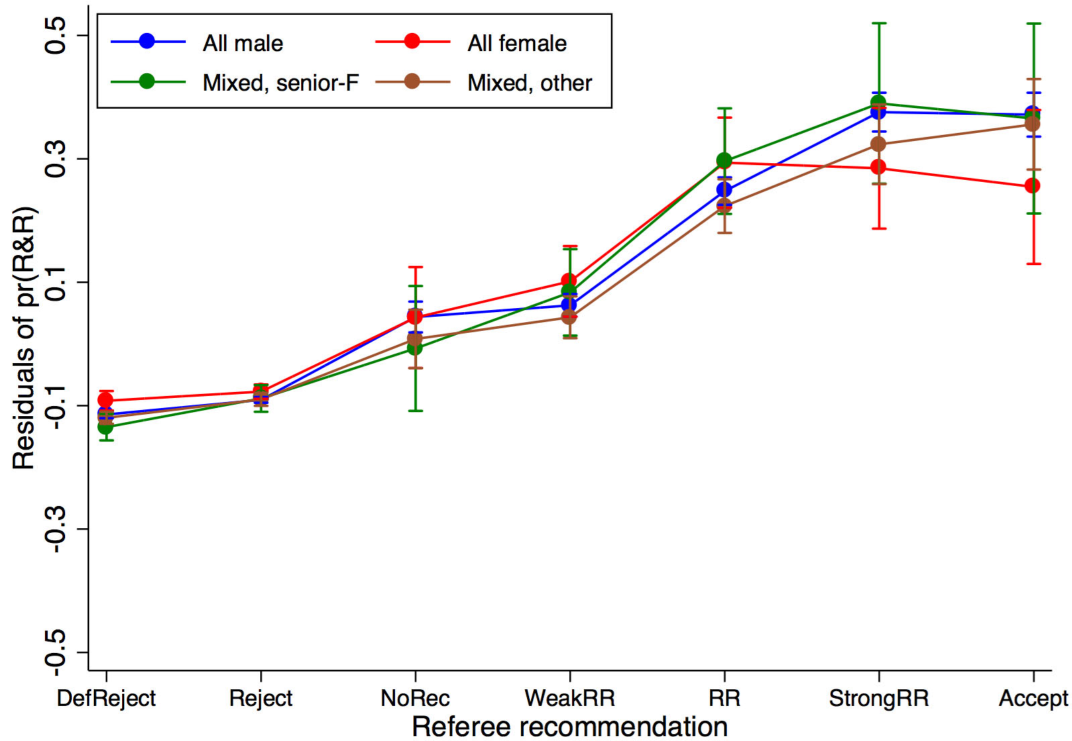
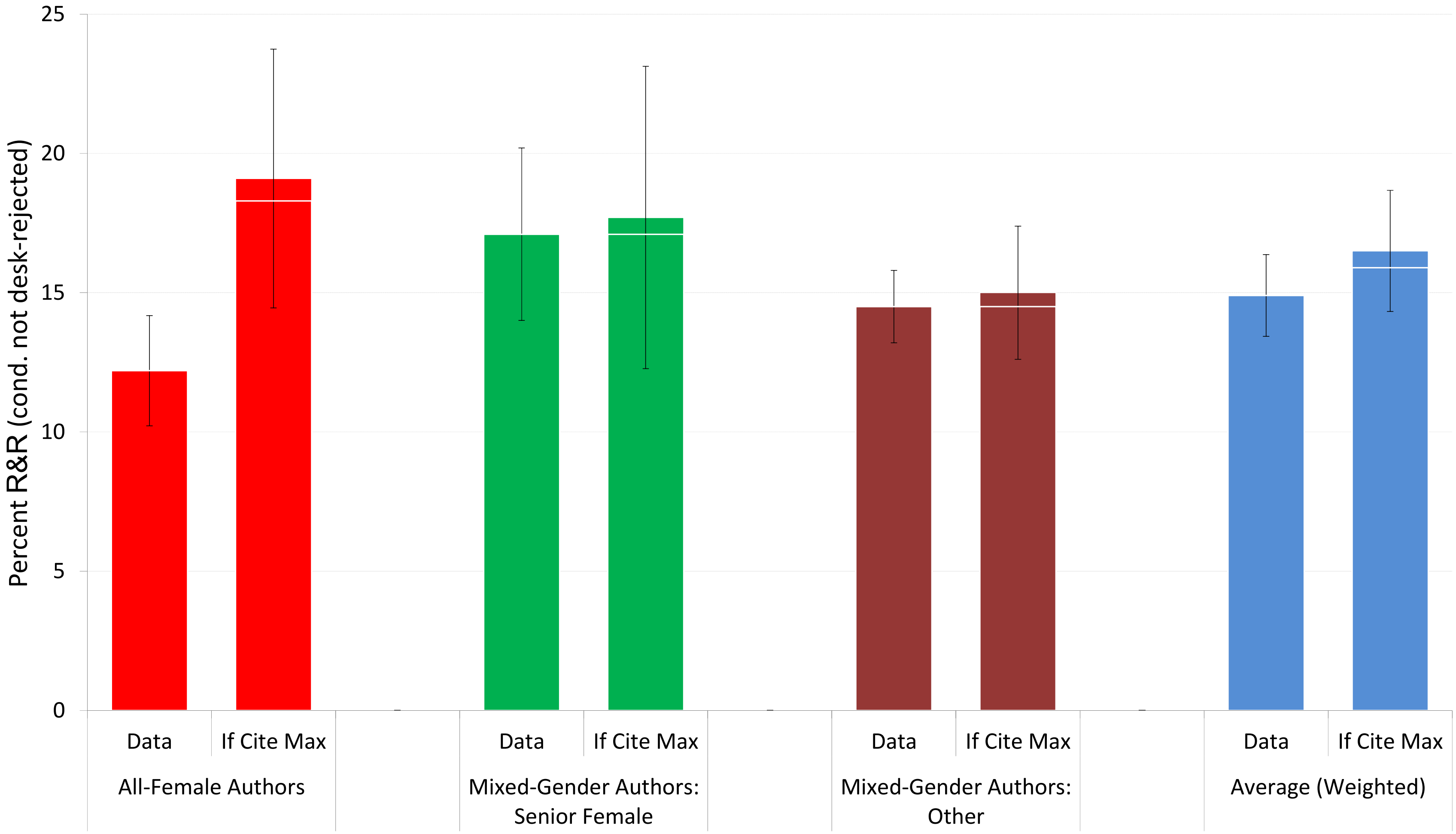


Figure 6c. Referee Recommendations and R&R Rate



Notes: Figures 6a and 6b show the weighted residuals of *asinh* (citations) for a paper receiving a given recommendation, while Figure 6c shows the residuals of pr(R&R) for a paper receiving a given recommendation. In both models, residuals are calculated by regression onto author publications, the number of authors, and fields. Figures 6a and 6c show the results separately by the gender break down of the author team. Figure 6b splits these two categories further into referees' gender. 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 6b omits confidence intervals for legibility.

Figure 7. Implication of Findings for R&R Rate for Female-Authored Papers



Notes: Figure 7 presents the implications of the results in Table 5 for the actual and counterfactual R&R rate for teams of authors. The figure breaks down papers by all-female authors, by mixed-gender teams with a senior female author and other mixed-gender teams. Within each group, the first bar plots the observed R&R rate for that group, conditional on a paper not being desk-rejected. The second bar reports the counterfactual R&R rate for that category of papers that we would expect to observe if the editorial process aimed to put the weight on the gender-author mix associated with citation maximization with respect to the author-gender mix variables (the details are in the text). This prediction is computed under the assumption that the journals can increase their R&R rate to accommodate the additional female-authored papers. The white line in the prediction indicates the level that would apply under a restriction that the overall R&R rate remains the same. The final set of columns presents an average over the first 3 sets of columns, weighting the different groups by the probability that a female author would have a paper in each category. The bars report 95% confidence intervals built from bootstraps.

Figure 8. Referee Informativeness, by Referee Gender and Publications

Figure 8a. Referee Recommendations and Citations

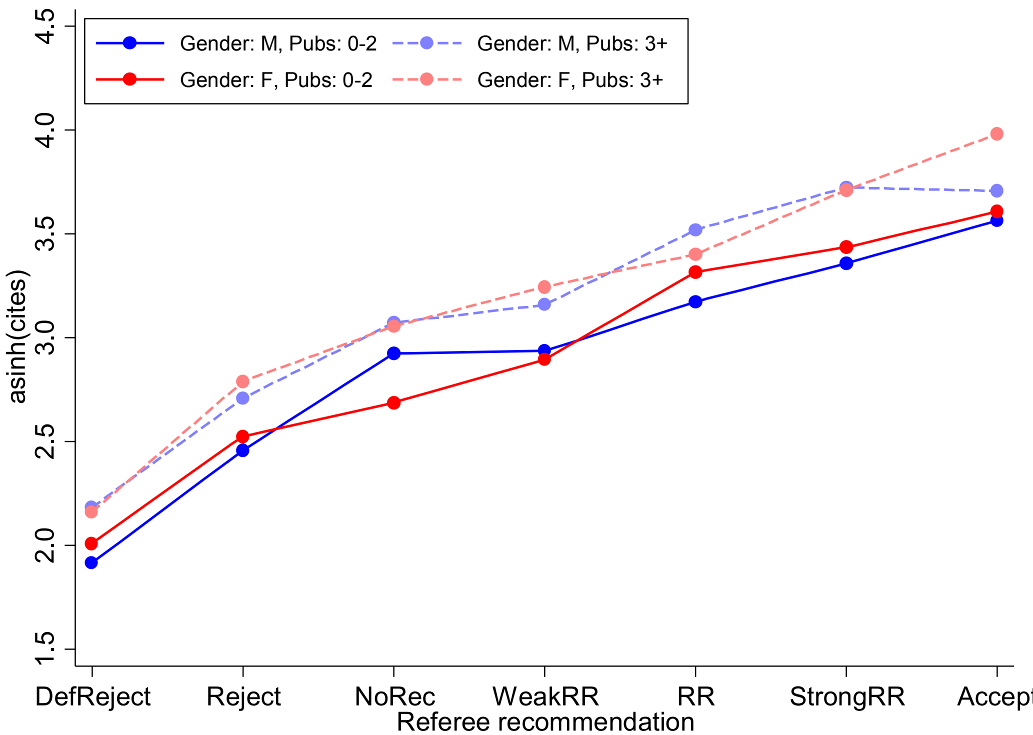
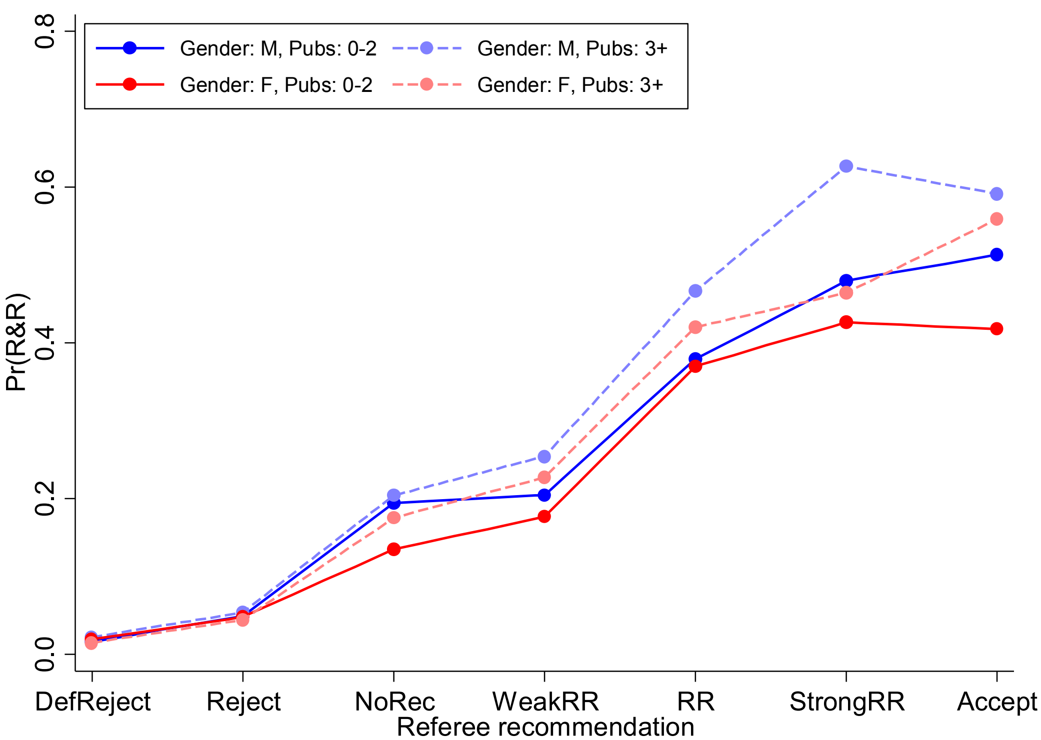


Figure 8b. Referee Recommendations and R&R Rate



Notes: Figure 8a shows the weighted *asinh* (citations) for a paper receiving a given recommendation. Figure 8b shows the R&R rate for a paper receiving a given recommendation. Both show the results separately by referee gender.

Figure 9. Other Editorial Outcomes: Referee and Editorial Delay

Figure 9a. Referee Response Time

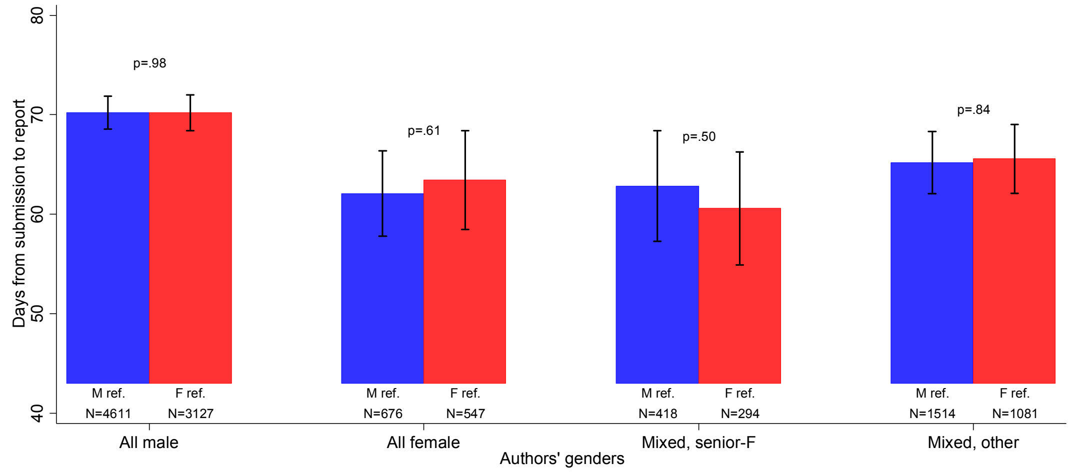


Figure 9b. Editor Response Time

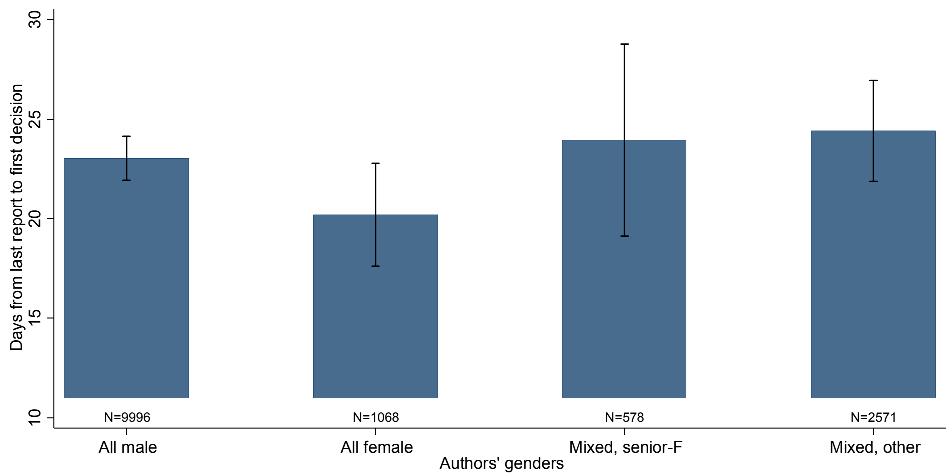
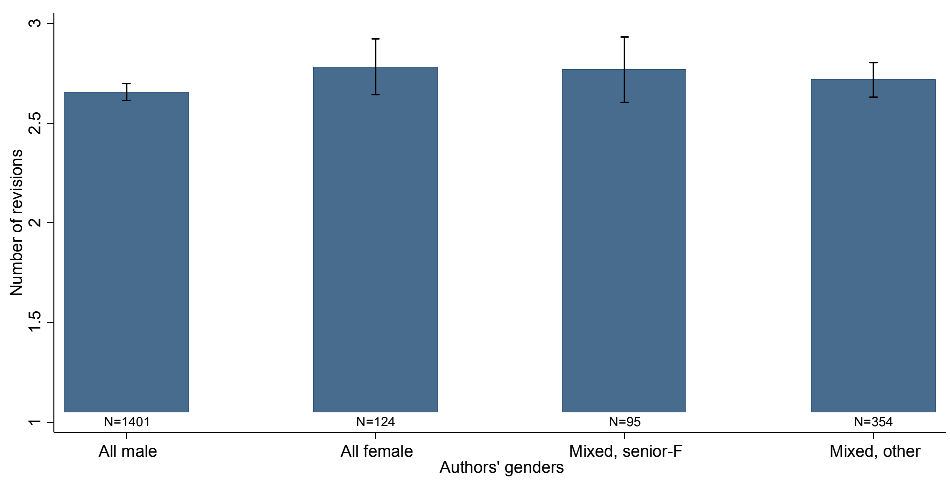


Figure 9c. Number of Rounds (for R&R papers)



Note: Figure 9a includes only 4,341 papers with both male and female referees. In Figure 9a, female referees are weighted at the paper level by N_{male} / N_{female}. Figure 9b omits papers when the editor decides before the last report arrives.

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

	All Papers						Non-Desk-Rejected Papers					
Sample:	All male	All female	Mix., F-led	Mix., other	Undet.	All	All male	All female	Mix., F-led	Mix., other	Undet.	All
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Google Scholar Citations												
Asinh Citations	2.11	1.97	2.72	2.41	1.27	2.11	2.72	2.61	3.09	2.92	2.15	2.74
	(1.83)	(1.80)	(1.85)	(1.82)	(1.57)	(1.83)	(1.84)	(1.85)	(1.81)	(1.81)	(1.77)	(1.84)
Editorial Decisions												
Not Desk-Rejected	0.59	0.56	0.67	0.61	0.41	0.58	1.00	1.00	1.00	1.00	1.00	1.00
Received R&R Decision	0.08	0.06	0.11	0.08	0.04	0.08	0.15	0.12	0.17	0.14	0.12	0.15
Authors' Genders												
All male	1.00	0.00	0.00	0.00	0.00	0.66	1.00	0.00	0.00	0.00	0.00	0.67
All female	0.00	1.00	0.00	0.00	0.00	0.08	0.00	1.00	0.00	0.00	0.00	0.07
Mixed, female-led	0.00	0.00	1.00	0.00	0.00	0.03	0.00	0.00	1.00	0.00	0.00	0.04
Mixed, other	0.00	0.00	0.00	1.00	0.00	0.16	0.00	0.00	0.00	1.00	0.00	0.17
Undetermined	0.00	0.00	0.00	0.00	1.00	0.07	0.00	0.00	0.00	0.00	1.00	0.04
Author Publications in 35 high-impact journals												
Publications: 0	0.46	0.69	0.00	0.33	0.70	0.46	0.32	0.59	0.00	0.22	0.53	0.32
Publications: 1	0.17	0.15	0.28	0.17	0.11	0.17	0.17	0.18	0.23	0.15	0.11	0.17
Publications: 2	0.10	0.07	0.24	0.13	0.07	0.10	0.12	0.09	0.22	0.13	0.11	0.12
Publications: 3	0.08	0.04	0.18	0.10	0.04	0.08	0.11	0.05	0.19	0.12	0.08	0.11
Publications: 4-5	0.09	0.04	0.17	0.12	0.04	0.09	0.13	0.07	0.20	0.16	0.08	0.13
Publications: 6+	0.10	0.01	0.13	0.14	0.04	0.10	0.16	0.02	0.17	0.22	0.09	0.15
Number of Authors												
1 author	0.44	0.76	0.00	0.00	0.34	0.37	0.36	0.72	0.00	0.00	0.26	0.31
2 authors	0.39	0.21	0.52	0.48	0.36	0.39	0.43	0.25	0.53	0.46	0.39	0.42
3 authors	0.15	0.03	0.39	0.39	0.23	0.19	0.18	0.03	0.37	0.40	0.27	0.22
4+ authors	0.03	0.00	0.09	0.13	0.07	0.05	0.03	0.00	0.10	0.14	0.08	0.05
Field of Paper												
Development	0.04	0.06	0.05	0.05	0.05	0.05	0.04	0.07	0.05	0.05	0.04	0.05
Econometrics	0.07	0.05	0.06	0.06	0.09	0.07	0.06	0.04	0.06	0.06	0.09	0.06
Finance	0.07	0.05	0.07	0.06	0.10	0.07	0.06	0.04	0.06	0.05	0.09	0.06
Health, Urban, Law	0.05	0.06	0.06	0.06	0.04	0.05	0.04	0.07	0.07	0.06	0.04	0.05
History	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.02	0.01	0.01	0.01	0.01
International	0.06	0.07	0.09	0.06	0.06	0.06	0.06	0.06	0.10	0.06	0.05	0.06
Industrial Organization	0.05	0.05	0.04	0.05	0.05	0.05	0.05	0.05	0.04	0.05	0.05	0.05
Lab/Experiments	0.02	0.02	0.04	0.04	0.01	0.02	0.02	0.02	0.04	0.04	0.02	0.03
Labor	0.10	0.17	0.13	0.13	0.07	0.11	0.11	0.20	0.14	0.15	0.10	0.12
Macro	0.11	0.08	0.08	0.08	0.13	0.10	0.11	0.06	0.08	0.08	0.10	0.10
Micro	0.11	0.08	0.10	0.11	0.08	0.11	0.11	0.08	0.09	0.11	0.09	0.11
Public	0.05	0.04	0.04	0.05	0.03	0.05	0.05	0.05	0.03	0.05	0.04	0.05
Theory	0.10	0.07	0.08	0.07	0.08	0.09	0.11	0.08	0.09	0.08	0.11	0.10
Unclassified	0.06	0.06	0.06	0.06	0.07	0.06	0.05	0.05	0.05	0.06	0.04	0.05
Missing Field	0.11	0.12	0.10	0.11	0.14	0.11	0.11	0.12	0.10	0.11	0.13	0.11
Gender-Field Variables												
Share female in fields	0.15	0.18	0.17	0.17	0.15	0.16	0.15	0.18	0.17	0.17	0.15	0.16
Gender-topic fields	0.03	0.08	0.05	0.05	0.03	0.04	0.03	0.08	0.06	0.06	0.04	0.04
Referee Recommendations												
Fraction Definitely Reject							0.12	0.12	0.10	0.12	0.19	0.12
Fraction Reject							0.54	0.56	0.54	0.54	0.51	0.54
Fraction with No Rec'n							0.06	0.06	0.04	0.05	0.06	0.06
Fraction Weak R&R							0.10	0.10	0.12	0.11	0.10	0.10
Fraction R&R							0.10	0.10	0.12	0.10	0.09	0.10
Fraction Strong R&R							0.04	0.04	0.04	0.04	0.03	0.04
Fraction Accept							0.04	0.03	0.03	0.03	0.02	0.03
Referee Publications in 35 high-impact journals												
Share of refs w/ 3+ publications							0.46	0.40	0.47	0.45	0.45	0.45
Referee genders (share per paper)												
Male							0.85	0.75	0.79	0.81	0.84	0.83
Female							0.14	0.24	0.20	0.18	0.14	0.15
Ambiguous							0.01	0.01	0.01	0.01	0.02	0.01
Number of Observations	19,814	2,273	921	4,723	2,159	29,890	10,199	1,097	585	2,612	654	15,147

Notes: Table presents information on mean characteristics of all submitted papers (Columns 1-6), and for non-desk-rejected papers (Columns 7-12). The latter sample also excludes papers with only 1 referee assigned. Author publications are based on publications in 35 high-impact journals (Online Appendix Table 1) in the 5 years prior to submission. In the case of multiple authors, the measure is the maximum over 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.

Table 2. Survey of Economists about Role of Author and Referee Gender

Surveyed Group:	All	Editors	Female Asst. Pr.	Female EconLit	Male Asst. Pr.	Male EconLit
	(1)	(2)	(3)	(4)	(5)	(6)
Sample Size						
Number surveyed	328	30	20	101	75	102
Number responded	141	14	9	51	26	41
Response Rate	0.43	0.47	0.45	0.50	0.35	0.40
Referee Assignment						
For two papers in the same field, are female-authored papers more likely to be assigned to female referees?						
More likely	0.19	0.36	0.11	0.25	0.17	0.08
Equally likely	0.77	0.64	0.89	0.67	0.83	0.90
Less Likely	0.04	0.00	0.00	0.08	0.00	0.03
Referee Assessment						
Consider the referee rec. for a female-authored paper with at least one male and at least one female referee.						
What percent of female referees are positive?	22.6	20.7	25.0	22.0	24.4	22.4
What percent of male referees are positive?	19.0	17.5	21.6	18.5	17.9	20.4
Consider the referee rec. for a male-authored paper with at least one male and at least one female referee.						
What percent of female referees are positive?	21.3	19.3	25.6	21.7	21.6	20.6
What percent of male referees are positive?	20.1	19.7	22.8	20.2	19.6	20.0
Editor Assessment						
Holding constant the prior publication record of the author(s), the field of the paper, and also the referee recs., do you think a female-authored paper has a higher, lower, or the same probability of receiving a R&R?						
More likely	0.11	0.14	0.11	0.02	0.15	0.20
About the same	0.67	0.57	0.67	0.63	0.69	0.73
Less Likely	0.22	0.29	0.22	0.35	0.15	0.07
Citation Discounting						
Conditional on field and quality, how large is the diff. in citations that a female-authored paper will receive?						
Mean citation gap in log points	-6.5	-3.8	-11.1	-10.2	-4.7	-3.5
Median citation gap in log points	0	0	-10	-10	0	0
Referee Informativeness						
For a given paper, is a positive recommendation from a female referee more informative about future citations, equally informative, or less informative than a positive recommendation from a male referee?						
More informative	0.08	0.00	0.11	0.08	0.12	0.07
About the same	0.86	0.93	0.56	0.82	0.88	0.93
Less informative	0.06	0.07	0.33	0.10	0.00	0.00
For a given paper, do you think that, on average, an editor is more, equally, or less likely to follow the recommendation of a female (relative to a male) referee in the R&R decision?						
More likely	0.03	0.00	0.00	0.04	0.04	0.02
About the same	0.75	0.93	0.78	0.67	0.81	0.76
Less likely	0.22	0.07	0.22	0.29	0.15	0.22
Mixed Gender Papers						
Consider an author team with both males and females, and the author with the most prior publications is female. Would you say that the patterns, in terms of the previous questions, would be more similar to:						
All-female	0.24	0.14	0.22	0.29	0.23	0.23
All-male	0.11	0.07	0.00	0.12	0.12	0.15
Halfway	0.41	0.50	0.33	0.33	0.46	0.47
It depends	0.23	0.29	0.44	0.25	0.19	0.15
If the author with the most prior publication is male, would you say that the patterns would be more similar to:						
All-female	0.01	0.00	0.00	0.00	0.04	0.00
All-male	0.56	0.64	0.89	0.65	0.50	0.40
Halfway	0.31	0.36	0.00	0.22	0.35	0.47
It depends	0.11	0.00	0.11	0.14	0.12	0.13

Notes: For legibility, questions are shortened from the original. Editor surveys were sent to the co-editors of the 4 journals; the number of editors surveyed set at 30 is an estimate. We count as completed surveys with at least 50% of the questions answered.

Table 3. Referee Assignment, Impact of Author Team Gender

Dependent Variable:	**Linear Probability Models**		
	Indicator for Female Referee		**Referee with 3+ Pub.**
	(1)	(2)	(3)
Authors' Genders (Omitted: All Male Authors)			
All Female Authors	0.111	0.074	-0.025
	(0.009)	(0.009)	(0.010)
Mixed-Gender Author Team	0.064	0.049	-0.006
senior author female	(0.011)	(0.011)	(0.013)
Mixed-Gender Author Team	0.043	0.026	-0.010
other	(0.006)	(0.006)	(0.007)
Undetermined Gender Team	0.014	0.004	-0.001
	(0.009)	(0.009)	(0.014)
Gender-field controls			
Share female in sub-fields		0.297	-0.036
		(0.043)	(0.049)
Fraction of gender-topic sub-fields		0.198	-0.034
		(0.019)	(0.021)
Mean of the Dependent Variable:	0.157	0.157	0.461
Controls for Author Publications	No	Yes	Yes
Controls for Referee Publications	No	Yes	No
Controls for No. of Authors	No	Yes	Yes
Controls for Field	No	Yes	Yes
Indicators for Journal-Year	Yes	Yes	Yes
R-squared	0.015	0.048	0.023
N	38,438	38,438	38,438

Notes: The sample is paper-referee observations for 15,147 papers with at least two referees assigned, excluding unknown gendered referees. The dependent variable in Columns 1-2 is an indicator for the referee being female, while the dependent variable in Column 3 is an indicator for the referee having at least 3 publications in the 35 publications in the previous 5 years. Standard errors clustered by paper in parentheses.

Table 4. Referee Recommendations, Impact of Author Team Gender

Specification:	OLS Models for Index of Referee Recommendations				Linear Probability Models for Receiving an R&R Recommendation			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Authors' Genders (Omitted: All Male Authors)								
All Female Authors	-0.043	0.018	0.019		-0.029	0.001	0.001	
	(0.014)	(0.014)	(0.027)		(0.008)	(0.008)	(0.017)	
Mixed-Gender Author Team senior author female	0.045	0.009	0.006		0.014	-0.002	-0.008	
	(0.020)	(0.020)	(0.037)		(0.012)	(0.012)	(0.025)	
Mixed-Gender Author Team other	-0.015	-0.030	-0.024		-0.011	-0.018	-0.019	
	(0.010)	(0.011)	(0.021)		(0.006)	(0.006)	(0.013)	
Undetermined Gender Team	-0.121	-0.073	-0.092		-0.049	-0.026	-0.036	
	(0.019)	(0.019)	(0.039)		(0.010)	(0.010)	(0.023)	
Referee Gender (Omitted: Male Referee)								
Female Referee			-0.060	-0.012			-0.026	-0.006
			(0.014)	(0.014)			(0.009)	(0.009)
Gender Interactions								
All Female Auth. X Female Ref.			0.020	0.011			0.002	-0.000
			(0.037)	(0.037)			(0.024)	(0.023)
Mixed Auth. (F-senior) X Female Ref.			-0.013	-0.029			-0.016	-0.010
			(0.047)	(0.049)			(0.031)	(0.032)
Mixed Auth. (other) X Female Ref.			-0.012	-0.015			-0.011	-0.006
			(0.028)	(0.028)			(0.017)	(0.018)
Undetermined Auth. X Female Ref.			0.067	0.047			-0.004	0.001
			(0.054)	(0.055)			(0.032)	(0.033)
Papers w/ both male & female refs	No	No	Yes	No	No	No	Yes	No
Paper Fixed Effects	No	No	No	Yes	No	No	No	Yes
Controls for Author Pub., No. of Authors, Field, Gender Comp., and Referee Pub.	No	Yes	Yes	-	No	Yes	Yes	-
Indicators for Journal-Year	Yes	Yes	Yes	-	Yes	Yes	Yes	-
R-squared	0.016	0.044	0.049	0.000	0.012	0.030	0.036	0.000
N	38,840	38,840	12,825	38,840	38,840	38,840	12,825	38,840

Notes: The index of referee recommendations is constructed using the coefficients in the cites model in Card and DellaVigna (2017). From Definitely Reject to Accept, the values are 0, 0.67, 1.01, 1.47, 1.92, 2.27, 2.33. Columns 3-4 and 7-8 also include a control for unknown-gender referee (coefficient not shown).

Table 5. Citations and Editor Decision, Impact of Author Team Gender

Specification:	OLS Models for Asinh of Google Scholar Citations				Probit Models for Receiving Revise-and-Resubmit		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Authors' Genders							
All Female	-0.07	0.24	0.22	0.22	-0.08	0.03	0.01
	(0.05)	(0.06)	(0.05)	(0.05)	(0.06)	(0.05)	(0.06)
Mixed-Gender Author Team senior author female	0.37	0.05	0.06	0.06	0.13	0.05	0.10
	(0.08)	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)
Mixed, other	0.26	-0.02	0.01	0.01	0.03	-0.08	-0.02
	(0.04)	(0.04)	(0.04)	(0.04)	(0.05)	(0.04)	(0.05)
Undetermined	-0.36	-0.39	-0.31	-0.31	-0.09	-0.10	-0.06
	(0.05)	(0.06)	(0.05)	(0.05)	(0.08)	(0.06)	(0.08)
Fractions of Referee Recommendations							
Reject	0.80		0.64	0.64	0.87		0.86
	(0.06)		(0.06)	(0.06)	(0.15)		(0.16)
No Recommendation	1.24		1.00	0.98	2.78		2.73
	(0.12)		(0.10)	(0.10)	(0.17)		(0.18)
Weak R&R	1.73		1.47	1.45	3.16		3.16
	(0.10)		(0.09)	(0.10)	(0.17)		(0.18)
R&R	2.33		1.93	1.89	4.63		4.61
	(0.10)		(0.09)	(0.13)	(0.20)		(0.21)
Strong R&R	2.72		2.32	2.26	5.57		5.56
	(0.15)		(0.13)	(0.22)	(0.21)		(0.21)
Accept	2.74		2.36	2.30	5.39		5.37
	(0.13)		(0.12)	(0.19)	(0.21)		(0.22)
Author Publications in 35 High-Impact Journals							
1 Publication		0.41	0.29	0.29		0.24	0.04
		(0.04)	(0.04)	(0.04)		(0.04)	(0.05)
2 Publications		0.65	0.49	0.49		0.37	0.19
		(0.04)	(0.04)	(0.04)		(0.05)	(0.07)
3 Publications		0.79	0.58	0.58		0.47	0.17
		(0.04)	(0.04)	(0.04)		(0.05)	(0.07)
4-5 Publications		1.06	0.81	0.80		0.64	0.33
		(0.06)	(0.06)	(0.06)		(0.05)	(0.06)
6+ Publications		1.31	0.99	0.99		0.82	0.41
		(0.05)	(0.05)	(0.05)		(0.06)	(0.08)
Gender-field controls							
Share female in sub-fields		-0.03	-0.02	-0.01		-0.48	-0.35
		(0.28)	(0.26)	(0.25)		(0.28)	(0.32)
Fraction of gender-topic sub-fields		-0.01	0.02	0.02		-0.09	-0.14
		(0.09)	(0.10)	(0.10)		(0.10)	(0.14)
R&R Indicator				0.06			
(Mechanical Publ. Effect)				(0.14)			
Control Function for Selection				0.32			
(Value Added of the Editor)				(0.08)			
Editor Leave-out-Mean R&R							3.43
Rate							(0.73)
Controls for No. of Authors & Field	No	Yes	Yes	Yes	No	Yes	Yes
Indicators for Journal-Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	15,147	15,147	15,147	15,147	15,147	15,147	15,147
R^2 / pseudo R^2	0.20	0.20	0.26	0.27	0.48	0.07	0.49

Notes: The sample is non-desk-rejected papers with at least two referees assigned. The control function for selection in Column 4 is calculated using predicted probabilities based on Column 7. Standard errors clustered by editor in parentheses.

Table 6. Citations and Editor Decision, Robustness

	Coefficients in Citation Model			Coefficients in Probit of R&R Decision			R&R Rate for All-Female Papers	
<u>Robustness Result:</u>	All-Female Auhors	Mixed-Gender, Senior Female	Mixed-Gender, Other	All-Female Authors	Mixed-Gender, Senior Female	Mixed-Gender, Other	Data	Cite-Max
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Benchmark	0.22	0.06	0.01	0.01	0.10	-0.02	0.122	0.191
	(0.05)	(0.07)	(0.04)	(0.06)	(0.07)	(0.05)	(0.01)	(0.02)
Controls								
Extra Controls for Author Prominence	0.22	0.08	0.01	0.01	0.14	-0.01	0.122	0.193
	(0.05)	(0.07)	(0.04)	(0.07)	(0.07)	(0.05)		
Extra Controls for Author and Institutional Prominence	0.17	0.04	0.00	-0.01	0.13	-0.01	0.122	0.183
	(0.05)	(0.07)	(0.04)	(0.07)	(0.07)	(0.05)		
Alternative Citation Measures								
Probit Model for Top-Cited Paper	0.13	0.10	-0.00	0.01	0.1	-0.02	0.122	0.190
	(0.05)	(0.06)	(0.03)	(0.06)	(0.07)	(0.05)		
OLS Regression of Citation Percentile	3.60	0.51	0.07	0.01	0.10	-0.02	0.122	0.195
	(0.85)	(1.10)	(0.64)	(0.06)	(0.07)	(0.05)		
Tobit Reg. of asinh SSCI Cites (2006-2008) (N=4507)	0.32	-0.04	-0.02	-0.01	0.41	0.10	0.124	0.215
	(0.15)	(0.27)	(0.11)	(0.11)	(0.11)	(0.06)		
By Year of Submissions								
Papers in Years 2003-2009	0.31	0.10	0.02	-0.00	0.36	-0.01	0.147	0.258
(N=7207)	(0.08)	(0.13)	(0.06)	(0.08)	(0.14)	(0.07)		
Papers in Years 2010-2013	0.13	0.02	-0.00	0.05	-0.13	-0.02	0.097	0.129
(N=7940)	(0.07)	(0.08)	(0.06)	(0.09)	(0.09)	(0.06)		
Papers in Years 2014-2017	-	-	-	0.09	0.06	-0.05	-	-
(N=6544)				(0.17)	(0.17)	(0.10)		
By Number of Authors								
Papers with 1 Author	0.17	-	-	0.11	-	-	0.122	0.150
(N=4639)	(0.07)			(0.08)				
Papers with 2 Authors	0.34	-0.08	0.01	-0.23	0.12	-0.13	0.128	0.308
(N=6406)	(0.08)	(0.10)	(0.06)	(0.12)	(0.11)	(0.08)		
Papers with 3 Authors	0.06	0.23	0.03	-0.15	0.10	0.05	0.079	0.115
(N=4102)	(0.23)	(0.09)	(0.05)	(0.49)	(0.10)	(0.07)		
By Author Publications								
Papers by Authors with	0.18	0.06	0.06	0.04	0.08	-0.03	0.113	0.167
0-3 Previous Pubs. (N=10771	(0.05)	(0.09)	(0.04)	(0.07)	(0.10)	(0.07)		
Papers by Authors with	0.49	0.05	-0.05	-0.14	0.15	-0.01	0.219	0.533
4+ Previous Pubs. (N=4376)	(0.10)	(0.10)	(0.07)	(0.22)	(0.13)	(0.07)		
By Share of Women in the Field								
Fields with Lower Share	0.20	0.03	-0.00	0.03	0.16	-0.06	0.142	0.215
of Women (N=7573)	(0.08)	(0.10)	(0.06)	(0.11)	(0.10)	(0.07)		
Field with Higher Share	0.21	0.06	0.00	-0.03	0.08	0.04	0.104	0.168
of Women (N=7529)	(0.08)	(0.08)	(0.06)	(0.10)	(0.10)	(0.06)		
By Number of Referees								
Papers with 1-2 Referees	0.29	0.07	0.03	0.07	0.14	0.03	0.105	0.203
(N=7940)	(0.07)	(0.12)	(0.05)	(0.08)	(0.15)	(0.08)		
Papers with 3+ Referees	0.12	0.04	-0.02	-0.04	0.08	-0.05	0.142	0.198
(N=7207)	(0.07)	(0.08)	(0.06)	(0.09)	(0.08)	(0.08)		
By Share of Women Referees								
Papers with All-Male	0.17	-0.05	-0.03	0.02	0.03	-0.08	0.133	0.183
Referees (N=10195)	(0.06)	(0.11)	(0.06)	(0.09)	(0.08)	(0.07)		
Papers with some Female	0.29	0.19	0.08	0.09	0.22	0.08	0.109	0.200
Referees (N=4952)	(0.07)	(0.09)	(0.05)	(0.11)	(0.11)	(0.11)		

Notes: The table reports the result of multiple robustness checks and sample splits. The coefficients in Columns 1-3 come from regressions with the same controls as in Table 5, Column 4. The coefficients in Columns 4-6 come from regressions with the same controls as in Table 5, Column 7. Column 7 reports the observed R&R rate (in the sample of non-desk-rejected papers) for all-female-authored papers, while Column 8 reports the counterfactual R&R rate for all-female papers if the R&R decision had the weights that maximize citation with respect to the author-gender variables (see Figure 7 for further detail).

Table 7. Citations, Results for R&Rs, Accepted Papers, and Published Papers

	OLS Models for Asinh of Google Scholar Citations				
Data Set:	Editorial Express Submissions, GS Cites in 2015			Published Papers in Econlit, GS Cites in 2018	
Sample of Papers:	Non-Desk-Rejected Papers	R&R Papers	Accepted Papers	Publications in our 4 Journals, 2008-2015	
	(1)	(2)	(3)	(4)	(5)
Authors' Genders					
All Female	0.22	0.26	0.24	0.39	0.30
	(0.05)	(0.13)	(0.13)	(0.15)	(0.15)
Mixed-Gender Author Team senior author female	0.06	-0.19	-0.11	0.10	0.16
	(0.07)	(0.17)	(0.16)	(0.20)	(0.20)
Mixed, other	0.01	0.15	0.09	0.09	0.08
	(0.04)	(0.11)	(0.11)	(0.09)	(0.09)
Undetermined	-0.31	-0.41	-0.40	-0.22	-0.21
	(0.05)	(0.16)	(0.19)	(0.27)	(0.28)
Author Publications in 35 High-Impact Journals (Max across Authors)					
1 Publication	0.29	0.22	0.28	0.05	
	(0.04)	(0.11)	(0.12)	(0.15)	
2 Publications	0.49	0.21	0.29	0.15	
	(0.04)	(0.12)	(0.14)	(0.16)	
3 Publications	0.58	0.52	0.55	0.58	
	(0.04)	(0.13)	(0.17)	(0.14)	
4-5 Publications	0.80	0.55	0.52	0.45	
	(0.06)	(0.13)	(0.15)	(0.14)	
6+ Publications	0.99	0.67	0.66	0.60	
	(0.05)	(0.14)	(0.16)	(0.15)	
R&R Indicator (Mechanical Publ. Effect)	0.06				
	(0.14)				
Control Function for Selection (Value Added of the Editor)	0.32	-0.14	-0.10		
	(0.08)	(0.18)	(0.27)		
Controls for Fraction Referee Rec:	Yes	Yes	Yes	No	No
Controls for No. of Authors	Yes	Yes	Yes	Yes	Yes
Controls for Field & Gender-Field	Yes	Yes	Yes	Yes	Yes
Indicators for Journal-Year	Yes	Yes	Yes	Yes	Yes
N	15,147	2,209	1,713	1,530	1,530
R^2 / pseudo R^2	0.27	0.26	0.30	0.24	0.22

Notes: The sample in Column 1 is 15,147 non-desk-rejected papers with at least two referees assigned. The sample in Column 2 is all papers with an R&R invitation, and the sample in Column 3 is all papers which are ultimately accepted within the time frame of the data. The sample in Columns 4-5 is instead from Econlit, tracking all papers published in the 4 journals in our sample in 2008-2015, which corresponds approximately to the submissions in our sample, assuming a 2-year delay in publication. For this data set, we measure Google Scholar citations in September 2018 (as opposed to mid 2015 for Columns 1-3).

Table 8. Desk Rejection, Impact of Author Team Gender

Specification:	OLS Reg.	Probit	OLS Reg.	
Dependent Variable:	Asinh of Citations	Indicator for Paper Not Desk Rejected	Asinh of Citations	
	(1)	(2)	(3)	(4)
Authors' Genders				
All Female	0.24	0.13	0.17	0.15
	(0.04)	(0.04)	(0.05)	(0.04)
Mixed-Gender Author Team	0.02	-0.03	0.10	0.15
senior author female	(0.05)	(0.09)	(0.05)	(0.05)
Mixed, other	-0.02	-0.03	0.12	0.15
	(0.03)	(0.04)	(0.04)	(0.03)
Undetermined	-0.50	-0.39	-0.41	-0.34
	(0.04)	(0.05)	(0.06)	(0.05)
Author Publications in 35 high-impact journals				
Publications: 1	0.53	0.40	0.56	0.55
	(0.05)	(0.04)	(0.06)	(0.05)
Publications: 2	0.83	0.62	0.90	0.87
	(0.07)	(0.04)	(0.09)	(0.07)
Publications: 3	0.98	0.81	1.06	1.01
	(0.08)	(0.06)	(0.10)	(0.09)
Publications: 4-5	1.28	1.05	1.41	1.32
	(0.10)	(0.08)	(0.14)	(0.11)
Publications: 6+	1.58	1.34	1.67	1.62
	(0.12)	(0.10)	(0.16)	(0.13)
Gender-field controls				
Share female in sub-fields	0.25	0.21		0.93
	(0.24)	(0.17)		(0.21)
Fraction of gender-topic sub-fielc	0.01	0.01		-0.06
	(0.06)	(0.06)		(0.07)
NDR Indicator	0.41		0.86	0.86
	(0.29)		(0.07)	(0.06)
Control Function for Selection into NE	0.27			
(Value Added of the Editor)	(0.15)			
Editor Leave-out-Mean NDR		2.78		
Rate		(0.36)		
Controls for No. of Authors and Field	Yes	Yes	No	No
Indicators for Journal-Year	Yes	Yes	No	Yes
Control for Cubic in P(NDR)	No	No	Yes	Yes
Number of Observations	29,890	29,890	29,890	29,890
R^2 / pseudo R^2	0.28	0.24	0.20	0.27

Notes: Dependent variable for OLS model in Columns 1 and 3-4 is asinh of Google Scholar citations. Dependent variable in probit model in Column 2 is indicator for avoiding desk rejection. The control function for selection in Column 1 is calculated using predicted probabilities based on Column 2. In Columns 3 and 4 we control for a cubic polynomial in the probability of non-desk-rejection, built using the specification in Column 2. Standard errors clustered by editor in parentheses.

Table 9. Effect of Referee Gender on Referee Informativeness and Weight

	NLS Models for Asinh of Google Scholar Citations			ML Probit Models for Receiving Revise-and-Resubmit Decision		
	(1)	(2)	(3)	(4)	(5)	(6)
Gender Slope Variables						
Female Referee		0.060	0.057		-0.028	-0.018
		(0.095)	(0.096)		(0.049)	(0.050)
Gender Level Controls						
Female Referee		-0.019	-0.016		-0.111	-0.143
		(0.110)	(0.111)		(0.146)	(0.148)
All Female Authors	0.221	0.219	0.218	0.011	0.039	0.042
	(0.053)	(0.055)	(0.055)	(0.064)	(0.067)	(0.068)
Mixed-Gender Author Team	0.058	0.054	0.051	0.099	0.106	0.102
senior author female	(0.068)	(0.068)	(0.068)	(0.069)	(0.067)	(0.066)
Mixed-Gender Author Team	0.013	0.009	0.009	-0.021	-0.020	-0.021
other	(0.040)	(0.039)	(0.039)	(0.049)	(0.049)	(0.050)
Other Slope Variables						
Referee Publications 3+		0.001	-0.010		0.187	0.163
		(0.059)	(0.056)		(0.032)	(0.032)
Asinh (No. Reports for Editor)			0.048			0.076
			(0.028)			(0.021)
Journal Fixed Effect	No	Yes	Yes	No	Yes	Yes
Field Fixed Effect	No	Yes	Yes	No	Yes	Yes
Level Additional Controls						
Share Referees with 3+ Pubs.		0.285	0.290		-0.299	-0.248
		(0.061)	(0.061)		(0.145)	(0.147)
Mean Asinh (No. Reports for Editor)			-0.029			-0.131
			(0.035)			(0.074)
Fractions of Referee Recommendations (Other Fractions Included, not Reported)						
R&R	1.886	1.821	1.790	4.593	4.155	4.024
	(0.126)	(0.235)	(0.241)	(0.214)	(0.433)	(0.421)
Author Publications (Other Indicators Included, not Reported)						
6+ Publications	0.996	0.953	0.952	0.415	0.394	0.399
	(0.049)	(0.049)	(0.048)	(0.079)	(0.078)	(0.078)
R&R Indicator	0.060	0.212	0.242			
(Mechanical Publ. Effect)	(0.142)	(0.131)	(0.132)			
Control Function for Selection	0.324	0.233	0.214			
(Value Added of the Editor)	(0.085)	(0.075)	(0.074)			
Editor Leave-out-Mean R&R				2.749	3.097	3.014
Rate				(0.721)	(0.762)	(0.766)

Notes: Standard errors clustered by editor in parentheses. For papers with more than 5 referees, referees after the fifth are randomly dropped.

Table 10. Referee Acceptance and Referee Delays, by Author and Referee Gender

Specification:	Linear Probability Model for Referee Accepting a Report Request			OLS Regression of Number of Days from Submission to Referee Report		
	(1)	(2)	(3)	(4)	(5)	(6)
Authors' Genders (Omitted: All Male Authors)						
All Female Authors	0.007	0.009		-4.20	-1.26	
	(0.008)	(0.008)		(0.95)	(0.97)	
Mixed-Gender Author Team	0.032	0.029		0.15	-0.24	
senior author female	(0.011)	(0.011)		(1.97)	(1.93)	
Mixed-Gender Author Team	-0.001	-0.000		-0.77	0.20	
other	(0.006)	(0.006)		(0.68)	(0.72)	
Undetermined Gender Team	-0.015	-0.007		-3.34	-1.80	
	(0.010)	(0.010)		(1.48)	(1.45)	
Referee Gender (Omitted: Male Referee)						
Female Referee	0.012	0.004	-0.000	-2.28	-0.53	0.35
	(0.005)	(0.006)	(0.008)	(0.55)	(0.55)	(0.76)
Gender Interactions						
All Female Auth. X Female Ref.			0.010			1.32
			(0.023)			(1.96)
Mixed Auth. (senior-F) X Female Ref.			-0.008			-2.85
			(0.029)			(2.66)
Mixed Auth. (other) X Female Ref.			-0.025			1.08
			(0.017)			(1.50)
Undetermined Auth. X Female Ref.			0.011			0.93
			(0.033)			(3.67)
Paper Fixed Effects	No	No	Yes	No	No	Yes
Controls for Referee Recommendation	No	No	No	No	Yes	Yes
Controls for Referee Publications	No	Yes	Yes	No	Yes	Yes
Controls for Author Pub & No. of Authors	No	Yes	-	No	Yes	-
Controls for Field & Gender-Field Ctrls	No	Yes	-	No	Yes	-
Indicators for Journal-Year	Yes	Yes	Yes	Yes	Yes	Yes
R-squared	0.013	0.018	0.005	0.29	0.31	0.01
N	60,445	60,445	60,445	38,825	38,825	38,825

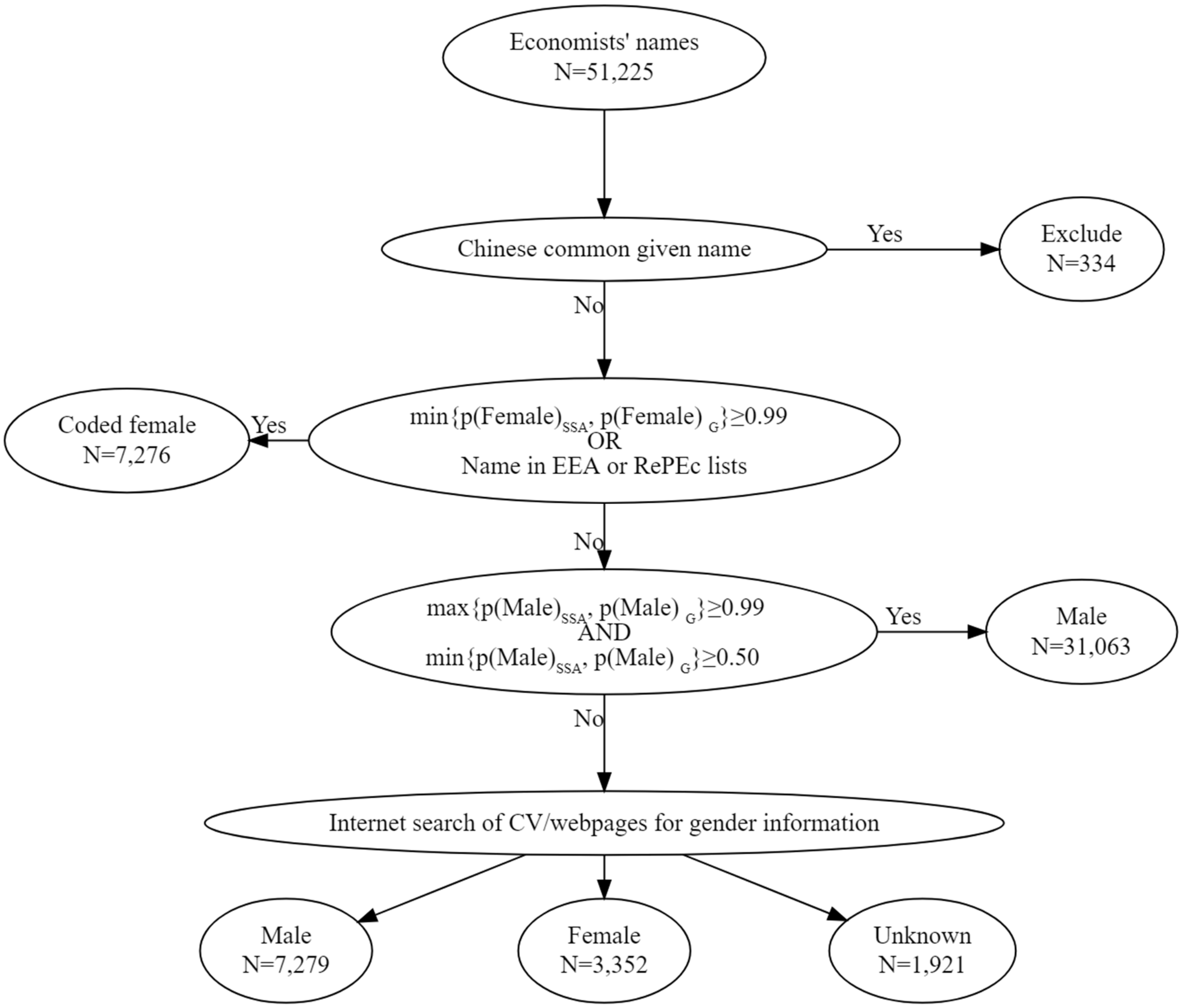
Notes: Standard errors clustered by paper in parentheses. The sample in Columns 1-3 is a referee-paper observations, including any referee invited to review a paper. The sample in Columns 1-6 is a referee-paper observations, including any referee who returned a review for that paper. Report time is calculated as the number of days from paper submission to referee report submission, rounded to the nearest 10.

Table 11. Decision Time and Duration of Revisions, by Author Team Gender

	Number of Days			No. of Rounds (for R&Rs)	Days Before Resub. (R&Rs)	Days from Resub. to Accept
	Sub. To Last Report Received	Reports Received to Editor Dec.	Sub. To Editor Dec.			
	(1)	(2)	(3)	(4)	(5)	(6)
Authors' Genders						
All Female	-1.10	-0.68	-1.85	0.05	-13.24	14.64
	(1.40)	(0.95)	(1.49)	(0.08)	(17.97)	(22.15)
Mixed-Gender Author Tear	0.60	-0.54	0.71	-0.05	-9.05	-10.37
senior author female	(2.03)	(2.29)	(3.18)	(0.08)	(16.11)	(20.37)
Mixed, other	1.18	0.19	1.82	0.01	-0.88	11.74
	(0.83)	(1.45)	(1.60)	(0.04)	(12.49)	(10.78)
Undetermined	-0.32	-0.33	-0.71	0.01	-38.74	28.11
	(1.97)	(2.23)	(2.50)	(0.11)	(22.76)	(34.87)
Fractions of Referee Recommendations						
Reject	12.90	1.57	14.80	0.24	12.87	10.38
	(2.21)	(1.68)	(2.82)	(0.32)	(46.98)	(82.21)
No Recommendation	35.11	9.67	45.23	0.22	-61.70	-22.07
	(5.21)	(4.95)	(7.45)	(0.29)	(53.74)	(72.79)
Weak R&R	26.18	22.81	49.34	0.35	-5.24	34.19
	(3.53)	(4.90)	(6.43)	(0.33)	(43.50)	(82.21)
R&R	43.77	20.03	64.10	0.38	-29.10	-5.04
	(4.87)	(9.57)	(12.72)	(0.33)	(52.52)	(73.08)
Strong R&R	37.49	15.29	53.41	0.15	-67.39	-65.84
	(5.00)	(11.09)	(13.55)	(0.34)	(56.76)	(81.54)
Accept	43.70	12.47	56.04	-0.11	-107.61	-120.97
	(6.62)	(11.66)	(15.15)	(0.32)	(57.32)	(73.06)
R&R Indicator	1.71	30.47	31.67			
(Mechanical Publ. Effect)	(1.62)	(8.91)	(9.25)			
Sample	Non Desk Rejected Papers			R&R Papers Only		
Controls for Author Pub., No. of Authors, Field & Gender-	Yes	Yes	Yes	Yes	Yes	Yes
Indicators for Journal-Year	Yes	Yes	Yes	Yes	Yes	Yes
Editor Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Mean of Dependent Variable:	94.9	23.1	116.9	2.7	253.5	218.3
N	15,147	14,859	15,147	2,046	1,996	2,003
R-squared	0.08	0.09	0.16	0.07	0.09	0.07

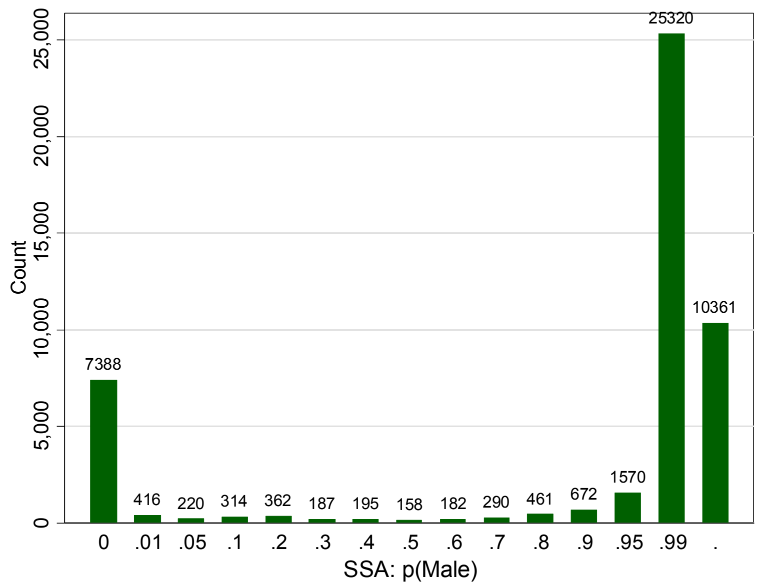
Notes: Decision time is calculated as the number of days from paper submission to referee report submission. This is rounded to the nearest 10. Editor fixed effects and clustered standard errors in parentheses. Column 2 excludes papers whose last reports arrive after the editor's decision.

Online Appendix Figure 1a. Coding Gender for Names



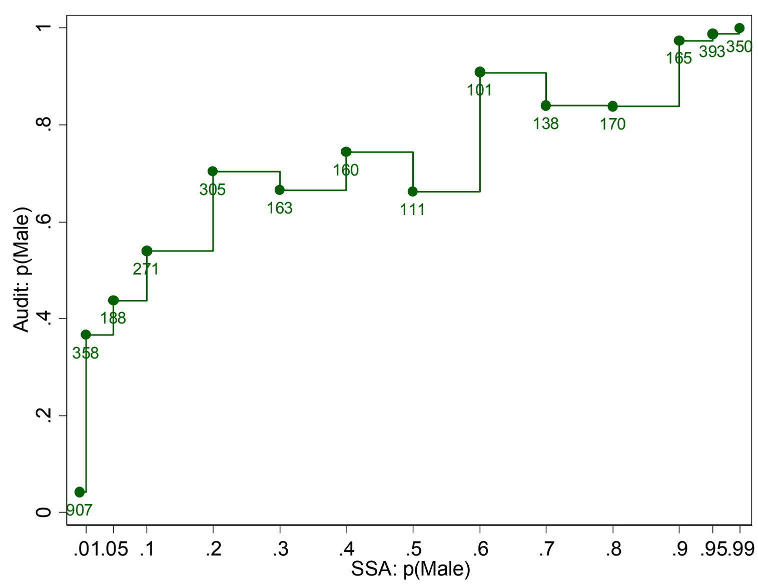
Note: Graph shows the process by which gender is assigned to names.

Online Appendix Figure 1b. Distribution of P(Male) According to SSA for Econlit Sample



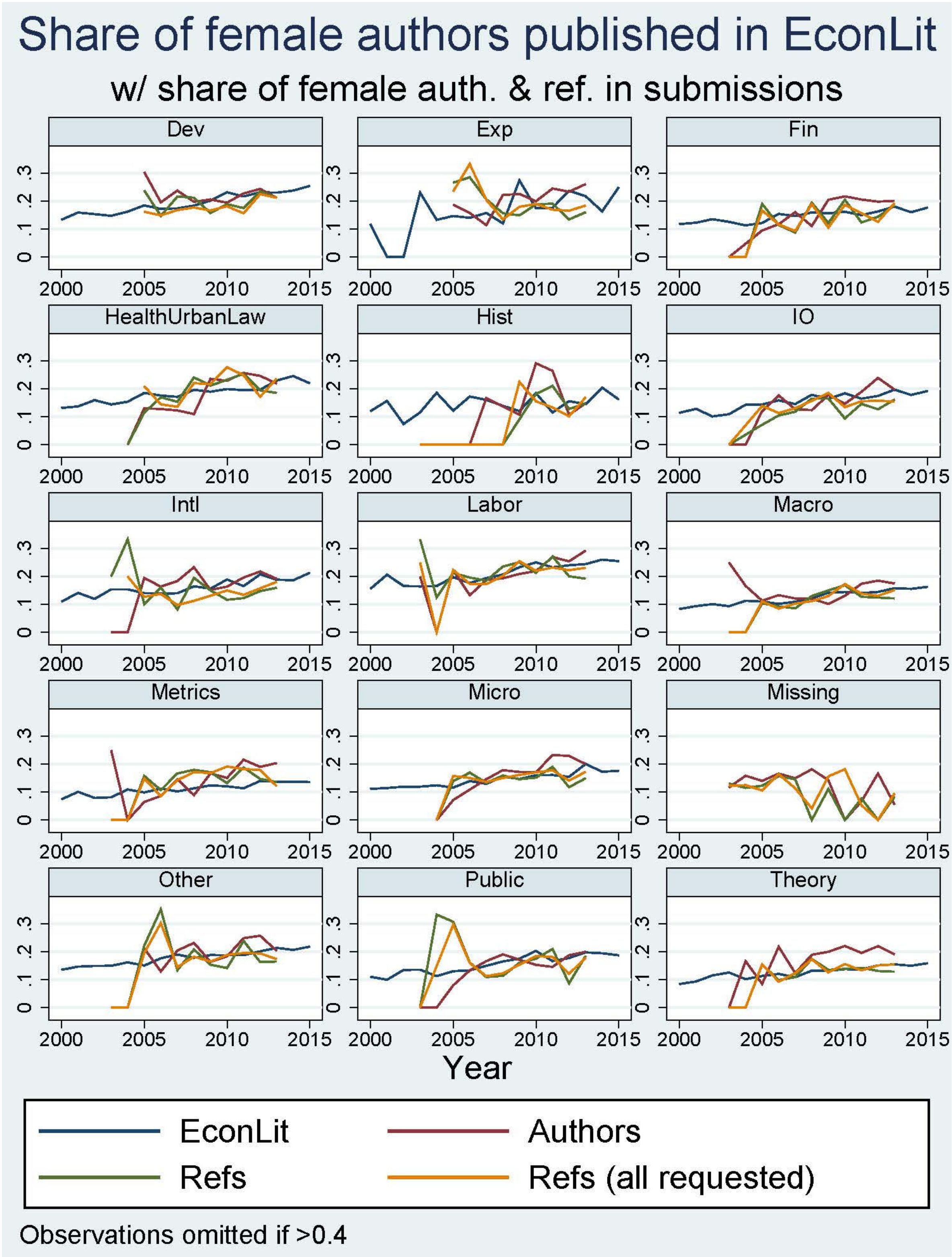
Note: Each observation is an author in a dataset of all papers published in 63 journals from 1991 to 2017 from Econlit. For each author, we code the probability that the author is male based on the first name, using an R routine that is based on the SSA data set of names. The graph indicates the p(male) as well as the number of observations in each bin. The last bin indicates cases in which there is no matching first name in the Census data.

O. A. Figure 1c. Share of Males in Audited Econlit Sample by Assessed P (Male) According to SSA



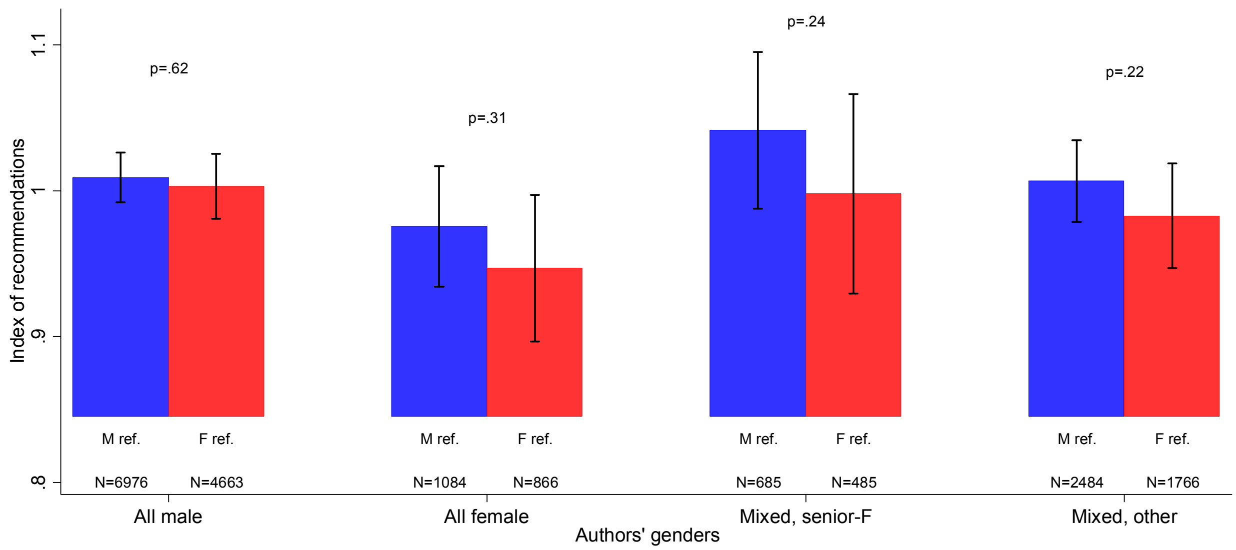
Note: Each observation is an author in a dataset of all papers published in 63 journals from 1991 to 2017 from Econlit. For each author, we code the probability that the author is male based on the first name, using an R routine that is based on the SSA data set of names. The plot then depicts, within each bin of the coded p(male), the share of male economists in the sample of names that the undergraduate students audited. The numbers in the graph report the number of economists in the audit data set. Notice that for economists in the *ConsistentM*, *ConsistentF*, or *SingleM* (see below) we sampled only a small random sample, while we attempted to sample all economists with intermediate probabilities; hence, the discrepancies in the cell numbers compared to Figure 1. The reported p(male) in the audit (the y axis) reweights observations by the sampling probability.

Online Appendix Figure 2. Share of Female Authors and Referees, by Field

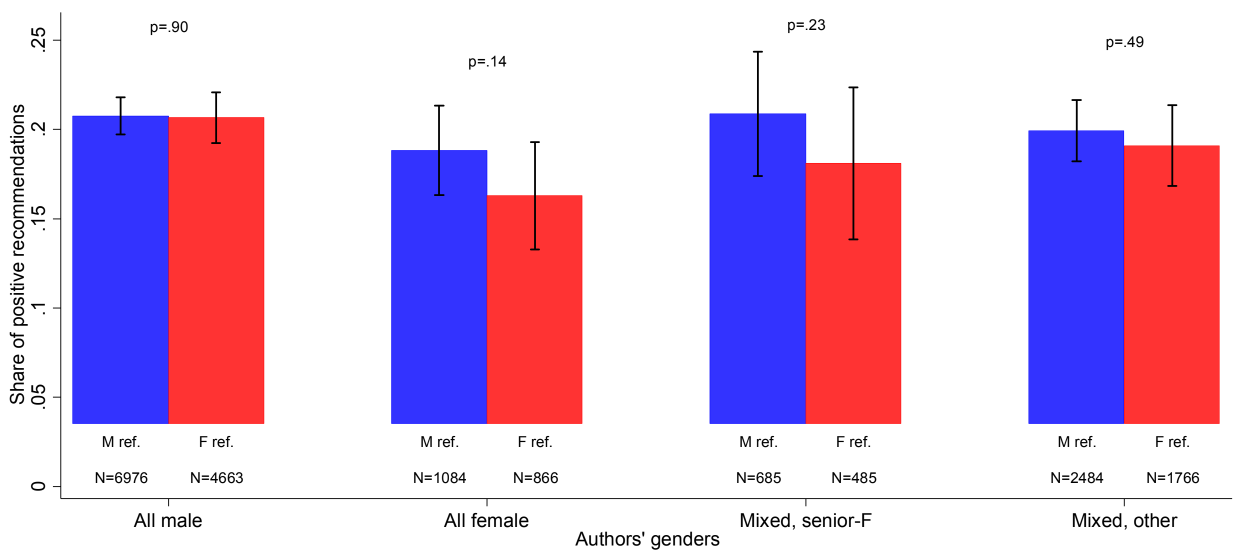


Online Appendix Figure 3. Referee Evaluation by Author Gender and Referee Gender, Extended Sample (Up to 2017)

Online Appendix Figure 3a. Index of Referee Recommendations



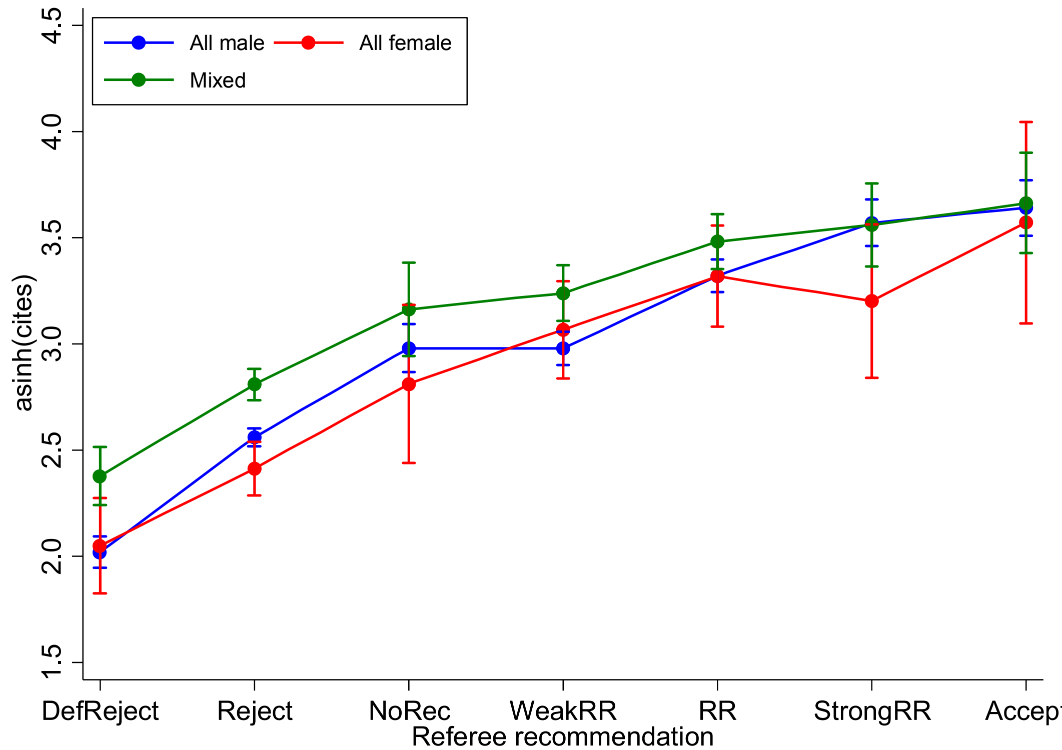
Online Appendix Figure 3b. Share of Positive Referee Recommendations



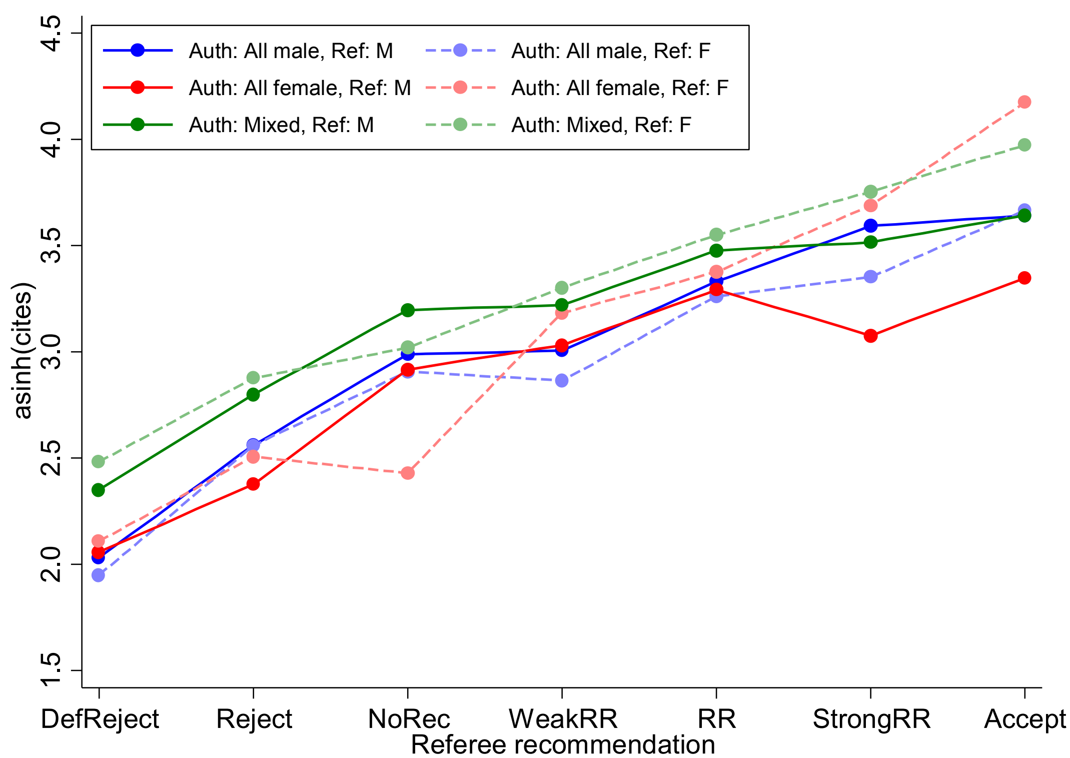
Notes: Online Appendix Figure 3a displays the mean recommendation given by referees based on gender. The index of referee recommendations is constructed using the coefficients in the cites model in Card and DellaVigna (2017). From Definitely Reject to Accept, the values are 0, 0.67, 1.01, 1.47, 1.92, 2.27, 2.33. The bands show 2 standard error intervals, clustered at the paper level. Includes only 6,585 papers with both male and female referees. Figure 3b shows the share of positive recommendations, defined as RR-Accept. In both panels, female referees are weighted at the paper level by N_{male} / N_{female}.

Online Appendix Figure 4. Differences in Citations and R&R Rate, by Author Gender

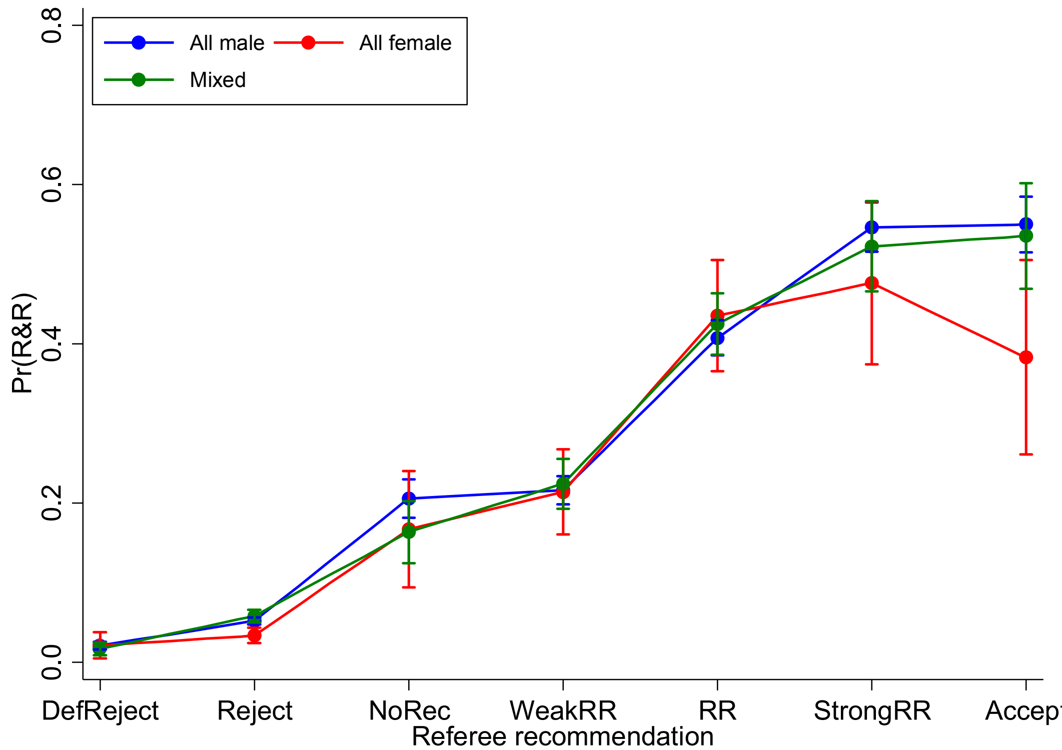
Online Appendix Figure 4a. Referee Recommendations and Citations



Online Appendix Figure 4b. Recommendations and Citations, by Author Gender and Referee Gender

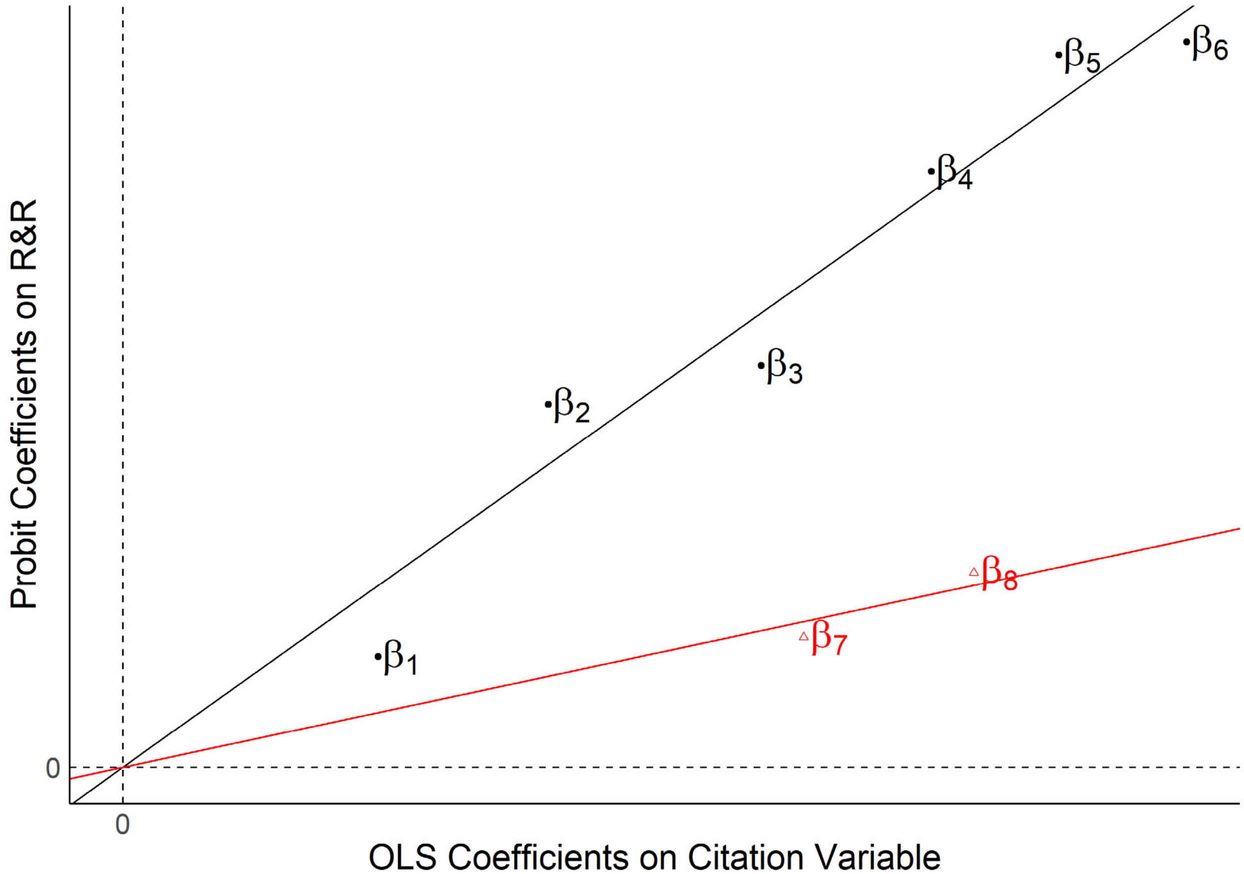


Online Appendix Figure 4c. Referee Recommendations and R&R Rate



Notes: Online Appendix Figures 4a and 4b show the weighted *asinh* (citations) for a paper receiving a given recommendation, while Figure 4c shows the R&R rate for a paper receiving a given recommendation. Figures 5a and 5c show the results separately by author gender. Figure 4b splits these two categories further into referees' gender. 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 4b omits confidence intervals for legibility.

Online Appendix Figure 5. Model Prediction: Predictors of Citation versus Predictors of Editor Decision



Notes: The Figure 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. The coefficient labels and values in the simulations are arbitrary.

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

Panel A. List of Journals Used in Publication Counts

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	

Panel B. List of Additional Journals Used to Generate List of Authors Coded for Gender

Economic Theory	Journal of Economics and Management Strategy
European Economic Review	Labour Economics
Quantitative Economics	Public Choice
Theoretical Economics	European Journal of Political Economy
Review of Economic Dynamics	Scandinavian Journal of Economics
Journal of Applied Econometrics	Regional Science and Urban Economics
Journal of Economic Perspectives	Mathematical Social Sciences
Economic Policy	International Tax and Public Finance
World Bank Economic Review	Environmental and Resource Economics
Journal of Law and Economics	Journal of Development Studies
Journal of Risk and Uncertainty	Energy Economics
Journal of Environmental Economics and Managemen	Journal of International Money and Finance
Journal of Economic Behavior and Organization	Journal of Money, Credit, and Banking
Journal of Theoretical Public Economics	Journal of Public Economic Theory

Notes: The 35 journals in Panel A are used to build measures of author and referee prominence, as the number of articles published in the previous 5 years in one of the journals by an author/referee. The additional journals in Panel B are used to build a database of economists, which we gender code.

Online Appendix Table 2. Citations and Editor Decision, Impact of Controls and Additional Measures of Prominence

	OLS Models for Asinh of Google Sc. Citations					Probit Models for Receiving R&R Dec.				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Authors' Genders										
All Female	-0.11	0.14	0.22	0.22	0.17	-0.08	0.02	0.02	0.01	-0.01
	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.06)	(0.06)	(0.06)	(0.07)	(0.07)
Mixed-Gender Author Team	0.33	0.14	0.07	0.08	0.04	0.15	0.11	0.12	0.14	0.13
senior author female	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)
Mixed, other	0.24	0.13	0.01	0.01	0.00	0.04	-0.00	-0.01	-0.01	-0.01
	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)
Undetermined	-0.37	-0.24	-0.30	-0.30	-0.27	-0.08	-0.04	-0.04	-0.03	-0.03
	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.08)	(0.08)	(0.08)	(0.08)	(0.09)
Author Publications in 35 High-Impact Journals (Max across Authors)										
1 Publication		0.37	0.22	0.22	0.22		0.04	-0.02	-0.06	-0.05
		(0.04)	(0.05)	(0.05)	(0.05)		(0.05)	(0.05)	(0.06)	(0.06)
2 Publications		0.59	0.34	0.33	0.32		0.19	0.08	-0.00	-0.01
		(0.04)	(0.05)	(0.05)	(0.05)		(0.06)	(0.07)	(0.07)	(0.07)
3 Publications		0.70	0.37	0.33	0.33		0.17	0.01	-0.11	-0.10
		(0.03)	(0.06)	(0.06)	(0.06)		(0.06)	(0.09)	(0.09)	(0.09)
4-5 Publications		0.95	0.51	0.43	0.42		0.33	0.10	-0.05	-0.05
		(0.05)	(0.09)	(0.09)	(0.09)		(0.05)	(0.10)	(0.11)	(0.11)
6+ Publications		1.15	0.53	0.36	0.33		0.42	0.07	-0.14	-0.15
		(0.05)	(0.11)	(0.12)	(0.11)		(0.07)	(0.13)	(0.13)	(0.14)
Author Publications in 35 High-Impact Journals, Mean across Authors										
Average Publications Across Coauthors			0.09	0.06	0.06			0.07	0.04	0.04
			(0.02)	(0.02)	(0.02)			(0.02)	(0.03)	(0.03)
Author Publications in Top 5 Journals (Max Across Authors)										
1 Publication				0.29	0.23				0.22	0.19
				(0.04)	(0.04)				(0.05)	(0.05)
2 Publications				0.41	0.30				0.27	0.22
				(0.04)	(0.04)				(0.08)	(0.08)
3+ Publications				0.53	0.34				0.45	0.37
				(0.07)	(0.06)				(0.08)	(0.07)
Author Publications in 35 High-Impact Journals, 6-10 years ago (Max Across Authors)										
1-3 Publications				-0.10	-0.07				0.15	0.17
				(0.03)	(0.03)				(0.05)	(0.05)
4+ Publications				0.03	0.04				0.11	0.12
				(0.05)	(0.04)				(0.06)	(0.06)
Rank of Authors' Institution										
US: 1-10					0.43					0.21
					(0.04)					(0.05)
US: 11-20					0.29					0.18
					(0.05)					(0.05)
Europe: 1-10					0.32					0.10
					(0.04)					(0.06)
Rest of World: 1-5					-0.16					0.10
					(0.09)					(0.09)
R&R Indicator	-0.02	0.05	0.07	0.03	0.05					
(Mechanical Publ. Effect)	(0.14)	(0.14)	(0.14)	(0.13)	(0.14)					
Control Function for Selection	0.42	0.33	0.32	0.33	0.30					
(Value Added of the Editor)	(0.08)	(0.08)	(0.08)	(0.08)	(0.08)					
Editor Leave-out-Mean R&R						3.38	3.41	3.42	3.39	3.42
Rate						(0.71)	(0.73)	(0.73)	(0.73)	(0.72)
Controls for Referee Recommendation	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls for No. of Authors	No	No	Yes	Yes	Yes	No	No	Yes	Yes	Yes
Controls for Field & Gender-Field Ctrls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Indicators for Journal-Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R^2 / pseudo R^2	0.22	0.27	0.27	0.28	0.29	0.48	0.49	0.49	0.50	0.50

Notes: The sample for all models is 15,147 non-desk-rejected papers with at least two referees assigned. Dependent variable for OLS models in Columns 1-5 is asinh of Google Scholar citations. Dependent variable in probit models in Columns 6-10 is indicator for receiving revise and resubmit decision. The control function for selection in Columns 1-5 is calculated using predicted probabilities based on Columns 6-10. Standard errors clustered by editor in parentheses.

Online Appendix Table 3. Citations and Editor Decision, Results Split by Number of Authors

	OLS Models for Asinh of Google Scholar Citations			Probit Models for Receiving Revise-and-Resubmit		
Number of Authors:	**1 Author**	**2 Authors**	**3+ Authors**	**1 Author**	**2 Authors**	**3+ Authors**
	(1)	(2)	(3)	(4)	(5)	(6)
Authors' Genders						
All Female	0.17	0.34	0.05	0.11	-0.23	-0.16
	(0.07)	(0.08)	(0.23)	(0.08)	(0.12)	(0.50)
Mixed-Gender Author Team senior author female		-0.08	0.24		0.12	0.11
		(0.10)	(0.10)		(0.11)	(0.10)
Mixed, other		0.01	0.04		-0.13	0.06
		(0.06)	(0.05)		(0.08)	(0.06)
Undetermined	-0.22	-0.28	-0.36	-0.04	-0.10	0.00
	(0.12)	(0.08)	(0.09)	(0.20)	(0.14)	(0.15)
Fractions of Referee Recommendations						
Reject	0.79	0.48	0.68	0.75	0.83	1.01
	(0.08)	(0.10)	(0.14)	(0.29)	(0.27)	(0.24)
No Recommendation	1.07	0.77	1.20	2.60	2.86	2.95
	(0.14)	(0.17)	(0.24)	(0.34)	(0.26)	(0.23)
Weak R&R	1.52	1.36	1.39	3.14	3.18	3.33
	(0.12)	(0.15)	(0.19)	(0.33)	(0.25)	(0.23)
R&R	2.37	1.57	1.75	4.85	4.69	4.63
	(0.22)	(0.18)	(0.18)	(0.40)	(0.31)	(0.24)
Strong R&R	2.99	1.82	2.02	5.88	5.71	5.44
	(0.31)	(0.28)	(0.31)	(0.48)	(0.33)	(0.32)
Accept	2.55	1.99	2.35	5.35	5.52	5.54
	(0.28)	(0.29)	(0.37)	(0.37)	(0.33)	(0.30)
Author Publications in 35 High-Impact Journals (Max across Authors)						
1 Publication	0.36	0.28	0.21	0.03	0.12	-0.09
	(0.07)	(0.06)	(0.10)	(0.09)	(0.09)	(0.14)
2 Publications	0.46	0.49	0.51	0.19	0.06	0.35
	(0.09)	(0.06)	(0.09)	(0.14)	(0.10)	(0.15)
3 Publications	0.40	0.60	0.65	-0.00	0.26	0.09
	(0.13)	(0.06)	(0.09)	(0.15)	(0.09)	(0.14)
4-5 Publications	0.83	0.76	0.89	0.41	0.39	0.22
	(0.10)	(0.07)	(0.11)	(0.14)	(0.10)	(0.14)
6+ Publications	0.68	0.98	1.14	0.26	0.45	0.40
	(0.19)	(0.08)	(0.10)	(0.21)	(0.10)	(0.12)
R&R Indicator	0.15	0.19	-0.09			
(Mechanical Publ. Effect)	(0.25)	(0.21)	(0.24)			
Control Function for Selection	0.37	0.22	0.38			
(Value Added of the Editor)	(0.15)	(0.13)	(0.14)			
Editor Leave-out-Mean R&R				3.27	4.09	2.80
Rate				(1.20)	(1.00)	(0.95)
Controls for Field & Gender-Field Ctrls	Yes	Yes	Yes	Yes	Yes	Yes
Indicators for Journal-Year	Yes	Yes	Yes	Yes	Yes	Yes
N	4,639	6,406	4,102	4,639	6,406	4,102
R^2 / pseudo R^2	0.26	0.25	0.27	0.53	0.50	0.48

Notes: The sample for all models is 15,147 non-desk-rejected papers with at least two referees assigned. Dependent variable for OLS models in Columns 1-3 is asinh of Google Scholar citations. Dependent variable in probit models in Columns 4-6 is indicator for receiving revise and resubmit decision. The control function for selection in Columns 1-3 is calculated using predicted probabilities based on Columns 4-6. Standard errors clustered by editor in parentheses.

Online Appendix Table 4. Models of Alternative Measures of Citations

	OLS Model for asinh(GS Citations)	OLS Model for Log(1+GS Citations)	OLS Model for GS Citation Percentile	Probit Model for Top Group of GS Citations	Probit Model for Top 2% of GS Citations
	(1)	(2)	(3)	(4)	(5)
Authors' Genders					
All Female	0.22	0.19	3.60	0.13	0.17
	(0.05)	(0.05)	(0.85)	(0.05)	(0.09)
Mixed-Gender Author Team senior author female	0.06	0.05	0.51	0.10	0.07
	(0.07)	(0.06)	(1.10)	(0.06)	(0.11)
Mixed, other	0.01	0.01	0.07	-0.00	0.00
	(0.04)	(0.03)	(0.64)	(0.03)	(0.05)
Undetermined	-0.31	-0.27	-5.05	-0.19	-0.18
	(0.05)	(0.04)	(0.81)	(0.07)	(0.11)
Fractions of Referee Recommendations					
Reject	0.64	0.54	10.55	0.30	0.30
	(0.06)	(0.05)	(0.95)	(0.08)	(0.13)
No Recommendation	0.98	0.84	16.04	0.55	0.51
	(0.10)	(0.09)	(1.52)	(0.11)	(0.21)
Weak R&R	1.45	1.24	23.32	0.79	0.72
	(0.10)	(0.09)	(1.48)	(0.11)	(0.18)
R&R	1.89	1.61	30.57	1.09	0.76
	(0.13)	(0.12)	(1.96)	(0.14)	(0.21)
Strong R&R	2.26	1.94	36.48	1.21	0.94
	(0.22)	(0.20)	(3.10)	(0.21)	(0.26)
Accept	2.30	1.99	36.46	1.34	1.19
	(0.19)	(0.18)	(2.47)	(0.20)	(0.26)
Author Publications in 35 High-Impact Journals					
1 Publication	0.29	0.25	4.52	0.19	0.20
	(0.04)	(0.04)	(0.70)	(0.05)	(0.11)
2 Publications	0.49	0.42	7.52	0.32	0.35
	(0.04)	(0.03)	(0.58)	(0.05)	(0.07)
3 Publications	0.58	0.50	9.08	0.32	0.39
	(0.04)	(0.03)	(0.55)	(0.05)	(0.08)
4-5 Publications	0.80	0.70	12.11	0.50	0.57
	(0.06)	(0.05)	(0.80)	(0.05)	(0.08)
6+ Publications	0.99	0.86	14.82	0.67	0.78
	(0.05)	(0.04)	(0.76)	(0.05)	(0.07)
Gender-field controls					
Share female in sub-fields	-0.01	-0.05	1.98	-0.45	0.09
	(0.25)	(0.21)	(3.69)	(0.26)	(0.51)
Fraction of gender-topic sub-fields	0.02	0.03	0.37	0.03	0.03
	(0.10)	(0.09)	(1.52)	(0.12)	(0.20)
R&R Indicator	0.06	0.11	-0.69	0.21	0.33
(Mechanical Publ. Effect)	(0.14)	(0.13)	(2.24)	(0.13)	(0.18)
Control Function for Selection	0.32	0.27	5.48	0.17	0.11
(Value Added of the Editor)	(0.08)	(0.08)	(1.28)	(0.08)	(0.10)
Controls for No. of Authors	Yes	Yes	Yes	Yes	Yes
Controls for Field	Yes	Yes	Yes	Yes	Yes
Indicators for Journal-Year	Yes	Yes	Yes	Yes	Yes
R^2 / pseudo R^2	0.27	0.28	0.20	0.15	0.16

Notes: The sample for all models is 15,147 non-desk-rejected papers with at least two referees assigned. Standard errors clustered by editor in parentheses.

Online Appendix Table 5. Models with Censoring of Citations

	OLS Model for asinh(GS Citations) All Years	Tobit Model for asinh(GS Citations) All Years	Tobit Model for asinh(SSCI Citations) 2006-2010	OLS Model for asinh(GS Citations) 2006-2010	Tobit Model for asinh(SSCI Citations) 2006-2008	OLS Model for asinh(GS Citations) 2006-2008
	(1)	(2)	(3)	(4)	(5)	(6)
Authors' Genders						
All Female	0.22	0.27	0.14	0.24	0.32	0.27
	(0.05)	(0.07)	(0.09)	(0.06)	(0.15)	(0.11)
Mixed-Gender Author Team senior author female	0.06	0.05	0.07	0.11	-0.04	0.12
	(0.07)	(0.08)	(0.18)	(0.10)	(0.27)	(0.13)
Mixed, other	0.01	0.01	-0.18	-0.02	-0.02	0.10
	(0.04)	(0.05)	(0.10)	(0.05)	(0.11)	(0.09)
Undetermined	-0.31	-0.41	-0.41	-0.40	-0.61	-0.52
	(0.05)	(0.07)	(0.25)	(0.06)	(0.27)	(0.11)
Fractions of Referee Recommendations						
Reject	0.64	0.86	0.93	0.69	0.80	0.48
	(0.06)	(0.08)	(0.17)	(0.09)	(0.21)	(0.14)
No Recommendation	0.98	1.26	1.91	1.07	1.58	0.87
	(0.10)	(0.12)	(0.27)	(0.12)	(0.26)	(0.15)
Weak R&R	1.45	1.84	1.95	1.53	1.58	1.31
	(0.10)	(0.11)	(0.24)	(0.12)	(0.28)	(0.14)
R&R	1.89	2.39	2.59	2.10	1.90	1.74
	(0.13)	(0.14)	(0.30)	(0.14)	(0.37)	(0.20)
Strong R&R	2.26	2.83	3.45	2.56	2.48	2.08
	(0.22)	(0.23)	(0.49)	(0.23)	(0.51)	(0.31)
Accept	2.30	2.87	3.94	2.65	3.17	2.41
	(0.19)	(0.19)	(0.41)	(0.17)	(0.39)	(0.23)
Author Publications in 35 High-Impact Journals						
1 Publication	0.29	0.38	0.31	0.25	0.46	0.28
	(0.04)	(0.05)	(0.08)	(0.05)	(0.13)	(0.11)
2 Publications	0.49	0.62	0.63	0.49	0.86	0.61
	(0.04)	(0.05)	(0.11)	(0.05)	(0.18)	(0.09)
3 Publications	0.58	0.72	0.65	0.62	0.76	0.66
	(0.04)	(0.05)	(0.09)	(0.04)	(0.16)	(0.09)
4-5 Publications	0.80	0.97	1.06	0.78	1.24	0.89
	(0.06)	(0.07)	(0.11)	(0.06)	(0.14)	(0.10)
6+ Publications	0.99	1.15	1.24	1.00	1.23	1.02
	(0.05)	(0.06)	(0.13)	(0.06)	(0.19)	(0.09)
Gender-field controls						
Share female in sub-fields	-0.01	0.10	0.74	-0.01	0.47	-0.46
	(0.25)	(0.31)	(0.61)	(0.31)	(0.97)	(0.50)
Fraction of gender-topic sub-field	0.02	0.04	0.04	0.19	-0.09	0.19
	(0.10)	(0.12)	(0.28)	(0.18)	(0.36)	(0.23)
R&R Indicator	0.06	-0.10	0.73	-0.05	1.63	0.23
(Mechanical Publ. Effect)	(0.14)	(0.16)	(0.35)	(0.17)	(0.40)	(0.23)
Control Function for Selection	0.32	0.44	0.52	0.49	0.07	0.30
(Value Added of the Editor)	(0.08)	(0.09)	(0.20)	(0.10)	(0.23)	(0.13)
Controls for No. of Authors	Yes	Yes	Yes	Yes	Yes	Yes
Controls for Field	Yes	Yes	Yes	Yes	Yes	Yes
Indicators for Journal-Year	Yes	Yes	Yes	Yes	Yes	Yes
Average R&R for All-Female Papers	0.122	0.122	0.122	0.122	0.124	0.124
Counterfactual R&R for All-Female Papers under Cite-Max	0.191	0.191	0.147	0.189	0.215	0.214
N	15,147	15,147	8,186	8,186	4,507	4,507
R^2 / pseudo R^2	0.27			0.25		0.24

Notes: The sample is non-desk-rejected papers with at least two referees assigned. Columns 3-4 restricts to years 2006-2010 and Columns 5-6 further restricts to years 2006-2008 to allow for time for SSCI citations to accrue. Standard errors clustered by editor in parentheses.

Online Appendix Table 6. Citations and Editor Decision, Heterogeneity

	OLS Models for Asinh of Google Scholar Citations			Probit Models for Receiving Revise-and-Resubmit		
	(1)	(2)	(3)	(4)	(5)	(6)
Authors' Gender						
All Female	0.20	0.10	0.28	0.06	0.05	0.03
	(0.05)	(0.10)	(0.08)	(0.07)	(0.17)	(0.08)
Mixed-Gender	0.06	0.10	0.07	0.01	0.01	0.06
	(0.04)	(0.08)	(0.05)	(0.06)	(0.08)	(0.07)
Undetermined	-0.31	-0.31	-0.31	-0.05	-0.05	-0.05
	(0.05)	(0.05)	(0.05)	(0.08)	(0.08)	(0.08)
Authors' Genders and Publications						
All Female *	0.13			-0.24		
(Max Publication >=3)	(0.15)			(0.18)		
Mixed-Gender *	-0.11			0.10		
(Female pub 3+, Male Pub<3)	(0.10)			(0.09)		
Mixed-Gender *	-0.18			-0.13		
(Female pub <3, Male Pub 3+)	(0.06)			(0.09)		
Mixed-Gender *	0.22			0.23		
(Female pub 3+, Male Pub 3+)	(0.13)			(0.11)		
Authors' Genders and Field						
All Female *		0.75			-0.27	
Share females in Sub-field		(0.61)			(1.06)	
Mixed-Gender *		-0.53			-0.05	
Share females in Sub-field		(0.49)			(0.57)	
Authors' Genders and Year of Submission						
All Female *			-0.13			-0.04
(Years of Submission 2010 on)			(0.10)			(0.11)
Mixed-Gender *			-0.09			-0.11
(Years of Submission 2010 on)			(0.08)			(0.07)
R&R Indicator	0.07	0.06	0.06			
(Mechanical Publ. Effect)	(0.14)	(0.14)	(0.14)			
Control Function for Selection	0.32	0.32	0.32			
(Value Added of the Editor)	(0.08)	(0.09)	(0.09)			
Editor Leave-out-Mean R&R				3.43	3.42	3.40
Rate				(0.74)	(0.73)	(0.73)
Controls for Author Publications	Yes	Yes	Yes	Yes	Yes	Yes
Controls for No. of Authors	Yes	Yes	Yes	Yes	Yes	Yes
Controls for Field & Gender-Field	Yes	Yes	Yes	Yes	Yes	Yes
Indicators for Journal-Year	Yes	Yes	Yes	Yes	Yes	Yes
N	15,147	15,147	15,147	15,147	15,147	15,147
R^2 / pseudo R^2	0.27	0.27	0.27	0.49	0.49	0.49

Notes: The sample for all models is non-desk-rejected papers with at least two referees assigned. Standard errors clustered by editor in parentheses. Dependent variable for OLS models in Columns 1-3 is asinh of Google Scholar citations. Dependent variable in probit models in Columns 4-6 is indicator for receiving revise and resubmit decision. The control functions for selection in Columns 1-3 are calculated using predicted probabilities based on Columns 4-6.

Online App. Table 7. Abstract Complexity, Impact of Author Team Gender

	Mesure of Complexity of Abstract			
	Gunning Fog	**Coleman-Liau**	**Gunning Fog**	**Coleman-Liau**
	(1)	(2)	(3)	(4)
Authors' Genders				
All Female	-0.05	0.11	0.50	0.18
	(0.07)	(0.05)	(0.30)	(0.19)
Mixed-Gender Author Team	0.29	0.12	0.07	0.09
senior author female	(0.11)	(0.07)	(0.31)	(0.22)
Mixed, other	0.12	0.04	-0.13	0.11
	(0.06)	(0.04)	(0.20)	(0.13)
Undetermined	0.32	-0.02	-0.51	0.16
	(0.08)	(0.05)	(0.33)	(0.25)
Author Publications in 35 High-Impact Journals (Max across Authors)				
1 Publication	-0.10	0.05	-0.24	-0.19
	(0.05)	(0.04)	(0.23)	(0.15)
2 Publications	-0.10	0.11	-0.11	-0.14
	(0.07)	(0.04)	(0.25)	(0.26)
3 Publications	-0.21	-0.03	-0.43	-0.34
	(0.08)	(0.05)	(0.26)	(0.20)
4-5 Publications	-0.23	-0.03	0.19	-0.04
	(0.07)	(0.05)	(0.24)	(0.16)
6+ Publications	-0.18	-0.06	-0.11	-0.33
	(0.07)	(0.05)	(0.23)	(0.15)
Sample	Rejected and Desk-Rejected Papers		R&R Papers Only	
Controls for Author Publications	Yes	Yes	Yes	Yes
Controls for No. of Authors	Yes	Yes	Yes	Yes
Controls for Field	Yes	Yes	Yes	Yes
Indicators for Journal-Year	Yes	Yes	Yes	Yes
Mean of the Dependent Variable:	19.4	15.3	19.3	15.4
N	27,545	27,545	2,366	2,366
R-squared	0.02	0.02	0.04	0.04

Notes: Dependent variables are measures of reading complexity. The Gunning fog index is as 0.4[(words/sentences) + 100(complex words/words)], where complex words are tri-syllabic words, excluding common suffixes and proper nouns. The Coleman-Liau index is calculated as 0.0588(letters/words) - 0.296(sentences/words) - 15.8. Robust standard errors in parentheses.