Are the Referees and Editors in Economics Gender Neutral? Pre-Analysis Plan*

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

June 2018

Abstract

Do men under-value the scientific contributions of women? We study the effects of gender on the evaluation of economic research using data on submissions to four leading journals, matched to referee recommendations, editorial decisions, and subsequent citations. A customized name-matching algorithm allows us to classify the genders of 97% of all authors and referees with an error rate of under 1%. About one-fifth of submitting authors are female, though rates vary widely across subfields. The fractions of female referees are similar. We begin by examining whether editors are more likely to match a female-authored paper with a female referee, suggesting an awareness of possible gender differences in recommendations. We then address four main sets of questions. First, do male and female referees assess papers differently, and does the gender composition of authors matter for how different referees rate a paper? Second, how reliable are the assessments of male and female referees in predicting future citations, and does this vary with the gender of the authors? Third, how do editors weigh the recommendations of different referees against the information contained in prior publications and other author characteristics, including gender? Fourth, are there gender-related differences in the time that referees take to make a recommendation, or that editors take to reach an initial decision? We compare our findings to the results from a survey of economists, and use the survey results to help interpret any gender gap (or lack thereof) in referees' and editors' decisions.

*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 are also grateful to Andrew Tai, Alden Cheng and Johannes Hermle for their extraordinary help.

1 Introduction

Women are under-represented in the top ranks of many professions, including academia. While numerous explanations have been offered for this gap (see e.g., Niederle and Vesterlund, 2010; Bertrand, Goldin, and Katz, 2010; Ceci et al., 2014), an abiding concern is that some combination of implicit and explicit biases causes men to systematically under-value the contributions of women. This concern is particularly salient in economics, where the vast majority of senior faculty, journal editors, and referees are male (see e.g., Ginther and Kahn, 2004; Bayer and Rouse, 2016).

Existing evidence on the presence of gender biases in the evaluation of economic research is mixed. Blank (1990) reported on the results of an experiment in which articles submitted to The American Economic Review were randomly assigned to a double- or single-blind category for review. She found no significant differences in the acceptance rates by gender for single- or double-blind reviewing. Broder (1993), based on reviews of grant proposals to the National Science Foundation (NSF), finds that female reviewers rate female-authored papers lower than do their male colleagues. Abrevaya and Hamermesh (2012), using data from a single journal, find that female referees are no more supportive than males of papers by female authors.[1] 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 females 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 (2012) 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 use a large and detailed database of submissions to four leading journals in economics – the *Journal of the European Economics Association*, the *Quarterly Journal of Economics*, the *Review of Economics and Statistics*, and the *Review of Economic Studies* – to study the role of gender in the evaluation of economics research. We use a combination of name-based algorithms and hand search to assign genders to the co-authors and referees of each submission.[2] We combine characteristics of the submission – including field, the gender composition of the authors, and their previous publication record – with the assessments of the referees, the decision of the editor, and ultimate citations received by the paper, regardless of whether it was accepted or not. We use these data to analyze gender differences in how papers are assigned to referees, how they are reviewed, and how the editor uses the inputs from the referees and other information about the paper to reach a revise and resubmit verdict. We also use the number of days referees take to submit their recommendations to test for gender differences in the public good provision of timely reports, as well as to test if the number of days of referee recommendations and editors' decisions differ for male

[1]It should be noted that the estimates by Abrevaya and Hamermesh, 2012 (Table 4) do not rule out moderately large effects (on the order of a 10 percentage point boost or penalty in the probability of a favorable recommendation) from having a female referee assess a female-authored paper.

[2]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.

and female authored papers, both in the first submission as well as in later revisions.

We complement our data base of journal submissions with a 14-question survey of economists covering the main topics above. The survey questions are included in this analysis plan. We intend to compare, both qualitatively and quantitatively, the results from our main analysis with the expectations of the survey population. We will also use the survey to elicit quantitative beliefs about the citation-quality link, which will then be used to interpret the relationships between referee recommendations, editor decisions, and realized citations.

This draft, which was written prior to the completion of our data collection procedures, outlines the key steps of our envisioned analysis. It describes the planned tables and figures that will summarize descriptive facts about the authors, referees, and decision processes at the four journals. It then lays out the main empirical specifications we will use to analyze the data and test for the presence of gender differences. By pre-specifying our research design we hope to address concerns over data mining and p-hacking (see Christensen and Miguel, forthcoming). Such concerns are particularly relevant for observationally based research designs on a topic like gender, since there is a presumption that "statistically significant" results will be more favorably received by academic and non-academic audiences.

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, using a combination of existing data and manual coding. We provide evidence on the accuracy of our gender coding and the error rates in classifying authors as males versus females.

In Section 3 we outline our planned descriptive overview of the submissions data. Our data base builds on the sample originally collected by Card and Dellavigna (2017) − hereafter, CDV − but adds three key pieces of information: the gender of each author; the previous publication record of each author[3]; and the gender of each referee. We also gathered more granular information on waiting times in the review process (e.g., the waiting time for reports from each referee and the time required by authors to revise their papers) enabling us to study the sources of any differential gap in the time between submission and later stages of the decision process. Finally, we gathered information on the complexity of the abstract.

Author gender information allows us to classify the vast majority of papers (>95%) into 3 broad categories: female-authored, male-authored, and mixed gender.[4] Preliminary data from one journal suggests that about two-thirds of papers are written by males (i.e., a single male author or an all-male team of coauthors), 8% are written by females, and just under a quarter are written by mixed-gender teams. We also use the publication records of each co-author to classify mixed gender teams by whether the most-published co-author is female or male. Similarly, we classify referees as female or male, and by their recent publication record.

In Section 4 we turn to the set of papers that are sent out for refereeing and analyze the "matching" process used by editors to assign papers to referees. (We defer an analysis of the desk rejection process to Section 6). Here a major factor is the field of the paper. Consistent with earlier studies

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

[4]We cannot assign gender for about 3% of authors. In our empirical analysis we therefore include additional categories for papers with a coauthor of unknown gender. The rate of missing information on the gender of referees is lower (1%).

(e.g., Chari and Goldsmith-Pinkham, 2017) we find that the fraction of female authors is lower in some fields (e.g., macro) and higher in others (e.g., labor economics). We examine whether the fraction of female referees assigned to papers in a field tracks the fraction of female authors, and whether editors are more likely to assign a female-authored paper to a female referee, even controlling for field. In fact, we propose to use this pattern of assignment as an indicator of how editors classify mixed-gender papers. Consider, for example, mixed-gender papers for which a female is the "senior author" – i.e., has the most prior publications – versus one in which a male is senior author. One possibility is that editors treat the former like all-female papers and the latter like all-male papers, lending support for a simplified binary classification of papers. We also test for any tendency of editors to assign papers with a highly prolific author or co-author to referees with more extensive publication records, and control for this in our analysis of gender-matching.

In Section 5 we study the recommendations of referees. We begin with a simple analysis that expresses the referee's recommendation as a function of observed paper characteristics, the characteristics of the authors, and the characteristics of the referee him- or herself. Our most general models will include paper fixed effects, which will allow us to abstract from unobserved features of the paper and focus on differential assessments by male and female referees *of the same paper.* This design builds on a similar specification by Abrevaya and Hamermesh (2012) but we will have a much larger submission sample and a different set of journals.

In Section 6 we then turn to the question of how referee assessments are related to future citations - a natural benchmark for judging the quality of papers. Here the key question is whether citations for female-authored papers are higher or lower, conditional on the recommendations of the referees. Such a gap can be interpreted as a measure of gender-bias (i.e. if higher/lower citations then negative/positive bias towards female authored papers). A major confound, however, is the possibility that female-authored papers are over/under-cited by other researchers (see CDV). To address this concern we conduct a survey of economists that elicits beliefs about the relative gap between citations and quality for female versus male-authored papers. We will use the results from this survey to interpret any gender gap in citations conditional on referee support.

We then focus on editors, who use information from the referee recommendations and their own reading of the paper to reach a decision to ask for a revision or not. Earlier work by CDV shows that editors tend to closely follow the averaged recommendations of the referees, with pseudo R-squared fits for the revise and resubmit (R&R) decision of around 0.5. We will estimate simple models of the revise and resubmit (R&R) decision that controlling for the averaged referee recommendations, the publication record of the authors, and the gender composition of the author team. This will allow us to asses whether female-authored papers are more or less likely to receive a positive R&R verdict. Again, we use the results from our survey to interpret the gap between citations given the referee recommendations and the editors' decisions.

In Section 7 we generalize our models of the link from referee recommendations to citations and the editors' decision by allowing for higher or lower information content in the recommendations of female versus male referees, and for potential differences in the two groups' enthusiasm for papers by authors of their own gender versus the opposite gender. We build a simple model of the weighting for the recommendations of different referees, allowing this weight to vary with gender and previous publication record of the referee, the gender of the authors, and interactions.

In Section 8 we will study the impacts of gender on delay times in the evaluation process, including effects of author gender on the time that referees take to return a recommendation and the subsequent time that editors take to reach a decision. We will also evaluate whether female referees tend to return their reports more or less quickly, and whether any gap is different when the paper is written by a female team versus a male team. These results will be potentially useful in interpreting any evidence of differential assessments of female authored papers by male versus female referees, and of any gender gap in editors' R&R decisions.

In Section 9 we consider additional envisioned components of the analysis. For example, we also compare the complexity of abstracts in papers written by male authors, versus female authors, inspired by Hengel (2017).

In Section 10 we outline the survey, which we aim to use in three ways. First, as discussed above, we use it to elicit the beliefs about how the disconnect between citation and quality varies for male-authored and female-authored papers. Second, we collect qualitative opinions about the role of gender in the editorial process in economics. Third, as in DellaVigna and Pope (forthcoming) we aim to compare how the beliefs compare to the actual features observed in the data.

In the rest of this pre-analysis plan, we present the format of how we envision the main results. For the pre-analysis content, we use data from only one journal, with the distribution of the author gender randomly assigned across papers to match the actual distribution of gender across papers. Thus, the tables and figures do not present actual results, but rather our plan of analysis, and an indication of the statistical power that we are likely to have with a sample that will be more than 4 times as large.

2 Assigning Gender to Economists

A major issue confronting any analysis of gender in the journal review process is that (to the best of our knowledge) no journals in economics collect information on the gender of authors or referees. In the absence of this information we follow the widespread practice of assigning gender based on names. In an effort to maximize the fraction of names with reliable gender information we developed a multi-step process for assigning gender 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 names of the authors and referees at each journal.

To conduct our analysis we began by assembling a "test" data set from EconLit of the names of all authors who have published in 53 economics journals (listed in Online Appendix Table 1) from 1990 to mid-2017. This yields a data set of 48,000 unique names. We supplement this list with the names of authors and referees from one of the four journals in our sample.

We then compared these names to four data sources:

1. The R-package "gender," which uses U.S. Social Security data on given names to calculate the fraction of people with a given name who are male, *p(Male)*.

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.[5]

4. A list of female members of the European Economic Association, an initiative of the Committee on Women in Economics.[6]

We classified an author name from the test data set as male if one of the US/German datasets assign $p(Male) \geq 0.99$ to the author's first (given) name and the other assigns $p(Male) \geq 0.50$. An analysis using hand lookups of an audit sample showed that this criteria yields an average false positive rate for being classified as male of less than 1% for names in the test data set.[7] We classify an author as female if both the US and German datasets assign $p(Male) \leq .01$ for the author's first name, **or** if the **full name** is present in either the RePEc or EEA lists of female economists. Again, an audit showed a false positive rate for being classified as female of less than 1% for names in the test data set.[8]

For all names in the test data set that could not be assigned to either male or female, we then used a team of undergraduate research assistants to hand code as many as possible, 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.[9]

Based on the results of this analysis, we proceed as follows: For each journal in our data base we first retrieved a list of the names of all authors and referees during the period the journal was using Editorial Express.[10] We then followed the same steps as in our test data set, first assigning gender using the four sources above, and then hand-coding the remaining names. We present more information on our process, the audits, and the final resolution of names in the Appendix.

Finally, for our main data extraction, we prepared a program that could run in the editorial office of each journal and access submissions as well as the journal-specific lists of author and refeee names with gender attached. This program prepared an anonymized data set with gender information on authors and referees for each submission that could be linked to the anonymized data set originally prepared by CDV.

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

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

[7] Specifically we instructed undergrad research assistants to searching the name, 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.

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

[9] Again we checked the reliability of the hand-coding process by having a fraction of names double-coded. We found that the coders agreed on gender 74% of the time; that one of two coders found enough evidence to determine a gender 14% of the time; that 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 to be correct.

[10] These names were extracted by staff at each journal and sent to us prior to running our main data extraction program with no other linked information.

3 Descriptive Overview

We begin by categorizing all submitted papers into four groups, based on the gender composition of the author team: 1) all male, 2) all female, 3) mixed gender, and 4) undetermined. The last group encompasses papers that cannot be conclusively categorized because of ambiguous names.

3.1 Summary Statistics

Table 1 presents a descriptive comparison of papers with different author gender compositions, using a test data set formed by randomly assigning gender to authors and referees for the submissions at one of the four journals in our sample.[11] We emphasize that the actual estimates in the table (and all subsequent tables and figures in this document) are uninformative. Nevertheless, the standard errors are potentially informative and should be roughly 2 times larger than we anticipate from our final sample. The table will highlight the differences in papers from male author teams, female teams, and mixed gender teams. We anticipate that less the 3% of papers will have an unknown gender for one or more co-authors.

Figure 1 will present the distribution of the editorial decisions for teams of authors of different genders (Figure 1a), the distribution of referee recommendations (Figure 1b), and the distribution of the number of papers published by the authors of submitted papers in the previous 5 years in a list of 35 high-impact journals (Figures 1c and 1d).

4 Gender of Referees vs. Authors

How does the gender composition of referees compare to that of authors? Figure 2 will show, for each of 12 main "fields" (based primarily on the first letter of the JEL code) the fraction of all authors in our EconLit data base who are female (by year of publication), the fraction of all authors of submitted papers in our preliminary data set who are female (by year of submission), and the fraction of all responding referees who are female (by year of submission). In general we expect that these will track closely within a given field, but will show relatively large differences between fields in the fraction of female authors and referees.

In Figure 3 we will examine whether editors are more likely to assign a paper by a female author to a female referee, even controlling for field. 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 (i.e., the coauthor with the most prior publications 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.

Table 2 presents a more detailed analysis. Using a simple linear probability model we will examine whether female-authored papers are more or less likely to be assigned a female referee, controlling for field and the publication records of the coauthors. We will also estimate similar models to

[11] We assign female at approximately the correct underlying fraction.

check whether papers written by a coauthor team with many prior publications is more likely to be assigned to at least one referee with many prior publications, and whether this process differs by the gender composition of the author team. Our classification of prior publication histories will follow (to the extent possible) the classification used in CDV. We anticipate that this set of results will appear in an appendix unless we find a large discrepancy between the rate at which papers by male and female coauthor teams are assigned to more prolific referees.

5 Referee Recommendations and Author Gender

Figure 4a presents our planned analysis of how the gender of referees affects the assessment of papers by coauthor teams with different gender mixes. We construct the figure by assigning an index to each categorical referee recommendation (e.g., "Reject") based on the predicted $asinh(citations)$ associated with that recommendation, as constructed using the coefficients in the cites model in Card and Dellavigna (2017).[12] We then tabulate mean assessments of papers that are assigned to both male and female referees for each referee gender group, separately by the author gender group. Observations are weighted by the number of referee reports for the paper to ensure that each paper receives equal weight. The bars will show ± 2 standard error intervals, constructed by clustering at the paper level.

The simple comparisons in Figure 4a will allow us to test whether female referees are more or less positive than males in their average assessments; and whether they are *relatively* more positive in their assessment of female-authored papers. A finding of a *relative assessment gap* could be interpreted as evidence that male referees impose a higher bar for female-authored papers, or that female referees apply a lower bar for female-authored papers. A null effect would indicate, to a first approximation, that neither force is at play. In Figure 4b we present similar comparisons but using the share of recommendations that are positive (R&R or stronger). The standard error bands in Figures 4a-b indicate that we will have substantial statistical power for our test since we will have a sample that is four times as large as our test sample.

Table 3 presents a series of OLS models of referee-specific recommendations that push further by adding additional controls for other characteristics of the author team and of the referee. We will fit two types of models: a set of probit models for recommendation of "revise and resubmit" or higher, and a set of ordered probit models for the referee's actual recommendation. (In this preliminary analysis we substitute linear regression models for the strength of the referee's recommendations based on the same index used to construct Figure 4a). Our primary interest here will be in the magnitude of the interactions between the gender composition of the author team and the gender of the referee. The most general models will include paper fixed effects: in this case the interaction effect is identified using only papers with at least one male and one female referee. These models compare assessments of referees for the same paper, providing the strongest test for any tendency of female referees to offer relatively more positive (or more negative) assessments of papers by female

[12] 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.

teams. We will cluster the standard errors in these analyses at the paper level to allow for correlations in the different referee's views about the same paper.

6 Evidence on Citations and Editor R&R Decisions

Panel a of Figure 6 plots mean $asinh(Google\,Scholar\,citations)$ by decile of the predicted probability of a revise and resubmit (R&R) decision for each of the three main author-gender groups.[13] We show the relationship separately for papers that were rejected and those that were invited for revision by the editor. The R&R prediction model underlying the figure includes referee recommendations and author and paper characteristics (including the gender composition of the author team), as well as journal and year effects.

Panel b of the figure presents a similar analysis for papers which received a desk rejection versus no desk rejection – though here the prediction model is only based on paper and author characteristics, since the desk reject decision is made prior to referee assignment. We include smoothing lines obtained via cubic fits to all data points.

The idea of these graphs is to show visually whether expected citations vary by author gender, conditional on the probability of R&R (or the probability of desk rejection) and the editor's decision. If editors treat papers by male and female authors similarly *and there is no gender bias in the citation process* then we would expect to see very similar average citations by gender of the author(s), once we condition on the probability of an R&R and the editor's ultimate decision.[14] If editors discriminate against female-authored papers, however, we expect to see **more** citations for female authored papers, conditional on the probability of R&R and the referee decision.

We will analyze the inter-relationships between referee assessments, the editor's decision, and citations using the framework developed in CDV. This leads to a parallel set of regression models for citations as a function of referee assessments and other characteristics of a paper (including gender of the authors), and for the editor's R&R decision as a function of the same variables, as shown in Table 4.

When citations are an unbiased (but potentially noisy) measure of quality, editorial bias against female-authored papers is revealed by a "proportional difference in differences" comparison that asks how female authorship affects citations versus the probability of an R&R verdict, **relative** to the benchmark comparison of how the referee recommendations affect these two outcomes. For example, suppose that papers that receive a unanimous recommendation of R&R from the referees have 100% more citations than those that receive a unanimous recommendation of reject (i.e., a coefficient of 1.0 in a model for $asinh(citations)$), and have a coefficient of 0.5 in a Probit model for the editor's R&R decision. Suppose in addition that papers written by an all-female team receive 10% more citations, conditional on the referee recommendations and other paper characteristics. (We note

[13] For added resolution among papers with the highest probability of an R&R, we split the top decile into two vingtiles.

[14] This same idea is widely used in the race discrimination literature. For example, Knowles, Persico and Todd (2001) compare the probabilities of finding drugs in the cars driven by black and white motorists who are stopped by police. Our case is different in that we see outcomes regardless of whether the paper was R&R (analogous to being stopped by police) or not.

that such a positive gap is indicative of referee discrimination against these papers). In this case, in the absence of any editorial bias, we would expect a coefficient of $0.05 = 0.10 \times \frac{0.5}{1.0}$ for female authored papers in the Probit model for the editor's decision. This coefficient is just large enough to offset the bias in the referee recommendations and ensure that conditional on the probability of an R&R decision and the editor's decision, female authored papers the same expected number of citations as male-authored papers.

There are two confounding issues in conducting and interpreting such a comparison. One is that papers that receive a more positive referee rating are more likely to get an R&R and more likely to be published. To the extent that published papers get more citations, this leads to an upward bias in the effect of the referee opinions on citations, distorting the benchmark comparison. To deal with this "publication bias", we will follow the control function approach developed in Card and Dellavigna (2017) which uses the relative leniency of different co-editors (measured by the leave-out mean R&R rate) as an exogenous determinant of the probability of R&R.

A second issue is that papers by female authors may be "under-cited" (or "over-cited") – in other words, there may be a gender gap in the relationship between true quality and citations. If so, this will confound the "difference in differences" comparison. If the extent of the gender bias in mapping from quality to citations is known, however, this can factored out of the comparison, allowing us to evaluate any gender bias in the referee/editing process. For example, if female authored papers receive 10% fewer citations than male authored papers of the same quality, then we have to inflate the estimated effect of female-authorship on expected citations by +0.10 before we conduct our difference in differences comparison. We will use results from our survey of economists, which will elicit opinions on the degree of over- or under-citing of female authored papers, to carry out such an adjustment.

Our approach will be to estimate the series of models shown in Table 4 for citations (columns 1-4) and the editor's R&R decision (columns 5-7), then use a particular estimate of the bias in the citation-quality link to form estimates of the degree of bias of referees against female-authored papers and the degree of bias of editors against female authored papers. To illustrate, suppose the working hypothesis is that female-authored papers receive 10% fewer citations than male-authored papers, conditional on quality. Then we can take the estimated coefficient for all-female papers from any of the citation models in columns 1-4 of Table 4 and add 0.10 to infer the extent to which referees undervalue papers by female authors. If the coefficient on female-authored papers from the richest model in column column 4 is 0.10, then we infer that referees discount the quality of these papers by 20% (summing the actual coefficient in the citation model and the 0.10 adjustment to inflate female citations by 10%). We note that we might expect to see some differences in the discounting of female-authored papers by male and female referees – we discuss this under the heterogeneity analysis below.

Having inferred the bias (if any) in the referee recommendations we can then estimate the degree of editor bias by conducting the difference-in-differences comparison discussed above. Suppose (as before) that the benchmark effects of a unanimous R&R referee recommendation in the citation model and the editor's decision model are 1.0 and 0.5, respectively, and suppose the coefficient for female-authored papers in the citation model is 0.10. Finally, assume that the bias in the citation process is 0.10 (i.e., female-authored papers are under-cited by 10%). Then the expected coefficient of female authored papers in the editor's R&R decision model, assuming the editor imposes the same quality bar for female and male authored papers, is $0.10 = (0.10 + 0.10) \times \frac{0.50}{1.0}$.

We will use a series of null hypotheses for the appropriate value of the degree of gender-bias in the citation process. As a first null, we will assume that the bias is 0. As a second null, we will use median elicited response from all survey respondents and use this to adjust the citation gap between male and female authored papers. As a third null we will use the median estimate among female respondents only. We will also conduct a bounding exercise to summarize the plausible range of implied biases of referees and editors.

Heterogeneity. In Table 5, we intend to consider three main directions of heterogeneity, which we pre-specify to address concerns about multiple testing. The first one is by the extent of publications by the author team. It is possible that the discrimination patterns, if any, differ for more senior women, versus more junior women. In Columns 1 and 4 we focus on that, examining separately papers depending on whether the authors have more, or less, than 3 publications in the recent past, a rough measure of publications.

In columns 2 and 5 we focus on a second measure, differences by field, and in particular by the share of female authors in a field. Some fields, like labor economics, have a relatively higher share of female authors while other fields, like theory and econometrics, have a low share (see Figure 2). We thus examine the interaction with the share of female authors of papers in a 5-year range in the JELs of submission of the paper, a variable that we generated using Econlit date (even as we do not keep the detailed JEL codes for the paper).

A third heterogeneity dimension is by the years in the sample. Since the citations are evaluated for all papers in mid 2015, papers submitted earlier, say up to 2009, have had a longer time to accumulate citations; over this longer period factors such as conference presentations and circulation of working papers are likely to matter less than for more recent papers, papers submitted from 2010 on. Thus, the the time periods allow us to consider whether the pattern of citation accumulation differ at different horizons. (It is of course possible that other difference exist between the two time periods) Thus, we estimate models with period interactions in Columns 3 and 6.

A fourth potentially important dimension of heterogeneity is the gender of the referees. The analysis in Table 3 is meant to discern whether (for example) female referees evaluate female-authored papers more highly than male referees. If we find such a gap, we will extend the analysis in Tables 4 and 5 in three ways. First, we will limit the analysis to papers that are evaluated only by male referees. We expect that this sample will include well over half of all papers. Second, we will use the estimated coefficients from the ordered-probit and OLS models for the referee evaluations to construct "adjusted" referee recommendations. Specifically, we will use the coefficients to subtract off any boost given by female referees to female-authored papers. Third, for our most general approach we will turn to the analysis outlined in Section 8, below, which explicitly distinguishes between male and female referees.

We will not consider the heterogeneity by journals, as part of our data agreement with the journals is that we will consider only all four journals jointly.

Desk-Rejection. While all the previous analysis has focused on non-desk-rejected papers, the same issues apply to desk-rejection. We propose to analyze desk-rejections in a parallel way as in Card and Dellavigna (2017), as seen in Table 6 and Figure 5b.

7 Evidence on Weight Placed on Referee Recommendations

In Figures 7a-b and in Table 7, we examine how editors treat referee recommendations by males vs. females. It is possible that editors treat certain referee recommendations as more, or less, informative than they are. In Card and Dellavigna (2017), for example, we show that referees with more recent publications are equally informative as referees with fewer publications, when we measure informativeness as the degree to which their recommendations predict future citations of the papers they referee; and yet, editors put more weight on their recommendations in their R&R decisions. In this section, we propose to study whether editors put the appropriate weight on reports by female referees as to reports by male referees.

We will consider two setups. The first is a nonlinear model like

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

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 category of recommendation, regardless of the gender of the referee (or of the author team). It then assumes that the outcome is affected by an average of "adjusted indexes" of the form:

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

where $Female_{ij}$ is an indicator for the gender of referee j of paper i (or, more generally, a vector of dummies with interactions between the gender of the referee and the gender of the authors 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 about all papers then we would expect to estimate a negative value for α_0 in the citation regression. Likewise, if female referees tend to compress their recommendations in a narrower range then male referees, then we would expect to estimate a value for the constant $\alpha_1 > 1$. We note that with this setup we can easily account for the possibility that female referees are more positive about female-authored papers by allowing the intercept shift for R_{ij} to depend on both referee and author gender.

As a simple alternative to this model, we will estimate models of the form:

$$
\begin{aligned}
Outcome_i = \delta_{M,reject} fr(M, reject)_i + \cdots + \delta_{M,accept} fr(M, accept) + \\
\delta_{F,reject} fr(F, reject)_i + \cdots + \delta_{F,accept} fr(F, accept) + \boldsymbol{\gamma X_i} + \varepsilon_i
\end{aligned}
$$

where $fr(g, rec)$ represents the fraction of referees of gender g who make recommendation rec. This specification can be generalized by allowing the coefficients $\delta_{g,rec}$ to vary with the gender of the author team.

8 Response Time, Editorial Delays

So far we have focused on the referee recommendations and editorial decisions, but another aspect that matters is the speed of decision-making. To the extent that there is discrimination against female authors, it may appear in the form of slower decisions, as Hengel (2017) argues. We thus consider the speed of decision-making of both referees and editors.

Referee Delays. For the referees, we measure the number of days from paper submission to the submission of their report, examining, as in the previous analysis, the interaction of author gender and referee gender. Figure 8a presents the proposed figure comparison, and Table 8 presents the corresponding table analysis.

Editorial Delays. We next consider the decision making time of the editor, as function of the author gender mix. In particular, in Figure 8b and in Table 9, Columns 1-3 we present the total decision-making time for the editor, decomposed also into referee delay and editorial delay since the last referee report.

Revisions. Finally, we also consider the lengths of revision of a paper, since it is possible that editors may impose a longer and harder revision paths on certain authors. We collected three measures of such delays: the number of rounds for an R&R (Column 4), the number of days the author takes in their first resubmission (possibly capturing the difficulty of the revision, a well as author delays of any other reason, Column 5), and the remaining time from the resubmission to the final acceptance (Column 6).

Considering the results from our analysis of referee assessments (outlined in Section 5), and our analysis of biases in referee opinions and editor decisions (section 6), and the patterns of delay by referees and editors, we will try to summarize the evidence in favor or against the hypothesis of systematic bias against female authors. We may find, for example, that male referees are relatively more negative about female authored papers; that referees as a whole (and particularly male referees) under-value the expected citations that accrue to female authored papers; that editors do not compensate for the relatively negative opinions of referees about female-authored papers; and that referees and editors impose longer delays on female authors. Such a consistent pattern would suggest relatively strong evidence of systematic bias against female authors in economics. Or we may find only small and unsystematic gender gaps in all these analyses, which would suggest that the biases are small.

9 Other Analysis

Abstract Complexity. We also consider, motivated by Hengel (2017) measures of abstract complexity for papers, comparing female-authored papers an male-authored papers (Table 10). In the data we observe the abstract of only the most recent version. We thus analyze separately the abstract for rejected and desk-rejected papers, versus papers which received an R&R. Hengel (2017) finds that revisions of papers by women are more clearly written.

Years 2014-17. The main analysis uses just submissions from the beginning of the data (typically 2006) until 2013. For robustness, we also plan to consider for the editorial decisions submissions taking place in 2014-17. While we cannot consider citation outcomes for such papers, since citation data was collected only once in mid 2015, we can still test whether the behavior of editors and referees has changed over the years, and use these additional years for extra sample. For example, the tests in Table 2, 3, Columns 5-7 of Table 4, Columns 4-6 of Table 5, Column 2 of Table 6, Columns 4-6 of Table 7, Table 8, Table 9, and Table 10 can be run for the whole sample, and for these additional years separately. For the results that do not rely on citations, we may use the full sample in the text for higher power.

Control Function. We also aim to use the control function approach develop in Card and Dellavigna (2017) to account for selection of papers into the sample. For example, if we find that there is a differential selection by gender of the authors into desk rejection, we will aim to correct for this in the main analysis.

Analysis of Citations. As we discussed above, an important parameter in the analysis is whether there is a difference in citation rates for male-authored and female-authored papers. For the main analysis, we have described the null hypotheses that we will use; in particular, we will use the information from a survey of economists about this parameter. If the beliefs of economists are particularly heterogeneous in this respect, we will also consider a comparison of citations of papers by male authors and female authors, to examine whether papers by female authors, within a narrow field, are more likely to be cited by other female authors, compared to papers by male authors Ferber (1986). This would suggest, given that female economists are a minority of economists, that female-authored papers suffer in citations, compared to male-authored papers.

10 Survey Evidence

The final data collection effort consists of a survey of editors and economists about the perception of gender differences in the publication process. In particular, we inquire about the perception of (i) whether female-authored papers are more likely to be assigned to female referees, and why; (ii) whether there is discrimination towards female authors; (iii) how male and female referees evaluate male-authored and female-authored papers; (iv) how informative male and female referees are, and how much editors follow their recommendations; (v) the perception of how mixed-gender teams are regarded, compared to all-male teams and all-female teams.

As we discussed above, an important part of the survey is the elicitation of the belief about a possible disconnect between citations and quality for female-authors papers, compared to male-authored-papers. We use the median elicited discount for an alternative test of the discrimination of female authors.

We also use the information on (v) in the perception of mixed-author teams, to guide our analyses of such teams. We ask survey respondents about their perception of mixed-author teams in which the most published author is female, and the ones in which the most published author is male. We will use the modal survey response to guide the analysis. If, for example, survey respondents believe that mixed-author teams in which the most published author is female are treated similarly to all-female authors, we will pool the two groups together in the main analysis.

The survey, which is approved under Berkeley IRB 2018-04-10955, will be sent to three groups: (i) editors at the 4 journals; (ii) two hundred economists in EconLit with at least 4 publications in the top-35 journals in the 5 years up to 2017; (iii) all assistant professors of economics in top-20 American schools and top-5 European schools with PhDs in 2015-17.

We selected this group to cover different relevant angles of the professions. The beliefs of the editors are obviously relevant given their role. The second group captures the belief of economists who have reached a publication record. The third group captures instead individuals who are just starting their career as economists. Within each group, the survey is completely anonymous so we do not keep track of the respondents. Within the second and third group, though, we keep track separately of female and male respondents and we over-sample female respondents in group (ii).

References

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

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, Claudia Goldin, and Lawrence F. Katz. 2010. ”Dynamics of the Gender Gap for Young Professionals in the Financial and Corporate Sectors.” *American Economic Journal: Applied Economics*, 2 (3): 228-55.

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.

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. 2017. “What do Editors Maximize? Evidence from Four Economics Journals.”

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. Forthcoming. “Transparency, Reproducibility, and the Credibility of Economics Research.” *Journal of Economic Literature*.

DellaVigna S, Pope D. What Motivates Effort? Evidence and Expert Forecasts. *Review of Economic Studies*. Forthcoming.

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.

Ferber, M.A., 1986. Citations: Are they an objective measure of scholarly merit?. Signs: Journal of Women in Culture and Society, 11(2), pp.381-389.

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.

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

Niederle, Muriel and Lise Vesterlund. 2010. “Explaining the Gender Gap in Math Test Scores: The Role of Competition.” *Journal of Economic Perspectives*, 24 (2): 129-144.

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

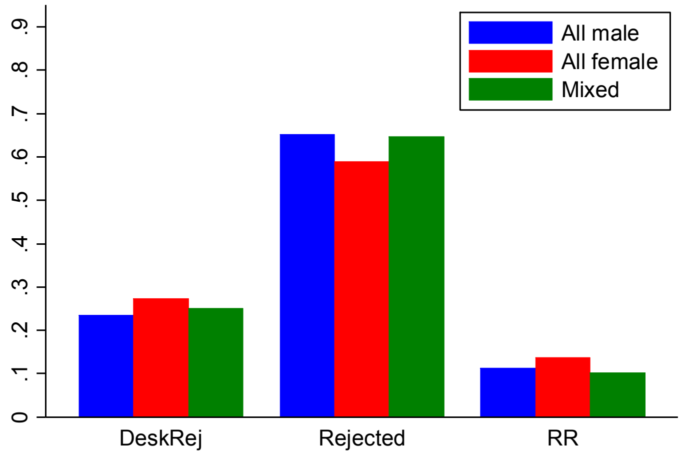


Figure 1b. Distribution of Ref. Recommendations

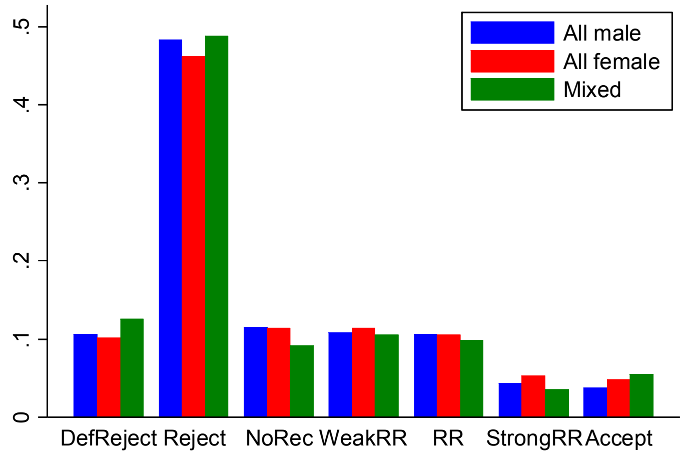


Figure 1c. Distribution of Author Publications

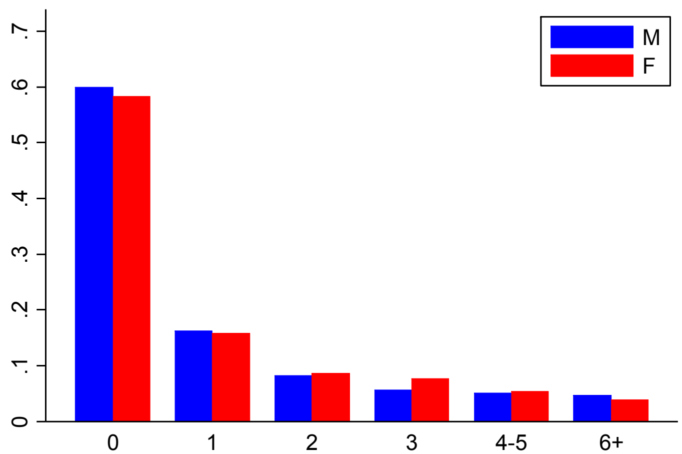
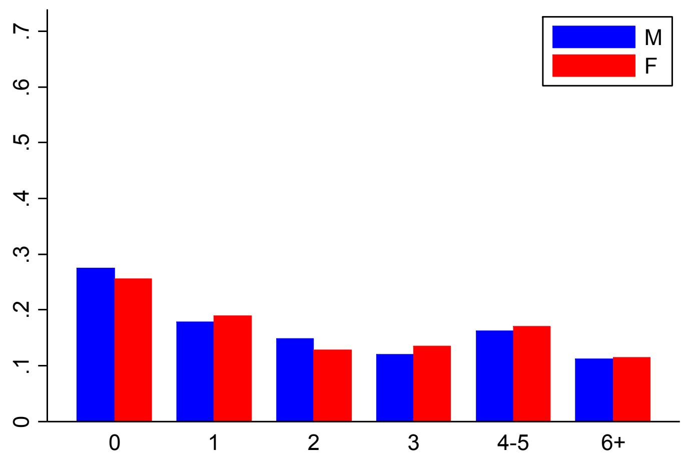


Figure 1d. Distribution of Referee Publications



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 2. Share of Female Authors and Referees, by Field

Figure 2a. Author Gender by Field

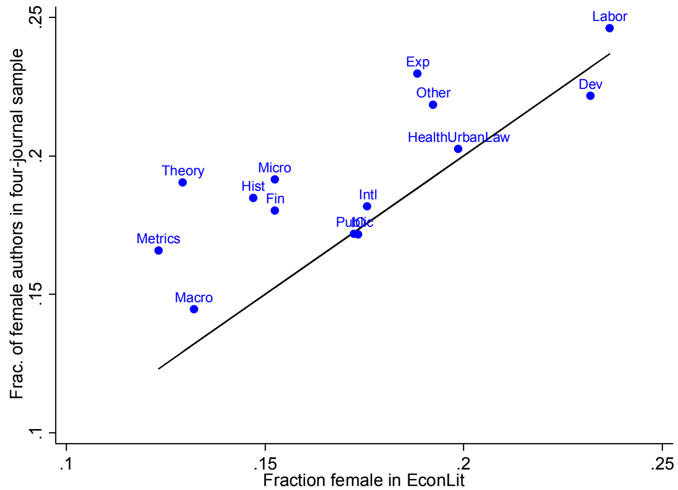


Figure 2b. Referee Gender by Field

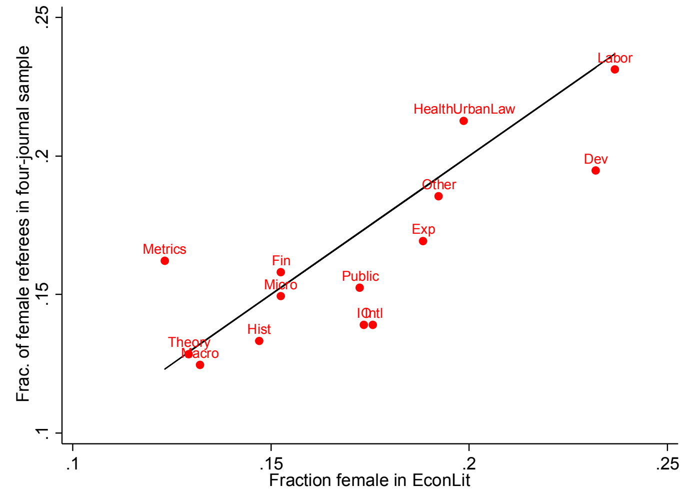
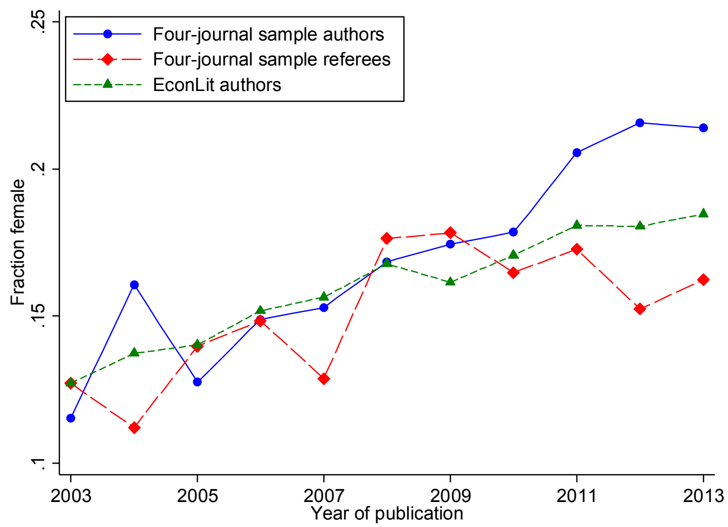
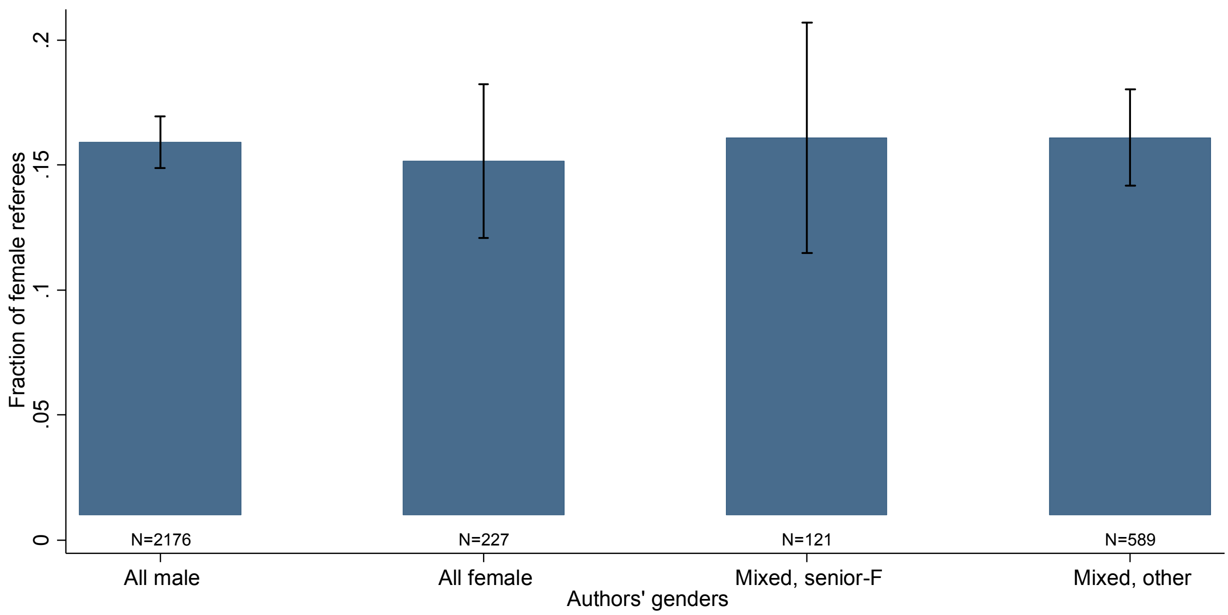


Figure 2c. Author and Referee Gender Over Time



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. Observations are at the author/referee-paper level and weighted by the inverse number of fields listed.

Figure 3. Referee Gender as Function of Author Gender



Note: Observations are at the referee times paper level.

Figure 4. Referee Evaluation by Author Gender and Referee Gender

Figure 4a. Index of Referee Recommendations

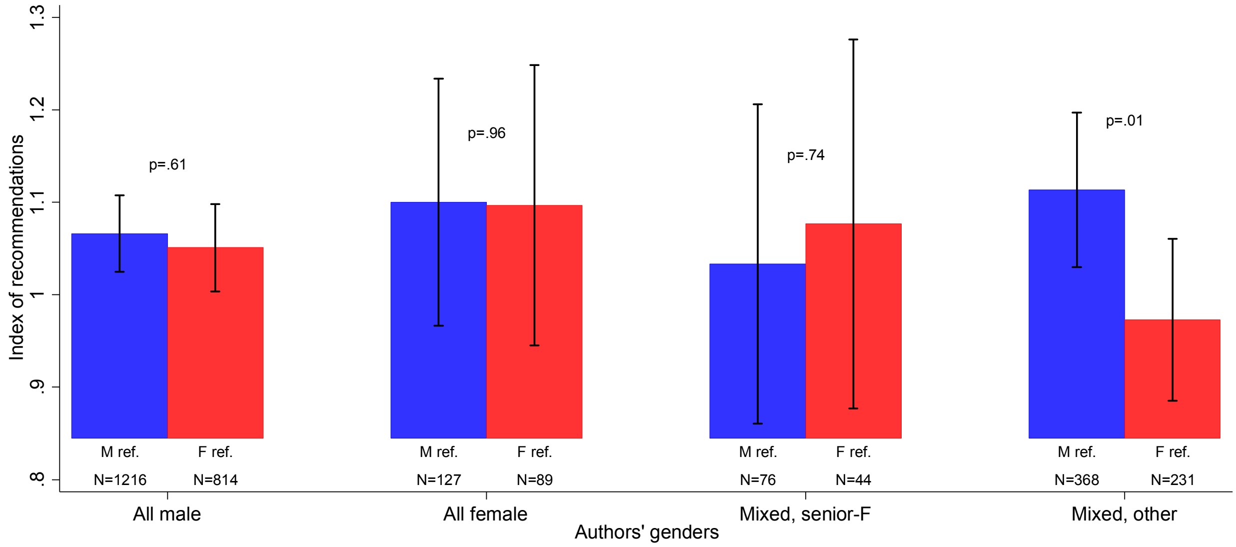
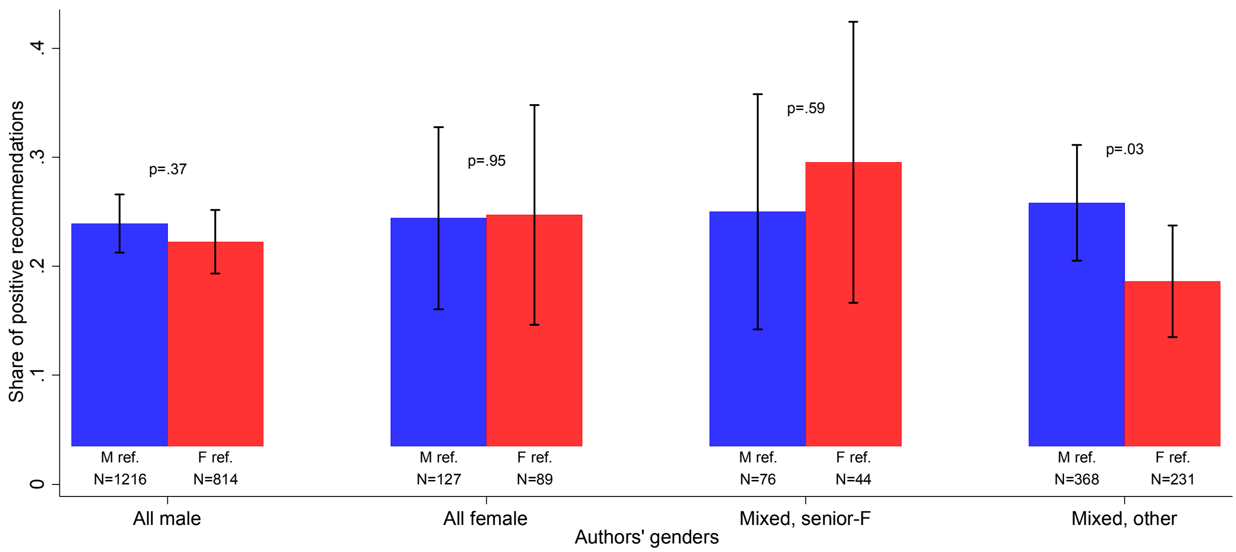


Figure 4b. Share of Positive Referee Recommendations



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 1,033 papers with both male and female referees. Figure 4b shows the share of positive recommendations, defined as RR-Accept.

Figure 5. Differences in Citations and R&R Rate, by Author Gender

Figure 5a. Referee Recommendations and Citations

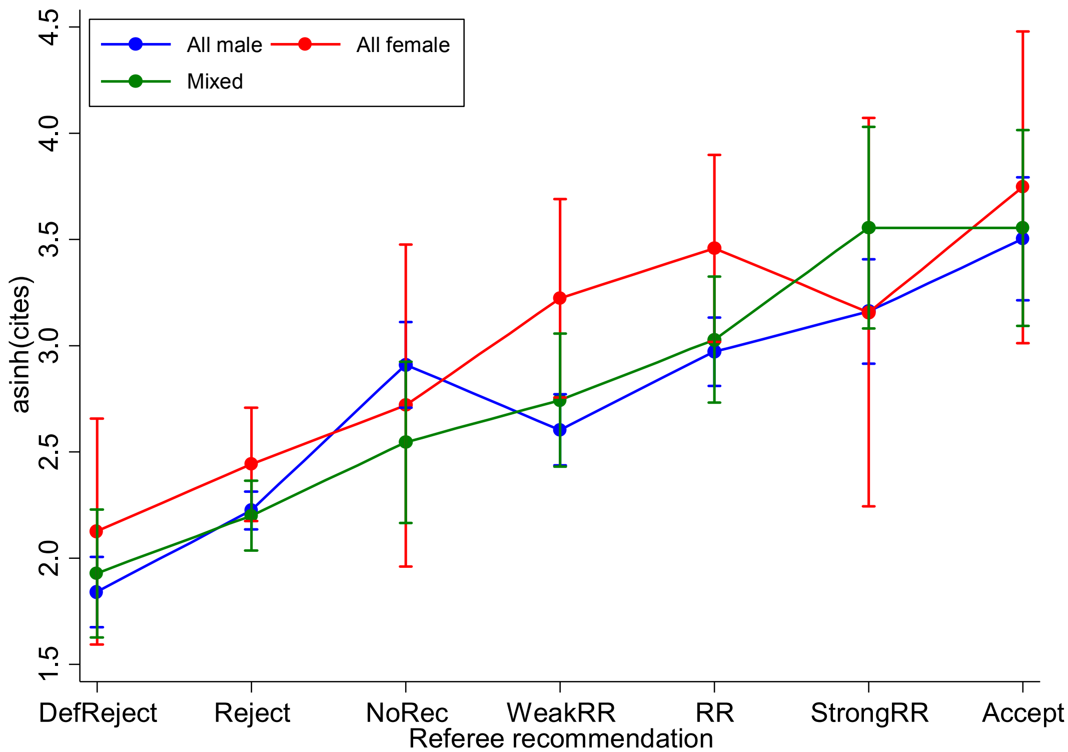


Figure 5b. Referee Recommendations and R&R Rate

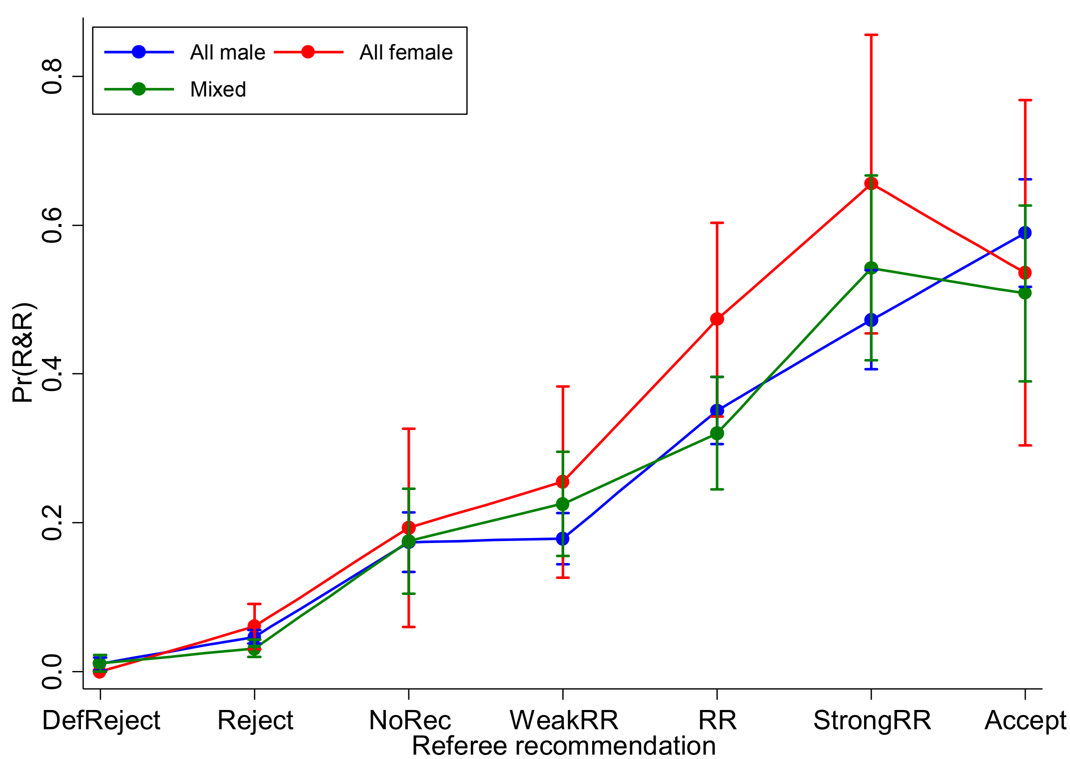
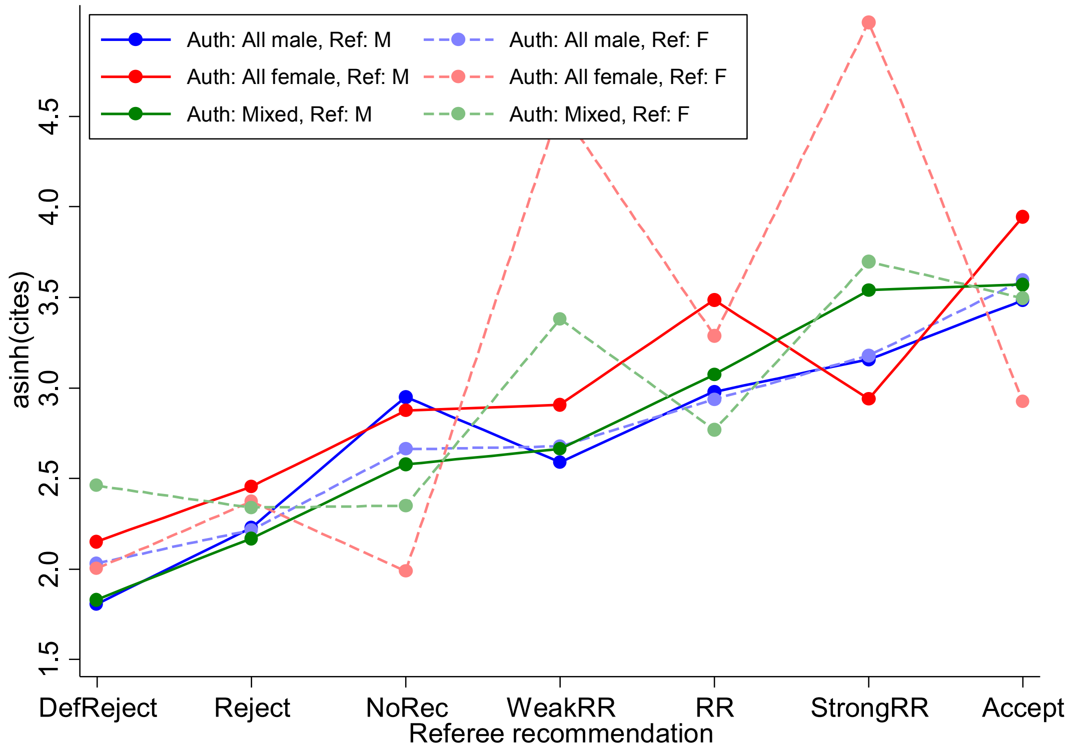


Figure 5c. Referee Recommendations and Citations, by Author Gender and Referee Gender



Notes: Figures 5a and 5c show the weighted *asinh* (citations) for a paper receiving a given recommendation, while Figure 5b shows the R&R rate for a paper receiving a given recommendation. Figures 5a and 5b show the results separately by author gender. Figure 5c 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 5c omits confidence intervals for legibility.

Figure 6. The Relationship Between the Editor's Decisions and Realized Citations

Figure 6a. R&R Papers versus Rejected papers, Split by Author Gender

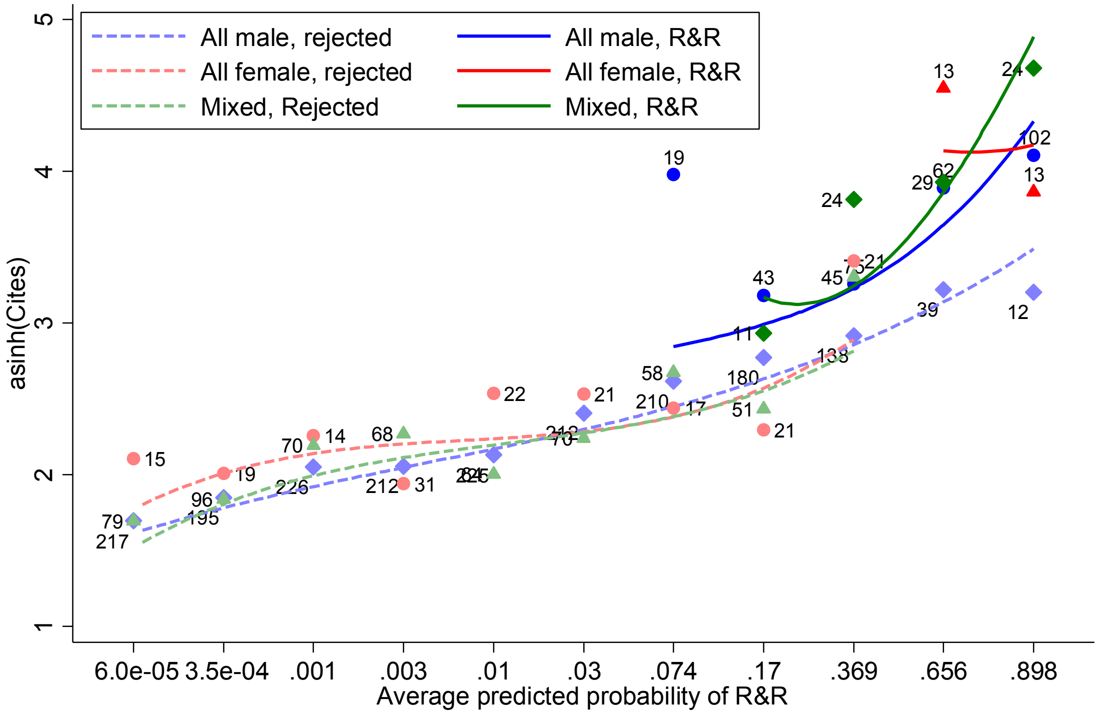
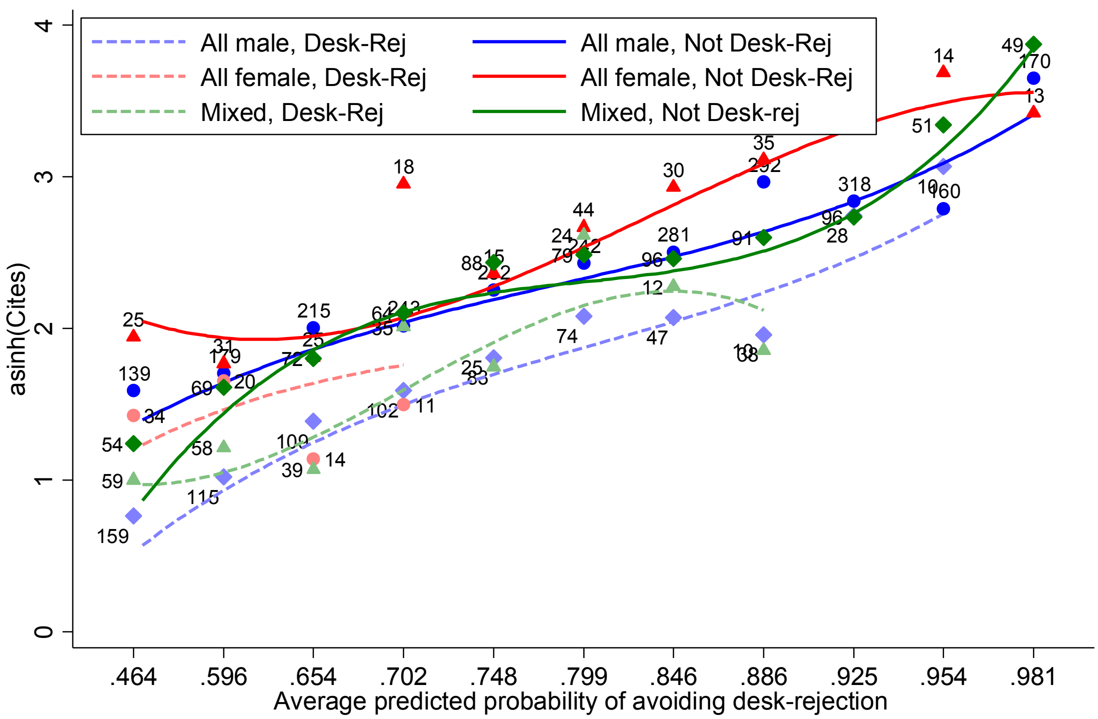


Figure 6b. Desk-Rejected Papers versus Rejected papers, Split by Author Gender



Notes: Figures 6a-b shows the average *asinh*(citations) by deciles of predicted probability of R&R where the top decile is further split into two ventiles, separately for papers that were rejected and those that the editor granted a revise-and-resubmit. Figure 6b does the same for the desk rejection decision. The smoothing lines are obtained via cubic fits to all data points.

Figure 7. Referee Informativeness, by Referee Gender

Figure 7a. Referee Recommendations and Citations

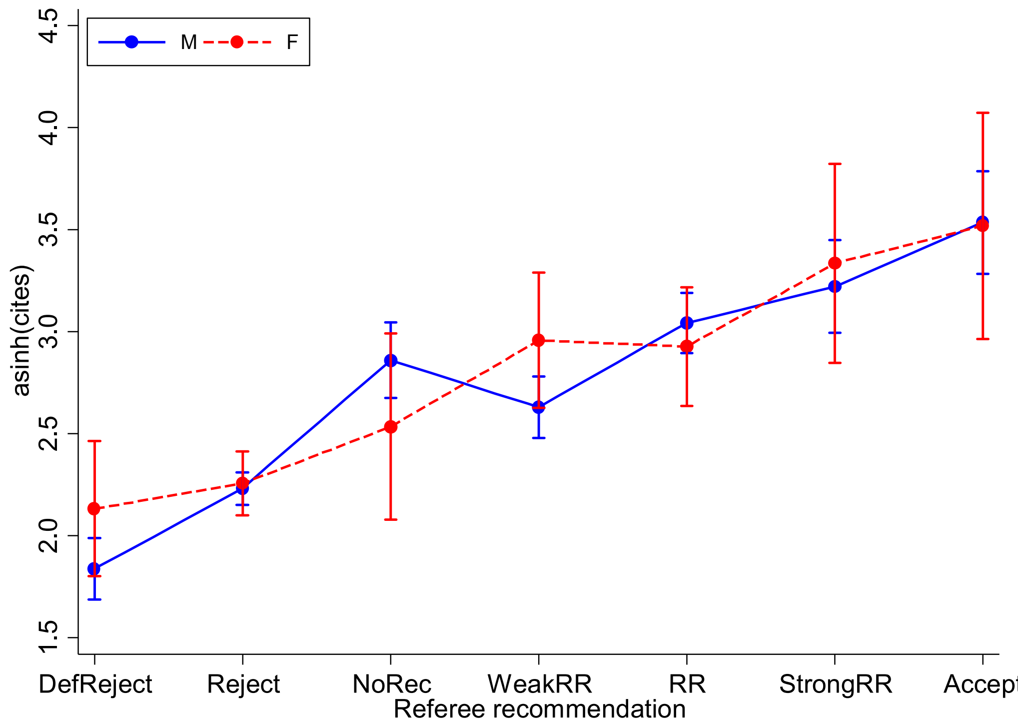
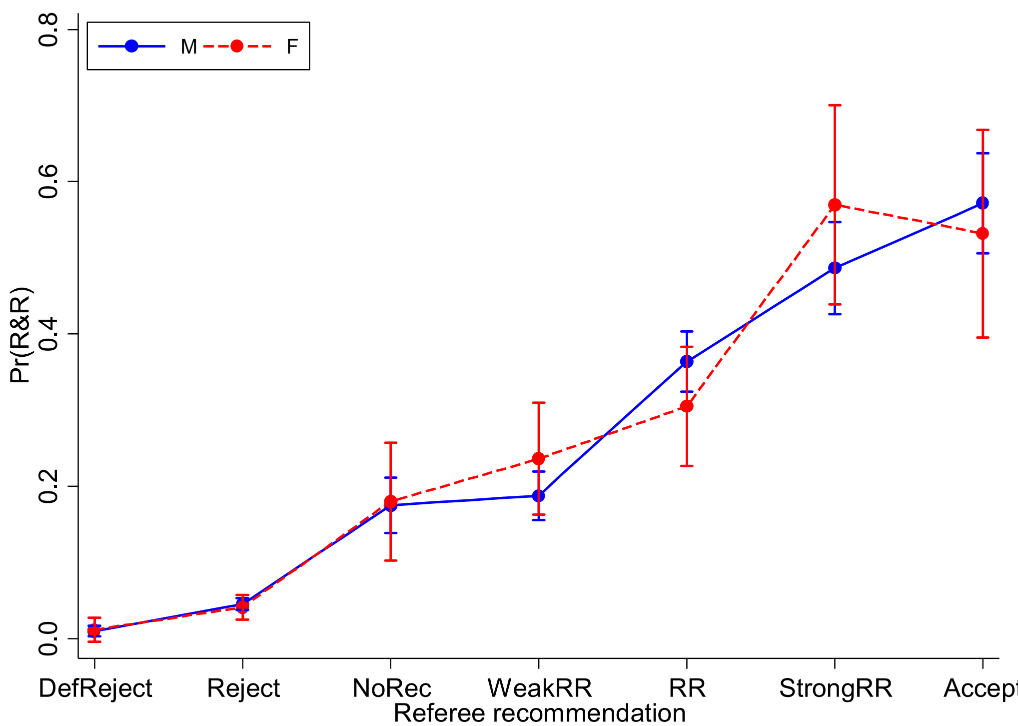


Figure 7b. Referee Recommendations and R&R Rate



Notes: Figure 7a shows the weighted *asinh* (citations) for a paper receiving a given recommendation. Figure 7b shows the R&R rate for a paper receiving a given recommendation. Both show the results separately by referee gender.

Figure 8. Other Editorial Outcomes: Referee and Editorial Delay

Figure 8a. Referee Response Time

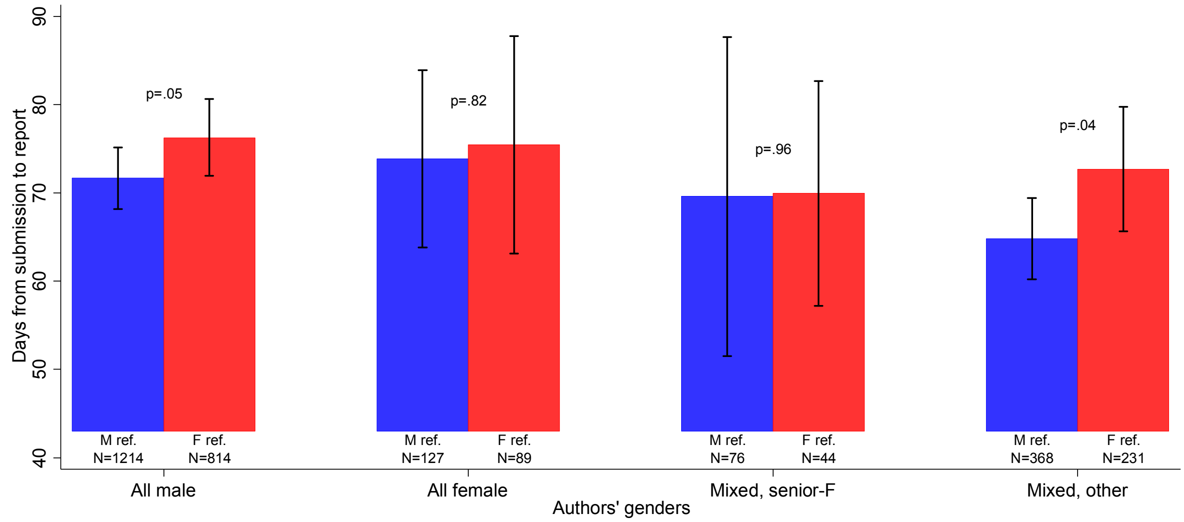


Figure 8b. Editor Response Time

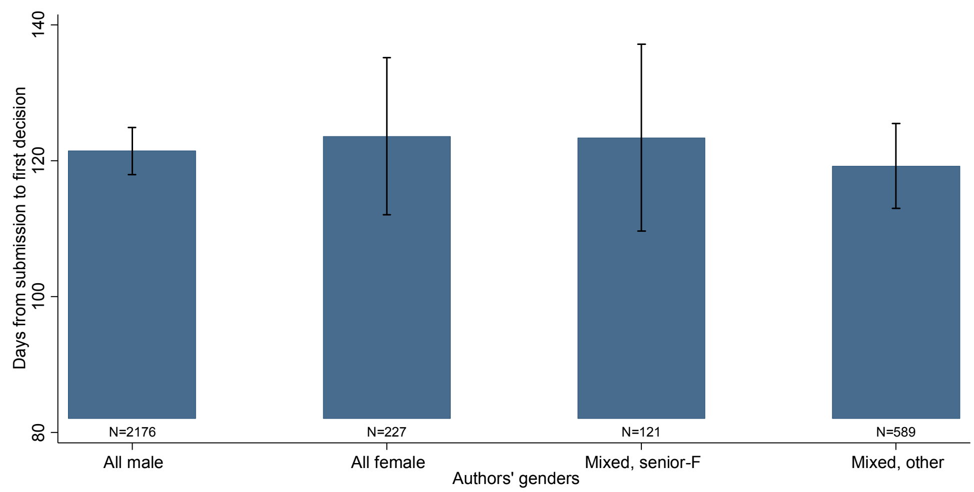
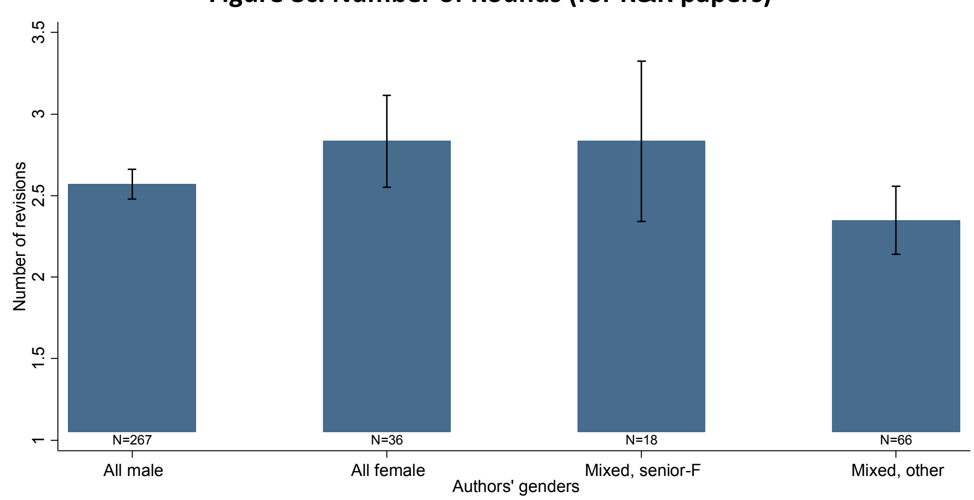


Figure 8c. Number of Rounds (for R&R papers)



Note: Figure 8a includes only papers with both male and female referees.

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	2.24	2.37	2.19	2.18	2.14	2.23	2.47	2.72	2.43	2.47	2.27	2.47
	(1.85)	(1.71)	(1.80)	(1.76)	(1.83)	(1.82)	(1.85)	(1.71)	(1.80)	(1.77)	(1.89)	(1.82)
Editorial Decisions												
Not Desk-Rejected	0.76	0.72	0.77	0.74	0.79	0.76	1.00	1.00	1.00	1.00	1.00	1.00
Received R&R Decision	0.11	0.13	0.11	0.10	0.10	0.11	0.14	0.18	0.17	0.13	0.12	0.14
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	0.66
All female	0.00	1.00	0.00	0.00	0.00	0.07	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	0.04	0.00	0.00	1.00	0.00	0.00	0.04
Mixed, other	0.00	0.00	0.00	1.00	0.00	0.18	0.00	0.00	0.00	1.00	0.00	0.18
Undetermined	0.00	0.00	0.00	0.00	1.00	0.05	0.00	0.00	0.00	0.00	1.00	0.05
Author Publications in 35 high-impact journals												
Publications: 0	0.44	0.47	0.49	0.46	0.49	0.45	0.38	0.39	0.46	0.39	0.42	0.39
Publications: 1	0.19	0.15	0.17	0.17	0.17	0.18	0.18	0.16	0.12	0.16	0.18	0.18
Publications: 2	0.12	0.12	0.09	0.09	0.08	0.11	0.13	0.11	0.11	0.12	0.08	0.12
Publications: 3	0.09	0.06	0.08	0.09	0.11	0.09	0.10	0.07	0.09	0.11	0.14	0.10
Publications: 4-5	0.08	0.10	0.08	0.10	0.06	0.09	0.10	0.14	0.11	0.12	0.08	0.10
Publications: 6+	0.08	0.10	0.09	0.09	0.09	0.08	0.10	0.13	0.11	0.11	0.10	0.10
Number of Authors												
1 author	0.34	0.31	0.35	0.34	0.34	0.34	0.32	0.30	0.31	0.29	0.34	0.31
2 authors	0.42	0.44	0.38	0.42	0.40	0.42	0.44	0.43	0.39	0.44	0.41	0.43
3 authors	0.20	0.21	0.24	0.20	0.22	0.20	0.20	0.25	0.26	0.23	0.21	0.21
4+ authors	0.04	0.03	0.03	0.04	0.04	0.04	0.04	0.03	0.03	0.04	0.04	0.04
Field of Paper												
Development	0.04	0.03	0.04	0.03	0.03	0.04	0.04	0.04	0.04	0.04	0.04	0.04
Econometrics	0.03	0.03	0.04	0.04	0.05	0.04	0.03	0.03	0.03	0.03	0.05	0.03
Finance	0.04	0.04	0.03	0.04	0.03	0.04	0.04	0.04	0.03	0.04	0.03	0.04
Health, Urban, Law	0.04	0.03	0.03	0.03	0.05	0.03	0.03	0.03	0.03	0.03	0.04	0.03
History	0.01	0.01	0.01	0.01	0.00	0.01	0.01	0.01	0.01	0.01	0.00	0.01
International	0.06	0.07	0.06	0.06	0.06	0.06	0.06	0.07	0.06	0.06	0.06	0.06
Industrial Organization	0.05	0.05	0.04	0.05	0.03	0.05	0.05	0.04	0.03	0.05	0.03	0.05
Lab/Experiments	0.03	0.03	0.04	0.03	0.02	0.03	0.03	0.04	0.05	0.02	0.02	0.03
Labor	0.11	0.12	0.10	0.11	0.12	0.11	0.11	0.14	0.10	0.12	0.11	0.12
Macro	0.10	0.10	0.10	0.09	0.07	0.10	0.10	0.12	0.12	0.10	0.09	0.10
Micro	0.10	0.09	0.11	0.10	0.12	0.10	0.10	0.08	0.10	0.11	0.11	0.10
Public	0.05	0.04	0.05	0.05	0.06	0.05	0.05	0.05	0.04	0.05	0.06	0.05
Theory	0.07	0.08	0.08	0.06	0.07	0.07	0.07	0.08	0.09	0.06	0.08	0.07
Unclassified	0.05	0.05	0.05	0.05	0.06	0.05	0.05	0.06	0.06	0.05	0.06	0.05
Missing Field	0.23	0.23	0.23	0.25	0.23	0.23	0.20	0.17	0.21	0.24	0.22	0.21
Referee Recommendations												
Fraction Definitely Reject							0.11	0.11	0.12	0.11	0.12	0.11
Fraction Reject							0.50	0.48	0.50	0.49	0.51	0.50
Fraction with No Rec'n							0.10	0.09	0.08	0.10	0.10	0.10
Fraction Weak R&R							0.11	0.11	0.11	0.11	0.11	0.11
Fraction R&R							0.11	0.12	0.12	0.10	0.10	0.11
Fraction Strong R&R							0.04	0.04	0.03	0.04	0.04	0.04
Fraction Accept							0.04	0.03	0.04	0.05	0.02	0.04
Referee Publications in 35 high-impact journals												
Share of referees with 3+ publications per paper							0.39	0.40	0.35	0.39	0.39	0.39
Referee genders (share per paper)												
Male							0.83	0.84	0.83	0.83	0.85	0.83
Female							0.16	0.15	0.16	0.16	0.14	0.16
Ambiguous							0.01	0.01	0.00	0.01	0.01	0.01
Years	2003-13	2003-13	2003-13	2003-13	2003-13	2003-13	2003-13	2003-13	2003-13	2003-13	2003-13	2003-13
Number of Observations	3,258	369	180	900	240	4,947	2,176	227	121	589	167	3,280

Notes: Table presents information on mean characteristics of all submitted papers (Columns 1-5), and for non-desk-rejected papers (Columns 6-10). The sample of non-desk-rejected papers also excludes papers with only 1 referee assigned. Author publications are based on publications in a set of 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. For example, a paper with JEL codes that match labor and theory will be coded 0.5 for labor and 0.5 for theory.

Table 2. Probability of Being Assigned a Female Referee, Impact of Author Team Gender

	Linear Probability Models for Female Referee	
	(1)	(2)
Authors' Genders (Omitted: All Male Authors)		
All Female Authors	0.01	0.01
	(0.01)	(0.01)
Mixed-Gender Author Team	0.02	0.02
female-led	(0.02)	(0.02)
Mixed-Gender Author Team	-0.01	-0.01
other	(0.01)	(0.01)
Undetermined Gender Team	-0.01	-0.01
	(0.02)	(0.02)
Controls for Author Publications	No	Yes
Controls for Referee Publications	No	Yes
Controls for No. of Authors	No	Yes
Controls for Field	No	Yes
Indicators for Year	Yes	Yes
R-squared	0.00	0.01
N	8,488	8,488

Notes: The sample for all models is referee reports for 3,280 papers with at least two referees assigned. The dependent variable is an indicator for the the referee being female. Standard errors clustered by paper in parentheses.

Table 3. Referee Recommendations, Impact of Author Team Gender

	Linear Probability Models for Receiving an R&R Recommendation or Better					OLS Models for Index of Referee Recommendations				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Authors' Genders (Omitted: All Male Authors)										
All Female Authors	-0.00	0.01	0.01	-0.00		-0.02	0.01	0.01	0.02	
	(0.02)	(0.02)	(0.02)	(0.05)		(0.03)	(0.03)	(0.04)	(0.07)	
Mixed-Gender Author Team	0.01	-0.03	-0.03	0.04		0.03	-0.05	-0.05	0.10	
senior author female	(0.02)	(0.02)	(0.03)	(0.05)		(0.04)	(0.04)	(0.04)	(0.08)	
Mixed-Gender Author Team	0.01	0.00	0.00	0.04		0.02	0.00	-0.00	0.03	
other	(0.01)	(0.01)	(0.02)	(0.03)		(0.02)	(0.02)	(0.03)	(0.05)	
Undetermined Gender Team	-0.06	-0.07	-0.08	-0.13		-0.07	-0.09	-0.11	-0.23	
	(0.02)	(0.02)	(0.02)	(0.04)		(0.04)	(0.04)	(0.04)	(0.07)	
Referee Gender (Omitted: Male Referee)										
Female Referee			-0.00	-0.01	0.01			0.00	-0.02	0.01
			(0.01)	(0.02)	(0.02)			(0.02)	(0.03)	(0.03)
Unknown-Gender Referee			0.06	0.21	0.07			-0.02	0.16	0.03
			(0.06)	(0.16)	(0.08)			(0.10)	(0.21)	(0.12)
Gender Interactions										
All Female Auth. X Female Ref.			0.00	-0.00	-0.00			0.02	0.02	0.01
			(0.05)	(0.07)	(0.06)			(0.08)	(0.10)	(0.10)
Mixed Auth. (F-senior) X Female Ref.			0.01	-0.07	-0.05			-0.02	-0.17	-0.12
			(0.05)	(0.07)	(0.07)			(0.09)	(0.11)	(0.11)
Mixed Auth. (other) X Female Ref.			0.01	-0.03	-0.02			0.02	-0.03	-0.01
			(0.03)	(0.04)	(0.04)			(0.05)	(0.06)	(0.06)
Undetermined Auth. X Female Ref.			0.11	0.16	0.18			0.14	0.25	0.22
			(0.07)	(0.06)	(0.06)			(0.12)	(0.10)	(0.11)
Only papers with both male and female referees	No	No	No	Yes	No	No	No	No	Yes	No
Paper Fixed Effects	No	No	No	No	Yes	No	No	No	No	Yes
Controls for Author Publications	No	Yes	Yes	Yes	-	No	Yes	Yes	Yes	-
Controls for No. of Authors and Field	No	Yes	Yes	Yes	-	No	Yes	Yes	Yes	-
Indicators for Year	Yes	Yes	Yes	Yes	-	Yes	Yes	Yes	Yes	-
Controls for Referee Publications	No	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes
R-squared	0.00	0.03	0.03	0.05	0.00	0.00	0.04	0.04	0.05	0.00
N	8,488	8,488	8,488	3,117	8,488	8,488	8,488	8,488	3,117	8,488

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.

Table 4. Citations and Editor Decision, Impact of Author Team Gender

	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	0.14	0.17	0.14	0.12	0.06	0.16	0.07
	(0.06)	(0.08)	(0.06)	(0.06)	(0.10)	(0.09)	(0.11)
Mixed-Gender Author Team	0.01	-0.16	-0.09	-0.12	0.41	0.07	0.37
senior author female	(0.13)	(0.14)	(0.13)	(0.13)	(0.27)	(0.15)	(0.31)
Mixed, other	0.01	0.04	0.01	-0.01	0.19	0.21	0.21
	(0.05)	(0.06)	(0.06)	(0.06)	(0.08)	(0.07)	(0.08)
Undetermined	-0.07	-0.17	-0.13	-0.12	0.08	-0.14	0.05
	(0.13)	(0.12)	(0.13)	(0.12)	(0.13)	(0.08)	(0.14)
Fractions of Referee Recommendations							
Reject	0.50		0.39	0.39	1.48		1.37
	(0.11)		(0.10)	(0.10)	(0.48)		(0.54)
No Recommendation	0.85		0.67	0.51	3.41		3.28
	(0.16)		(0.15)	(0.13)	(0.49)		(0.62)
Weak R&R	1.41		1.15	0.95	3.81		3.66
	(0.19)		(0.15)	(0.14)	(0.55)		(0.62)
R&R	2.00		1.67	1.18	5.22		5.07
	(0.19)		(0.15)	(0.15)	(0.67)		(0.72)
Strong R&R	2.18		1.91	1.14	6.34		6.33
	(0.16)		(0.15)	(0.32)	(0.67)		(0.72)
Accept	2.77		2.38	1.56	6.55		6.46
	(0.32)		(0.32)	(0.52)	(0.81)		(0.83)
Author Publications in 35 High-Impact Journals							
1 Publication		0.35	0.24	0.25		0.19	-0.05
		(0.09)	(0.09)	(0.09)		(0.10)	(0.12)
2 Publications		0.64	0.48	0.45		0.50	0.37
		(0.08)	(0.08)	(0.08)		(0.11)	(0.16)
3 Publications		0.69	0.48	0.46		0.55	0.18
		(0.10)	(0.10)	(0.10)		(0.05)	(0.06)
4-5 Publications		0.83	0.63	0.57		0.72	0.54
		(0.09)	(0.07)	(0.07)		(0.11)	(0.14)
6+ Publications		1.30	0.95	0.84		1.18	0.78
		(0.16)	(0.14)	(0.13)		(0.11)	(0.09)
R&R Indicator				-0.17			
(Mechanical Publ. Effect)				(0.17)			
Control Function for Selection				0.84			
(Value Added of the Editor)				(0.28)			
Editor Leave-out-Mean R&R							1.48
Rate							(1.86)
Controls for No. of Authors	No	Yes	Yes	Yes	No	Yes	Yes
Controls for Field	No	Yes	Yes	Yes	No	Yes	Yes
Indicators for Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	3,280	3,280	3,280	3,280	3,280	3,280	3,280
R^2 / pseudo R^2	0.17	0.16	0.22	0.23	0.48	0.09	0.51

Notes: The sample for all models is 3,280 non-desk-rejected papers with at least two referees assigned. Dependent variable for OLS models in Columns 1-4 is asinh of Google Scholar citations. Dependent variable in probit models in Columns 5-7 is indicator for receiving revise and resubmit decision. 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 5. Citations and Editor Decision, Heterogeneity Analysis

	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	0.09	0.34	0.30	0.02	0.41	0.20
	(0.06)	(0.16)	(0.17)	(0.15)	(0.35)	(0.22)
Mixed-Gender	-0.03	0.05	-0.14	0.25	-0.04	-0.14
	(0.07)	(0.24)	(0.09)	(0.10)	(0.27)	(0.09)
Undetermined	-0.12	-0.31	-0.13	0.05	-0.05	0.04
	(0.12)	(0.15)	(0.11)	(0.14)	(0.13)	(0.13)
Authors' Genders and Publications						
All Female *	0.46			0.50		
(Max Publication >=3)	(0.32)			(0.26)		
Mixed-Gender *	0.21			-0.03		
(Female pub 3+, Male Pub<3)	(0.37)			(0.31)		
Mixed-Gender *	-0.05			-0.04		
(Female pub <3, Male Pub 3+)	(0.15)			(0.14)		
Mixed-Gender *	0.19			-0.13		
(Female pub 3+, Male Pub 3+)	(0.31)			(0.37)		
Authors' Genders and Field						
All Female *		-1.40			-1.89	
Share females in Field		(1.03)			(1.91)	
Mixed-Gender *		0.06			1.41	
Share females in Field		(1.38)			(1.06)	
Authors' Genders and Year of Submission						
All Female *			-0.23			0.04
(Years of Submission 2010 on)			(0.23)			(0.25)
Mixed-Gender *			0.24			0.09
(Years of Submission 2010 on)			(0.10)			(0.13)
R&R Indicator	0.85	0.76	0.83			
(Mechanical Publ. Effect)	(0.28)	(0.25)	(0.29)			
Control Function for Selection	-0.18	-0.09	-0.17			
(Value Added of the Editor)	(0.16)	(0.16)	(0.16)			
Editor Leave-out-Mean R&R				1.38	-0.79	1.52
Rate				(1.87)	(1.82)	(1.84)
Controls for No. of Authors	Yes	Yes	Yes	Yes	Yes	Yes
Controls for Field	Yes	Yes	Yes	Yes	Yes	Yes
Indicators for Year	Yes	Yes	Yes	Yes	Yes	Yes
N	3,280	2,592	3,280	3,280	2,592	3,280
R^2 / pseudo R^2	0.23	0.24	0.23	0.51	0.52	0.51

Notes: The sample for all models in Columns 1, 3, 4, and 6 is 3,280 non-desk-rejected papers with at least two referees assigned. The sample for Columns 2 and 4 is further restricted to 2,592 papers with fields recorded. 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.

Table 6. Desk Rejection, Impact of Author Team Gender

Specification: / **Dependent Variable:**	**OLS Regression** / **Asinh of Citations**	**Probit** / **Indicator for Paper Not Desk Rejected**
	(1)	(2)
Authors' Genders		
All Female	0.15	-0.11
	(0.06)	(0.09)
Mixed-Gender Author Team	-0.04	-0.10
senior author female	(0.12)	(0.08)
Mixed, other	0.01	-0.07
	(0.05)	(0.06)
Undetermined	-0.06	-0.24
	(0.08)	(0.08)
Author Publications in 35 high-impact journals		
Publications: 1	0.42	0.21
	(0.11)	(0.09)
Publications: 2	0.69	0.50
	(0.15)	(0.08)
Publications: 3	0.75	0.61
	(0.16)	(0.09)
Publications: 4-5	0.85	0.86
	(0.19)	(0.13)
Publications: 6+	1.38	1.15
	(0.29)	(0.17)
NDR Indicator	0.28	
(Mechanical Publ. Effect)	(0.66)	
Control Function for Selection into NDR	0.13	
(Value Added of the Editor)	(0.38)	
Editor Leave-out-Mean NDR		3.47
Rate		(0.84)
Controls for No. of Authors	Yes	Yes
Controls for Field	Yes	Yes
Indicators for Year	Yes	Yes
Number of Observations	4,947	4,947
R^2 / pseudo R^2	0.21	0.12

Notes: The sample for all models is 4,947 papers. Dependent variable for OLS model in Column 1 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. Standard errors clustered by editor in parentheses.

Table 7. 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		-0.003	-0.201		0.301	0.162
		(0.301)	(0.312)		(0.158)	(0.157)
Gender Level Controls						
Female Referee		0.201	-0.186		-0.182	-0.142
		(0.212)	(0.194)		(0.518)	(0.505)
All Female Authors	0.308	0.089	0.006	0.117	0.123	0.054
	(0.094)	(0.095)	(0.096)	(0.128)	(0.143)	(0.137)
Mixed-Gender Author Team	-0.180	-0.107	-0.084	0.077	0.197	0.182
female-led	(0.234)	(0.244)	(0.246)	(0.183)	(0.211)	(0.214)
Mixed-Gender Author Team	0.232	-0.104	-0.231	-0.153	-0.143	-0.023
other	(0.063)	(0.068)	(0.071)	(0.130)	(0.129)	(0.138)
Other Slope Variables						
Referee Publications 3+		-0.093	-0.113		0.217	0.189
		(0.208)	(0.190)		(0.125)	(0.112)
Asinh (No. Reports for Editor)			0.205			0.127
			(0.062)			(0.048)
Journal Fixed Effect	-	-	-	-	-	-
Field Fixed Effect	No	Yes	Yes	No	Yes	Yes
Level Additional Controls						
Share Referees with 3+ Pubs.		0.190	0.212		-0.541	-0.312
		(0.141)	(0.136)		(0.525)	(0.465)
Mean Asinh (No. Reports for Editor)			-0.175			-0.268
			(0.034)			(0.165)
Fractions of Referee Recommendations (Other Fractions Included, not Reported)						
R&R	1.179	2.161	2.131	5.112	4.821	4.476
	(0.137)	(0.396)	(0.413)	(0.765)	(1.129)	(1.096)
Author Publications (Other Indicators Included, not Reported)						
6+ Publications	0.901	0.912	0.887	0.812	0.793	0.789
	(0.113)	(0.103)	(0.105)	(0.104)	(0.119)	(0.122)
R&R Indicator	0.820	0.830	0.878			
(Mechanical Publ. Effect)	(0.278)	(0.242)	(0.245)			
Control Function for Selection	-0.198	-0.123	-0.126			
(Value Added of the Editor)	(0.159)	(0.131)	(0.125)			
Editor Leave-out-Mean R&R				0.923	0.876	0.721
Rate				(1.477)	(1.447)	(1.522)

Table 8. Referee Decision Time, Impact of Author and Referee Gender

	Number of Days from Paper Submission to Referee Report			
	(1)	(2)	(3)	(4)
Authors' Genders (Omitted: All Male Authors)				
All Female Authors	-1.94	-2.19	-2.14	
	(2.17)	(2.16)	(2.16)	
Mixed-Gender Author Team	-1.74	-0.87	-0.95	
female-led	(3.33)	(3.20)	(3.20)	
Mixed-Gender Author Team	-0.27	0.32	0.28	
other	(1.74)	(1.69)	(1.69)	
Undetermined Gender Team	-0.95	0.32	0.23	
	(3.15)	(3.13)	(3.14)	
Referee Gender (Omitted: Male Referee)				
Female Referee			2.10	1.94
			(1.88)	(1.74)
Unknown-Gender Referee			-3.11	1.89
			(4.63)	(5.46)
Gender Interactions				
All Female Auth. X Female Ref.			0.95	2.01
			(5.76)	(5.27)
Mixed Auth. (senior-F) X Female Ref.			0.12	-0.15
			(5.99)	(4.77)
Mixed Auth. (other) X Female Ref.			-1.61	-0.63
			(3.39)	(3.57)
Undetermined Auth. X Female Ref.			-6.10	0.21
			(6.89)	(5.32)
Paper Fixed Effects	No	No	No	Yes
Controls for Summary Evaluations	No	Yes	Yes	Yes
Controls for Author Publications	No	Yes	Yes	Yes
Controls for Referee Publications	No	Yes	Yes	Yes
Controls for No. of Authors	No	Yes	Yes	Yes
Controls for Field	No	Yes	Yes	Yes
Indicators for Year	Yes	Yes	Yes	Yes
R-squared	0.15	0.17	0.17	0.01
N	8,481	8,481	8,481	8,481

Notes: Report time is calculated as the number of days from paper submission to referee report submission, rounded to the nearest 10.

Table 9. Decision Time and Duration of Revisions, Impact of Author Team Gender

	Number of Days: Sub. To Last Report	Number of Days: Reports to Editor	Number of Days: Sub. To Editor	No. of Rounds (for R&Rs)	Days Before Resub. (R&Rs)	Days from Resub. to Accept
	(1)	(2)	(3)	(4)	(5)	(6)
Authors' Genders						
All Female	0.38	2.19	1.29	0.17	-0.96	30.41
	(2.79)	(4.92)	(3.97)	(0.12)	(30.41)	(34.68)
Mixed-Gender Author Team	-2.09	5.34	3.70	0.14	71.82	-2.66
senior author female	(2.10)	(5.80)	(5.05)	(0.12)	(51.53)	(31.60)
Mixed, other	-1.85	1.10	-0.93	-0.10	17.77	5.65
	(1.40)	(3.11)	(3.29)	(0.08)	(29.42)	(12.28)
Undetermined	-2.46	-2.55	-4.17	-0.00	131.41	16.50
	(3.30)	(3.28)	(3.63)	(0.24)	(90.39)	(42.16)
Fractions of Referee Recommendations						
Reject	15.50	7.12	24.18	0.59	140.81	43.78
	(4.01)	(5.67)	(8.44)	(0.35)	(99.57)	(169.55)
No Recommendation	29.96	15.19	46.18	0.50	61.73	43.52
	(9.67)	(9.13)	(11.18)	(0.37)	(101.70)	(187.18)
Weak R&R	35.17	32.11	69.26	0.57	142.00	77.85
	(4.91)	(11.33)	(13.46)	(0.34)	(104.01)	(195.60)
R&R	51.32	31.48	84.31	0.74	208.43	96.22
	(6.20)	(11.89)	(11.11)	(0.31)	(120.93)	(183.69)
Strong R&R	54.95	22.37	80.78	0.65	90.57	56.15
	(11.21)	(16.33)	(24.54)	(0.42)	(75.10)	(204.89)
Accept	44.28	9.99	56.29	0.49	108.80	-43.80
	(9.84)	(14.74)	(19.63)	(0.45)	(122.65)	(185.65)
R&R Indicator			13.18			
(Mechanical Publ. Effect)			(5.80)			
Sample		All Papers			R&R Papers Only	
Controls for Author Publications	Yes	Yes	Yes	Yes	Yes	Yes
Controls for No. of Authors	Yes	Yes	Yes	Yes	Yes	Yes
Controls for Field	Yes	Yes	Yes	Yes	Yes	Yes
Indicators for Year	Yes	Yes	Yes	Yes	Yes	Yes
Editor Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Mean of the Dependent Variable:	94.5	31.4	120.9	2.6	209.1	178.9
N	3,280	2,892	3,280	405	384	389
R-squared	0.11	0.15	0.23	0.10	0.25	0.25

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.

Table 10. 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	0.30	0.18	1.13	0.53
	(0.22)	(0.16)	(0.77)	(0.38)
Mixed-Gender Author Team	-0.03	0.06	-1.16	-0.85
senior author female	(0.29)	(0.21)	(0.63)	(0.60)
Mixed, other	0.05	0.13	-0.48	-0.04
	(0.15)	(0.10)	(0.40)	(0.27)
Undetermined	0.39	0.20	-0.93	-1.03
	(0.25)	(0.17)	(0.79)	(0.55)
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 Year	Yes	Yes	Yes	Yes
Mean of the Dependent Variable:	19.1	15.2	19.1	15.2
N	2,810	2,810	455	455
R-squared	0.03	0.02	0.08	0.07

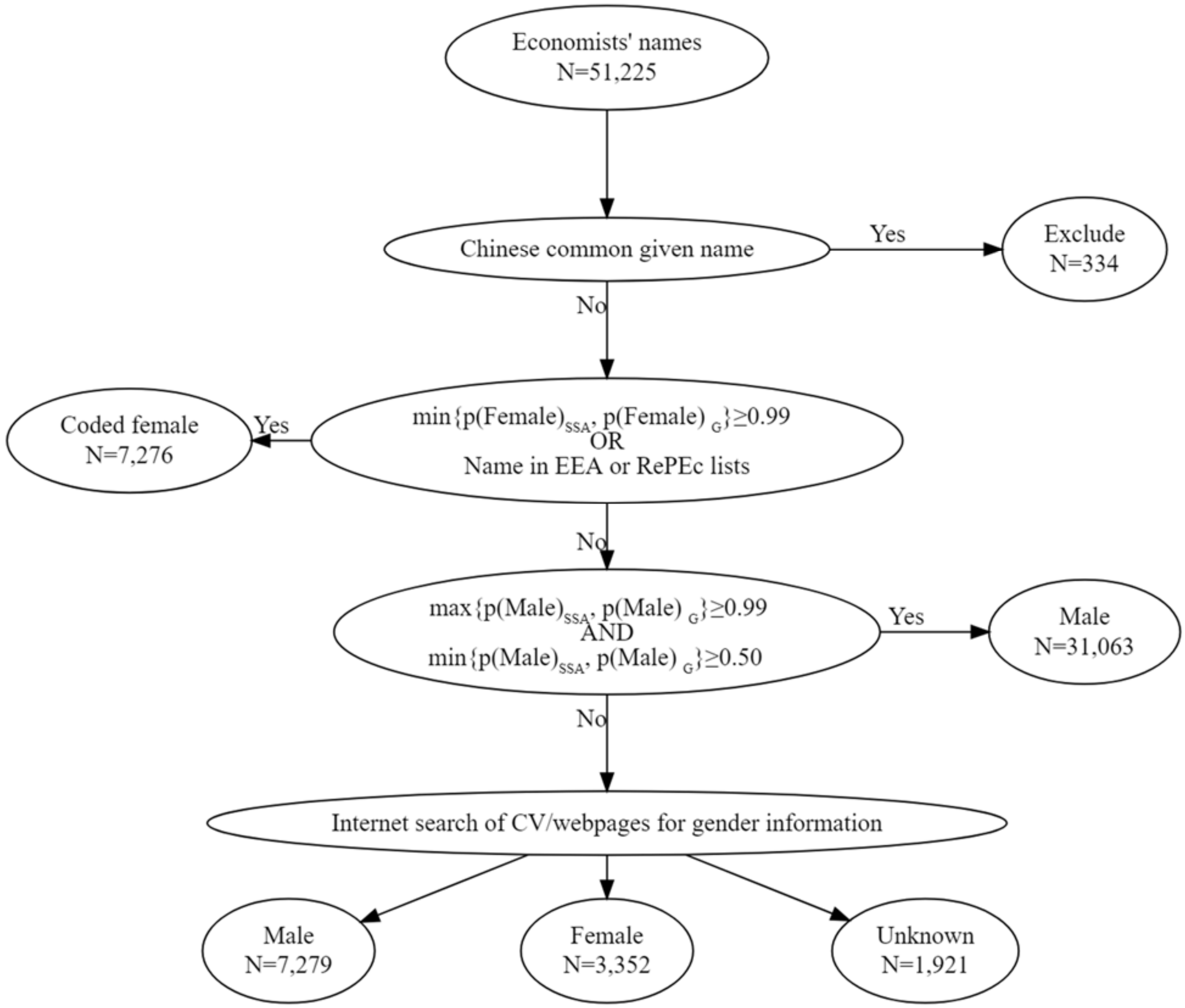
Notes: Dependent variables are measures of reading complexity. The Gunning fog index is as 0.4[(words/sentences) + 100(complex words/words)], where complex words are tri-syllabic words, excluding common suffixes and proper nouns. The Coleman-Liau index is calculated as 0.0588(letters/words) - 0.296(sentences/words)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

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

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	

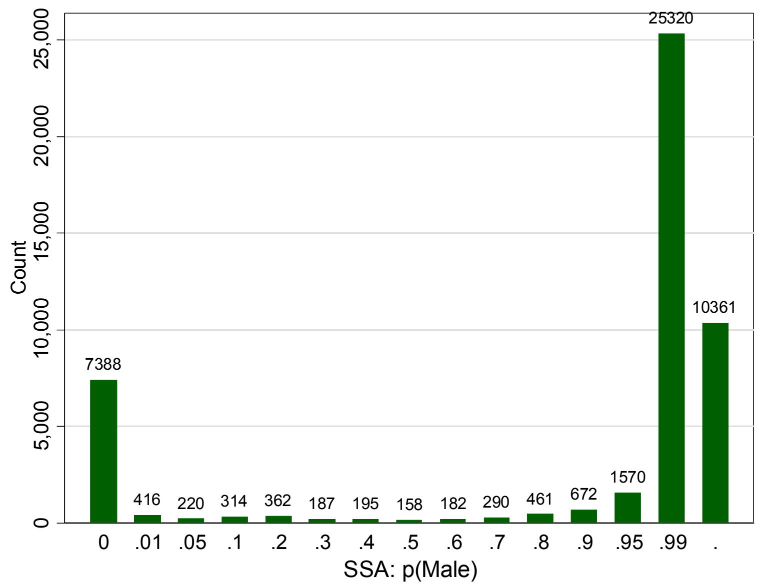
List of Additional Journals Used to Generate List of Authors Coded for Gender	
Economic Theory	Journal of Economics and Management Strategy
European Economic Review	Labour Economics
Quantitative Economics	Public Choice
Theoretical Economics	European Journal of Political Economy
Review of Economic Dynamics	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 Management	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

Online Appendix Figure 1a. Coding Gender for Names



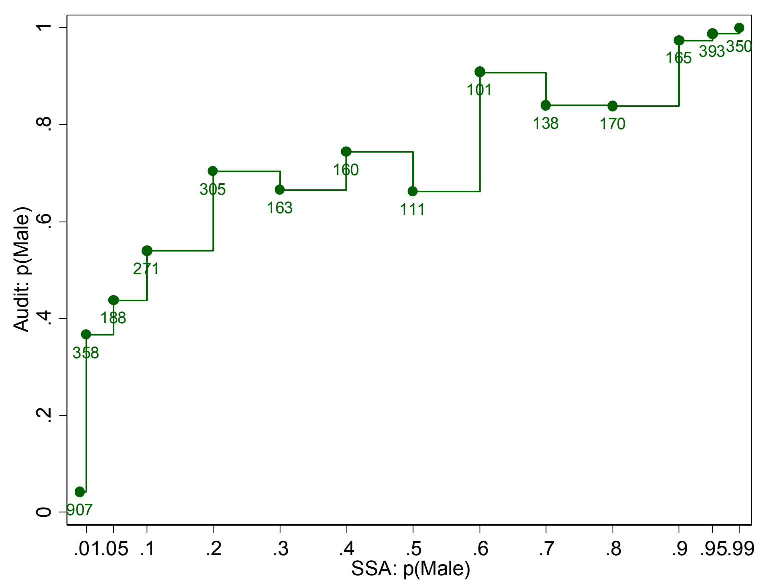
Note: Graph shows the process by which gender is assigned to names.

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



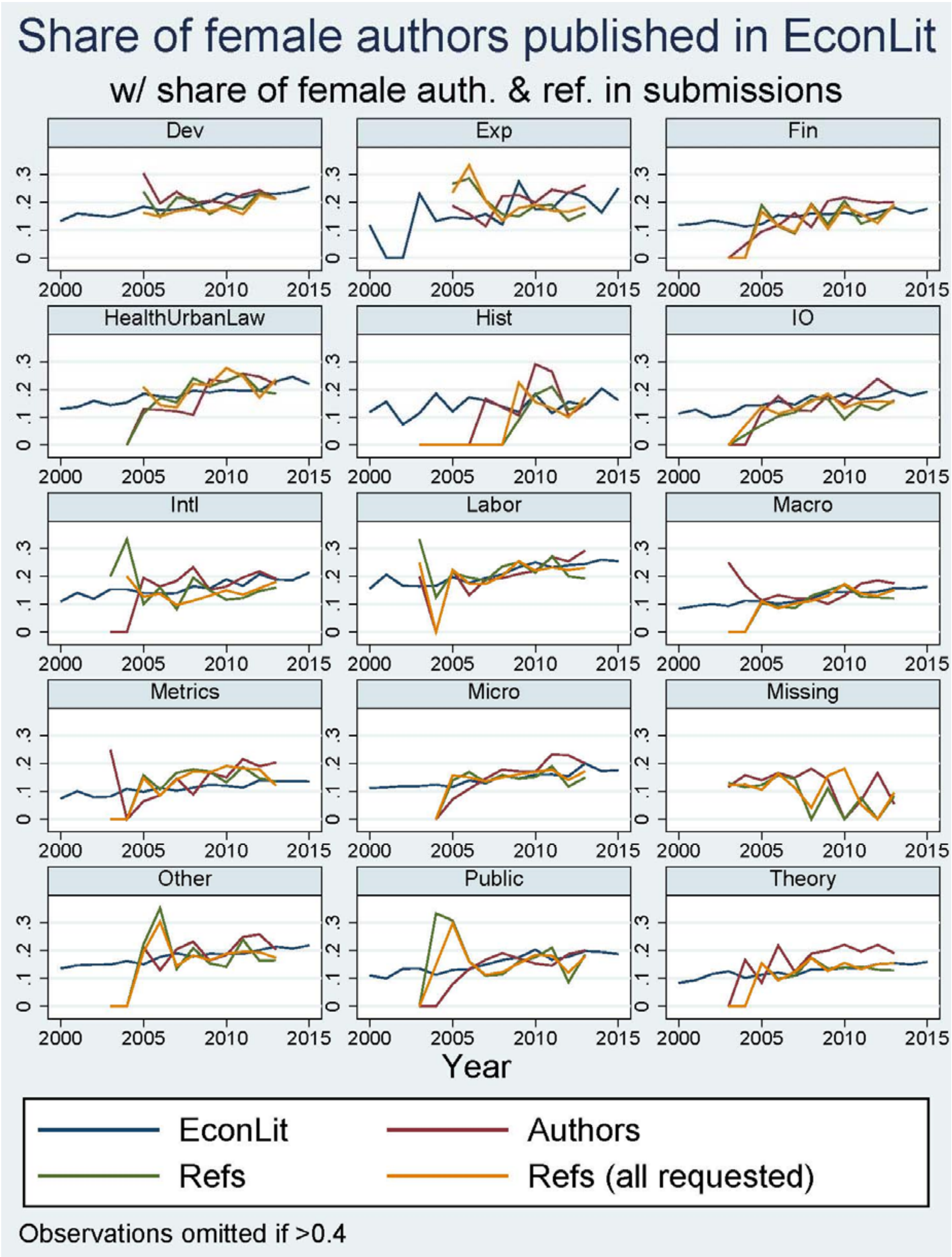
Note: Each observation is an author in a dataset of all papers published in 63 journals from 1991 to 2017 from Econlit. For each author, we code the probability that the author is male based on the first name, using an R routine that is based on the SSA data set of names. The graph indicates the p(male) 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



Intro

.

This survey is being conducted by David Card, Stefano DellaVigna, Patricia Funk, and Nagore Iriberri. In this survey, we ask about the impact of gender (of authors and referees) in the editorial process. We will analyze evidence in this regard for four journals (see below), and the survey responses will allow us to relate our findings with the expectations of economists.

For reference, we are analyzing submissions to four journals (The Quarterly Journal of Economics, The Review of Economic Studies, The Journal of the European Economic Association and The Review of Economics and Statistics) between the date the journal first started using Editorial Express (typically 2006) and 2013, and measuring citations with Google Scholar citations to those submissions in mid 2015. This is the same sample as in Card and DellaVigna (2018), supplemented with gender information. For the purpose of this survey, we are considering only papers that are not desk-rejected.

If you agree to participate in this research, we will ask you to complete the following survey. Otherwise, you may close this page to exit. The survey consists of 14 quantitative and qualitative questions and should take about 5-10 minutes to complete.

There is no direct benefit or compensation to you from taking part in this study, but it is hoped that the research will enhance our understanding of the impact of gender in the editorial process.

As with all research, there is a chance that confidentiality could be compromised; however, we are taking precautions to minimize this risk. The survey is anonymous. We do not identify any individual respondents or collect individual identifiers. Furthermore, the anonymous responses will be securely stored on Qualtrics servers and will only be accessed by the researchers and their assistants. After the research is complete, the responses will be stored as described above for potential future research or replication.

Participation in research is completely voluntary. You are free to decline to take part in the project. You can decline to answer any questions and are free to stop taking part in the project at any time.

If you consent to participating in this survey, which is approved by the UC Berkeley IRB as protocol 2018-04-10955, please save a copy of this page for future reference and continue by clicking the arrow below.

Thank you so much!
-David Card, Stefano DellaVigna, Patricia Funk, and Nagore Iriberri

---

Referee assignment

.
In order to simplify, please assume that either all the authors of a given paper are female ("female authored") or all the authors are male ("male authored").

Q1. In your experience, for two papers in the same field (e.g., macro, labor, etc.), are female-authored papers more likely to be assigned to female referees than male-authored papers?

More likely

About the same

Less likely

. If you answered more or less likely, in your opinion, which might be the most important potential reasons for the higher or lower share of female referees for female-authored papers?

Author gender

.

Again, in order to simplify, please assume that either all the authors of a given paper are female ("female authored") or all the authors are male ("male authored").

Q2. Compare a female-authored paper to a male-authored paper in a given field. Holding constant how a referee perceives the quality of a paper, do you think the average referee is likely to give a more positive, about equal, or less positive referee recommendation to the female-authored paper?

More positive to the female-authored paper

About the same

Less positive to the female-authored paper

Q3. Consider the referee recommendations for a **female-authored paper** which has at least one male and at least one female referee. On average across all papers, 20 percent of referee recommendations are positive (that is, R&R, Strong R&R, or Accept).

What percent of recommendations by female referees are positive?

What percent of recommendations by male referees are positive?

Q4. Consider now the referee recommendations on a **male-authored paper** which has at least one male and at least one female referee. On average across all papers, 20 percent of referee recommendations are positive (that is, R&R, Strong R&R, or Accept).

What percent of recommendations by female

referees are positive?
What percent of recommendations by male referees are positive?

Q5. Holding constant the prior publication record of the author(s), the field of the paper, and also the *referee recommendations*, do you think a female-authored paper has a higher, lower, or the same probability of receiving a revise and resubmit (R&R) decision?

Higher

About the same

Lower

Q6. Do you have any thoughts you would like to contribute on how referees and editors handle papers by female authors versus papers by male authors?

Citations

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?

More

About the same

Fewer

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.

Answer in log points

.

If you answered more or fewer in Q7: what do you think are the potential explanations?

Referee gender

.

In this section we are interested in the degree to which referee recommendations are predictive of future citations. We measure the informativeness of referee recommendations with the extent to which recommendations predict (later) citations for a paper. (We measure citations with asinh of citations). When we ask you to compare the informativeness of male and female referees, please assume that we hold constant the publication record of the two referees.

Q9.
For a given paper, do you think that, on average, a positive recommendation from a female referee is more informative about future citations, equally informative, or less informative than a positive recommendation from a male referee?

More informative

About the same

Less informative

Q10. 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

About the same

Less likely

.

Now we ask the same questions, but about female-authored papers.

Q11. For a **female-authored paper**, do you think that, on average, a positive recommendation from a female referee is more informative about future citations, equally informative, or less informative than a positive recommendation from a male referee?

More informative

About the same

Less informative

Q12.

For a **female-authored paper**, do you think that, on average, an editor is more or less likely to follow the recommendation of a female (relative to a male) referee in the R&R decision?

More likely

About the same

Less likely

Gender composition of author teams

Q13. So far, we considered only teams of all-male authors versus teams of all-female authors. Consider now an author team which includes 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

The patterns for all-male authors

The patterns for all-female authors

About halfway

It depends

Q14. Consider now an author team which includes both males and females, and the author with the most prior publications is **male**. Would you say that the patterns, in terms of the previous questions, would be more similar to

The patterns for all-male authors

The patterns for all-female authors

About halfway

It depends

. If you answered "it depends" in Q13 or Q14, can you tell us briefly how?

Validation Comments

. How many years ago did you obtain your PhD?

. What is your approximate field within economics?

. Thank you for taking the time to respond to this survey. If you have any additional comments, feel free to enter them in the box below: