

NBER WORKING PAPER SERIES

ARE REFEREES AND EDITORS IN ECONOMICS GENDER NEUTRAL?

David Card
Stefano DellaVigna
Patricia Funk
Nagore Iriberry

Working Paper 25967
<http://www.nber.org/papers/w25967>

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
June 2019

We thank the editors and staff of the Journal of the European Economics Association, the Quarterly Journal of Economics, the Review of Economics and Statistics, and the Review of Economic Studies for their invaluable assistance and support. We also thank Dan Hamermesh, Lawrence Katz, Andrei Shleifer, Lise Versterlund, and Heidi Williams for helpful comments, and Manu Garcia for sharing code. We are also grateful to Luisa Cefala, Margaret Chen, Alden Cheng, Manu Garcia, Johannes Hermle, Giovanni Kraushaar, Christopher Lim, Andrew Tai, and a team of undergraduate research assistants for their extraordinary help. Nagore Iriberry acknowledges financial support from grants ECO2015-66027-P MINECO/FEDER and IT869-13. Patricia Funk acknowledges financial support from the Swiss National Science Foundation (grant 178887). The analysis plan is posted on the AEA registry under AEARCTR-0003048. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2019 by David Card, Stefano DellaVigna, Patricia Funk, and Nagore Iriberry. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Are Referees and Editors in Economics Gender Neutral?
David Card, Stefano DellaVigna, Patricia Funk, and Nagore Iriberry
NBER Working Paper No. 25967
June 2019
JEL No. J16

ABSTRACT

We study the role of gender in the evaluation of economic research using submissions to four leading journals. We find that referee gender has no effect on the relative assessment of female-versus male-authored papers, suggesting that any differential biases of male referees are negligible. To determine whether referees as a whole impose different standards for female authors, we compare citations for female and male-authored papers, holding constant referee evaluations and other characteristics. We find that female-authored papers receive about 25% more citations than observably similar male-authored papers. Editors largely follow the referees, resulting in a 6 percentage point lower probability of a revise and resubmit verdict for female-authored papers relative to a citation-maximizing benchmark. In their desk rejection decisions, editors treat female authors more favorably, though they still impose a higher bar than would be implied by citation-maximization. We find no differences in the informativeness of female versus male referees, or in the weight that editors place on the recommendations of female versus male referees. We also find no differences in editorial delays for female versus male-authored papers.

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

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

Patricia Funk
Università della Svizzera Italiana
Via Giuseppe Buffi 13
6904 Lugano
Switzerland
patricia.funk@usi.ch

Nagore Iriberry
University of the Basque Country
Bilbao
Spain
nagore.iriberri@gmail.com

A randomized controlled trials registry entry is available at
<https://www.socialscienceregistry.org/trials/3048>

1 Introduction

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

Existing evidence on gender biases in the evaluation of economic research is mixed. Blank (1991) randomly assigned submissions at the *American Economic Review* to referees, with or without masking the author’s name and affiliation. She found a large but imprecisely estimated positive effect of blinding on the acceptance rate of female-authored papers. Broder (1993) analyzed reviews of National Science Foundation proposals, finding that female reviewers give lower average ratings to female-authored proposals. Abrevaya and Hamermesh (2012) find no significant gap in the relative evaluations of male- versus female-authored papers by male and female referees at an anonymous economics journal. Chari and Goldsmith-Pinkham (2017) similarly find no disparity in the acceptance rates of female- and male-authored papers for NBER conferences. Hengel (2018), however, finds that female authors face a higher bar in the review process. Focusing on the general climate in economics, Wu (Forthcoming) shows that online discussions of female economists often contain negative personal comments. Nevertheless, Donald and Hamermesh (2006) conclude that members of the American Economics Association exhibit a *positive* preference for female candidates for the Association’s executive board.

In this paper we analyze the role of gender in the evaluation process, using anonymized data on nearly 30,000 submissions to four leading economics journals: the *Journal of the European Economics Association*, the *Quarterly Journal of Economics*, the *Review of Economics and Statistics*, and the *Review of Economic Studies*. We combine paper characteristics – including the gender and previous publication record of the authors – with information on the gender of the referees assigned to the paper, their summary recommendations, the editor’s decision, and the ultimate citations received by the paper, regardless of whether it was accepted or not.¹ We use these data to analyze gender differences in how papers are assigned to referees, how they are reviewed, and how editors use referee inputs to reach a revise and resubmit (R&R) verdict.² We also examine the effects of author gender on desk-rejection decisions and editorial delays. Our analysis follows the

¹We do not have access to any textual information in the referee report, the editorial letter, or the paper itself. Our access agreement required us to extract the data from each journal’s archive using a stand-alone program that created a file with limited information on each paper, and no identifying information on authors or referees.

²Since the vast majority of editors in our 2003-2013 sample period were male, it is not possible to examine the impact of editor gender. The evidence in Bransch and Kvasnick (2017) suggests that having female editors does not increase the share of female-authored papers published in top journals.

analysis plan AEARCTR-0003048, which we drafted prior to the completion of our data collection to address concerns over data mining (see Christensen and Miguel, 2018).

We complement our submissions data with information from a survey of 141 economists, allowing us to compare our findings with the expectations of practitioners, as in DellaVigna and Pope (2018). We also elicit beliefs about gender differences in the link between citations and quality to aid in the interpretation of our findings.

We begin in Section 2 with an overview of our submissions database. This builds on the 4-journal sample collected by Card and DellaVigna (forthcoming), hereafter CDV, adding information on the gender of authors and referees and the publication record of individual co-authors.³ We obtained the names of authors and referees from each journal prior to our data extraction, then used a combination of first-name coding and internet search to assign genders. We validate our approach using a sample of published papers in the same journals, showing that we can assign gender to the authors of 97% of these papers with misclassification rates of less than 1 percent.

We find that two-thirds of the submissions in our database were written by all-male teams of authors, 8% by all-female teams, and 19% by mixed-gender teams. We distinguish mixed-gender teams by whether the most-published (“senior”) co-author is female (3% of all submissions) or not (16%), yielding four gender-mix categories for authors.

In Section 3 we analyze the matching process by which papers are assigned to referees. As in earlier studies (e.g., Dolado et al., 2012; Lundberg, 2017; Chari and Goldsmith-Pinkham, 2017), we find that the share of female authors varies widely across fields, with roughly proportional variation in the share of female referees. Even controlling for field and other factors, however, editors are 7 percentage points (50%) more likely to assign a female-authored paper to a female referee, suggesting that they are sensitive to gender-related issues in the review process.

In Section 4, we conduct a simple audit-style comparison of the summary evaluations submitted by female and male referees. Our most general models include paper fixed effects, allowing us to quantify differences in assessments of *the same paper* by referees of each gender. Consistent with the findings of Abrevaya and Hamermesh (2012) we find that the relative evaluations of female and male-authored papers by male and female referees are very similar.

Although these findings rule out any *relative* bias in males’ assessments of female-authored papers, they do not allow us to determine whether male and female authors face the same standards for publishing their work. To make further progress we need to make *between-paper* comparisons, accounting for differences in the quality of female- and male-authored papers. While there is no perfect measure of quality, we observe citations, which are correlated with quality and are highly relevant to publishers and editors. We therefore use citations as a noisy measure of quality, taking account of potential sources of divergence between the two.

We frame our analysis using the simple model developed in CDV. In this model, referee preferences for papers by different author groups can be inferred from the relationship between citations,

³CDV only collected the publication record of the co-author with the most previous publications. We also gathered more granular information on waiting times in the review process, information on the gender composition of the sub-field of the paper, and information on the complexity of the abstract.

referee recommendations, and author characteristics. If referees are unbiased judges, papers by male and female authors that receive similar referee evaluations will receive similar citations. Likewise, editors’ preferences for papers by different author groups can be inferred by comparing the relative effects of author characteristics and referee recommendations on citations and R&R probabilities. If editors select papers to maximize expected citations, the relative weights accorded to the referee recommendations and author characteristics in the R&R decision model will be proportional to their relative weights in the citation model.

There are two important caveats. First, if published papers receive more citations, conditional on quality, there is a potential “publication bias” in the relationship between paper characteristics and realized citations. As in CDV, we address this by including in the citation models an indicator for R&R status and a control function that corrects for endogeneity in the editor’s decision.

Second, there are potential gender biases in citations. Indeed, our survey respondents believe that females receive about 6% fewer citations than males, holding constant the quality of their papers. If males receive more citations, a finding of equal citations for male- and female-authored papers with similar referee recommendations implies that referees are setting a higher bar for female-authored papers. Assuming that female authors receive fewer citations conditional on quality, we obtain a *bound* on the gender gap in the referee recommendations.

We find that female-authored papers receive 22 log points (s.e.=0.05) *more* citations than male-authored papers, controlling for the referee evaluations. Our estimate of this gender gap is robust to alternative measures of citations and to a variety of alternative specifications. The magnitude of the gap *does* depend on whether we control for the prior publications of authors, since females have fewer prior publications and prior publications strongly predict citations. It also falls slightly (to 17 log points, with s.e.=0.05) if we add controls for the institutional affiliation of authors.⁴

For mixed-gender papers with a senior male co-author we find no difference in citations relative to those authored by all-male teams, consistent with our survey respondents’ expectations that such papers are treated about the same as male-authored papers. For mixed-gender papers with a senior female co-author, however, we find a 6 log point (s.e.=0.07) citation premium. We cannot reject that this premium is one-half as large as the premium for papers written by all-female teams, again consistent with our survey respondents’ expectations about how such papers are treated.

In Section 5 we study the R&R decisions of editors, given the referee recommendations and the characteristics of authors and their papers. On average editors tend to follow the referees’ recommendations, putting essentially no weight on author gender in their R&R decisions. This means that they are *over-rejecting* female-authored papers relative to a citation-maximizing benchmark.

We then use our modeling framework to compute how the R&R rate would change if editors were to reset the gender effects in their decision model to maximize expected citations. We estimate that the R&R rate for female-authored NDR papers would increase from 12.3% to 18.5%, with no change for mixed-gender papers. Assuming that about 30% of the papers written by a female economist

⁴We use as our benchmark specification the one without institutional prominence since it was the one pre-specified in the analysis plan, but we consider a large number of alternative specifications in our robustness tables.

have female co-authors and 70% have at least one male co-author (the shares in our sample), her average R&R rate would rise from 14.3% to 15.9%. We also perform a similar calculation, assuming that editors reset both the gender-related and publication-related decision weights to maximize expected citations. Under this scenario, the R&R rate for papers by female economists would rise from 14.3% to 16.0%, driven by a higher R&R rate for mixed-gender papers.

There are two main explanations for our finding that female-authored papers receive more citations, conditional on the referee evaluations. The first is that referees hold female authors to a higher bar, perhaps because of stereotype biases. The second is that female-authored papers have characteristics that lead to higher citations but are not as highly rewarded in the review process. For example, female authors may tend to write more empirically-oriented papers, or concentrate on certain topics within broad field categories that referees undervalue relative to expected citations.⁵

To provide some evidence on this second explanation, we conduct an in-depth analysis of 1,719 papers published in the four journals in our sample between 2008 and 2015 (approximately the same period that accepted papers in our main sample would be published). In this sample we find a citation premium of 12 log points (s.e.=0.13) for female-authored papers and 24 log points (s.e.=0.13) for mixed-gender papers with senior female authors, controlling for the same covariates as in our main analysis, apart from the referee evaluations. These premiums are much less precisely estimated than those from our main sample (which is nearly ten times larger), affirming the advantages of our main approach. Nonetheless, they replicate the key patterns, and are qualitatively consistent with the findings of Hengel (2019) for a different sample of published papers.

We then add controls for features of the published papers that are unavailable in our main sample. First, we include detailed 2-digit JEL field controls.⁶ These controls nearly double the R^2 in the citation regression, but barely affect the gender coefficients. Second, we include a series of controls to capture the theoretical versus empirical content of a paper based on word counts (words such as “proposition” and “standard error”) and on qualitative ratings. We find that papers with more formal modeling receive fewer citations, even controlling for field, and that female-authored papers tend to have higher empirical content and less formal modeling. Thus, controlling for measures of content leads to somewhat smaller citation premiums, 9 log points (s.e.=0.13) for female-authored papers versus a baseline of 12 log points, and 13 log points (s.e.=0.15) for mixed-gender papers with senior female authors versus a baseline of 24 points. This provides suggestive evidence that paper characteristics play some role in explaining the gender gap in citations, though given the imprecision of the estimated premiums we cannot reach a definitive conclusion.

Returning to our main sample of submissions, we examine the desk-rejection process, which provides information on editors’ preferences free from any referee inputs. Across all submissions

⁵A third explanation, that this gap reflects a tendency for female authors to submit papers that have been circulating longer, is unlikely given that we find similar findings using the SSCI citations, which accrue only to published papers.

⁶For confidentiality reasons, the field controls in our main database are limited to 13 broad groups. Thus, papers with code *J16* (economics of gender) are grouped with those in *J31* (wage level and structure), though our main analysis does include a control for the fraction of female authors in each 2-digit JEL code, and another for “gender-related” subfields.

(refereed or not), female-authored papers receive 24 log points more citations than male-authored papers, conditional on other controls. An editor who sets a bar for desk rejection based on expected citations should therefore give positive weight to female authorship. Consistent with this prediction, we find that editors desk reject fewer female-authored papers, holding constant other characteristics. The gap, however, is smaller than predicted by a citation-maximizing benchmark.

We then address two further issues in the editorial process: Are some referees more reliable in judging quality (as revealed by citations)? Do editors pay more attention to more reliable referees? CDV find that the recommendations of more and less prolific referees are equally predictive of future citations, yet editors tend to place more weight on recommendations from referees with more prior publications. In the case of gender, we find that male and female referees are equally informative, and that editors place equal weight on their recommendations in reaching an R&R decision. Thus, editors appear to be gender-neutral in their use of referee inputs.

Finally, in Section 6 we study gender-related differences in the speed of the review process. We find no gender differences in the time that referees take to return a recommendation, or in the time that editors take to reach a decision, or in the total length of time between submission and acceptance for published papers. We conclude that female and male authors experience similar editorial delays, in contrast to Hengel (2018) who studies gender differences in the time from submission to acceptance at one journal.

In light of these results, in our concluding section we offer a partial reconciliation of the myriad findings in the literature. Some of these differences appear due, at least in part, to the different strategies used to identify potential gender discrimination. One strategy, used by Broder (1993) and Abrevaya and Hamermesh (2012), is an audit-style comparison of the recommendations provided by different reviewers of male- and female-authored papers. These studies, like us, find no evidence of differential biases against female-authored papers by male referees. This suggests that the animus documented by Wu (2018) against female economists in online discussion boards (which is widely attributed to male commentators) is absent in the review process.

A second strategy is to compare outcomes for male- and female-authored papers without explicit quality controls, e.g, the analysis of NBER submissions by Chari and Goldsmith-Pinkham (2017). Like these authors, we find that female- and male-authored papers have similar R&R rates when we do not control for prior author publications. A third strategy is to compare outcomes conditional on quality controls. We are aware of only one prior analysis using this design – Donald and Hamermesh (2006) – which comes to the opposite conclusion as us, albeit in a different setting (the election of AEA officers). Finally, a fourth strategy is to compare outcomes when author identity (including gender) is blinded. The only such study we are aware of, by Blank (1991), finds a relatively large but imprecisely estimated effect of blinding on the acceptance rate of female-authored papers. Blank’s findings are consistent with ours, but her sample lacked the power to precisely estimate gender differences. We return to the implications of our findings in the conclusions.

2 Data Extraction, Gender Coding, and Summary Statistics

2.1 Data Extraction

Our data are derived from information stored in the Editorial Express (EE) system used by each of the four journals. For confidentiality reasons we wrote a program that could be run by journal staff to create an anonymized data base, combining information in the EE system with gender information from pre-coded lists of author and referee names (see below). We are grateful to the four journals for agreeing to provide data access.⁷

Our database builds on the submissions extract created by CDV in mid-2015. Google Scholar (GS) has created new barriers to accessing its data base in the past few years. We therefore elected to match our new data base back to the CDV data set and use the citations originally collected by CDV.⁸ We include all submissions from the first year each journal started using the EE system up to 2013 (as in CDV), leaving at least 1.5 years for citations to be realized.

Gender Coding. Prior to running our extraction program we obtained a list of the names of all authors and referees from each journal for the relevant years. We then developed a protocol for assigning gender to the names on these lists. The protocol, laid out in Online Appendix Figure 1, relies on a combination of (1) public lists of given names including the fractions of people in the US and Germany with that name who are male; (2) lists of female economists’ names; (3) a list of common Chinese given names; (4) internet search by a team of research assistants. We developed the protocol using a test data set of the names of 48,000 authors of articles published between 1990 and mid-2017 in a set of 63 economics journals (Online Appendix Table 1).

Our protocol begins by assigning “unknown gender” to common Chinese first names, since these names can be used by both males and females, and there are often multiple economists with the same Chinese name. This exclusion affects less than 1% of names. We then classify an author as **female** if *both* the US and German lists report that less than 1% of people with that first name are male, or if the full name is present in one of the lists of female economists. Likewise, we classify an author as **male** if one of the US or German names lists shows that over 99% of people with that name are male and the other shows at least 50% are male.⁹ Finally, a team of undergraduate research assistants looked up all names that remained unassigned.¹⁰ Overall we are able to assign gender to about 97% of names in our test data set. As explained in Section 5.2, an audit of the

⁷The data agreement with the journals has two conditions: (i) no analysis should present separate results by journals, and (ii) unlike the CDV data set, this supplemented data set will not be posted, even upon publication.

⁸Since CDV did not retain paper identifiers, we used a fuzzy match algorithm based on all the identifying variables stored in the CDV data base. This yields perfect matches for all non-desk rejected papers, but many-to-many matches for some desk-rejected papers. For papers with multiple matches, we calculate our primary measure of citations as a simple average of $\text{asinh}(\text{citations})$ across all possible matches, though our results are virtually identical if we retain all possible matches and weight by the inverse number of matches for a given paper.

⁹These cutoffs were derived using the Econlit test data set. Since only 20% of economists are female, we have to set higher cutoffs for assigning female gender than male to (roughly) equate the misclassification rates.

¹⁰If the assistant initially assigned to the name could not find a match, it was passed on to a second assistant. We had two separate assistants code a subsample of names that could not be assigned by the first-name procedure. The coders disagreed only 1% of the time; in 11% of cases neither coder was able to find a name match; and in 14% of cases only one of two coders found enough evidence to determine a gender.

genders assigned by our procedure for authors of 1,719 *published* articles in the four journals of our main database yields an error rate of under 1% (Online Appendix Table 11).

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

Survey. To help interpret our findings, we conducted a survey of editors and economists, asking about their perception of gender differences in the publication process. The survey, approved under Berkeley IRB 2018-04-10955, was sent to three groups: (i) editors and co-editors at the 4 journals in our sample; (ii) a stratified random sample of 200 economists (100 male and 100 female) with at least 4 publications in our top-35 journal set from 2013 to 2017; (iii) all assistant professors of economics in the top 20 US schools and top 5 European schools with PhDs in 2015-17. The views of editors and co-editors are obviously relevant given their role in the publication process. The other two groups were selected to represent the views of economists with considerable recent publication experience, and of promising researchers at the start of their careers.¹¹ Our response rates were reasonably high, ranging from 35% to 50% depending on the group.

Table I summarizes the responses to questions focusing on (i) how papers by mixed-gender teams are treated in the review process; (ii) whether female-authored papers are more likely to be assigned to female referees; (iii) how male and female referees evaluate male- and female-authored papers; (iv) the likelihood that the editor gives an R&R to male- versus female-authored papers; (v) the extent to which citations vary with author gender, holding constant the quality of a paper; (vi) the informativeness of male and female referees; and (vii) the degree to which editors follow the recommendations of male and female referees.

We use these answers in three ways. First, following our analysis plan, we use the beliefs about mixed-gender teams to motivate our classification into “senior female” and “senior male” teams. Second, we use beliefs about the potential differences in citations for male-authored and female-authored papers to help interpret the gaps in citations in our analysis. Third, we use the answers as “priors” to benchmark our findings, as in DellaVigna and Pope (2018).

2.2 Summary Statistics

Table II presents summary statistics for the 15,147 submitted papers in our database that were not desk-rejected (NDR) and were assigned to at least two referees. We present summary statistics for the full set of 29,890 submissions in Online Appendix Table 2.¹²

We classify papers into five groups based on the author-gender mix: (i) all-male teams (67% of NDR papers); (ii) all-female teams (7%); (iii) mixed gender teams with a senior female co-author, i.e., the co-author with most prior publications is female (4%); (iv) other mixed gender teams (17%); and (v) gender undetermined, i.e., at least one co-author with unassigned gender (4%).

¹¹Within each group, we did not keep track of individual respondents. Within the second and third group, however, we referred male and female respondents to different URL’s to keep track of gender.

¹²Among the non-desk-rejected papers, we exclude papers that were assigned to only one referee, since this process (which is especially common at the *Review of Economic Studies*) appears to be a form of desk-rejection.

Note that mixed gender teams in which none of the authors has any prior publications are assigned to the “other mixed gender” category, leading to a higher average publication record for the mixed gender teams with a senior female author than for any of the other groups (see below).

The top row of the table shows our benchmark measure of citations, the inverse hyperbolic sine ($asinh$) of GS citations, collected in mid 2015.¹³ Citations are highest for mixed-gender papers (columns 3 and 4), followed by all-male papers (column 1), with all-female papers (column 2) at the bottom. We plot the cumulative distributions of $asinh(citations)$ by author-gender group in Online Appendix Figure 2a, and of residualized citations after adjusting for our key control variables (including prior author publications) in Online Appendix Figure 2b. The adjustment reverses the ranking, with higher residualized citations at most quantiles for female-authored papers.

Next we show the probability of an R&R verdict. Consistent with their relative citations, NDR papers by mixed-gender teams have the highest probability of an R&R, followed by all-male papers, with papers by female authors at the bottom. The editorial outcomes for all submissions (including desk rejections), summarized in Figure I Panel A, show a parallel pattern, with the lowest rates of desk rejection and highest unconditional rates of R&R for mixed-gender teams.

We measure the productivity of authors by the number of publications in 35 high-impact journals over the 5 years prior to submission (see Online Appendix Table 1 for the list of journals). On average across all authors in our submissions data, about 50% of males and 65% of females have no prior publications (see Figure I Panel C). We assign previous publications to each *paper* using the publication record of the most prolific co-author. As shown in Table II, authors of mixed gender papers have the highest numbers of prior publications. This is particularly true for mixed gender teams with a senior female co-author, though part of the gap is mechanical given how we assign papers for which none of the authors has any prior publications.

The number of coauthors is a key characteristic of papers that is highly correlated with author gender, since the likelihood of an all-female team of 2+ authors is relatively low in a field where only 20% of authors are female. Indeed, the rate of single authorship is over twice as high among all-female papers (72%) than all-male papers (36%) in the NDR sample.¹⁴

Another important characteristic of papers is field. We classify submissions into 13 broad groups based on their JEL code(s) and include separate codes for papers with no JEL codes, and for those that fall outside the main groups.¹⁵ Author gender is correlated with field: female authors are more prevalent in the empirical micro fields like labor and development, and less prevalent in

¹³We use the $asinh$ transformation to accommodate zero citations. Bellemare and Wichman (forthcoming) derive the elasticity of y with respect to x for a specification like $asinh(y) = \beta x + u$. For a continuous x this is $(\beta x/y)\sqrt{y^2 + 1} \approx \beta x$ for $y \geq 2$. To address concerns about 0 citations, we present Poisson and Tobit models below, as well as models using $\ln(1 + y)$ as the dependent variable.

¹⁴Interestingly, the fraction of single-authors who are female is very close to the fraction of authors of 2-author papers who are female, suggesting that females are no more likely to work alone. There is, however, some evidence of assortative matching of co-authors by gender. For example, among 2-authored papers the fraction written by two females is 4.1%, higher than the 2.5% rate expected under random matching.

¹⁵We classify a paper with J JEL codes as being assigned with weight $1/J$ to each of the listed fields. Some 11% of papers in our database do not have JEL information at submission and are thus coded as having missing field; most of these observations are due to two of the journals not collecting JEL information over 2 years.

econometrics, finance, macro, and micro theory. Reassuringly, the share of female authors in different fields in our database is highly correlated with the share among papers in the EconLit database we use to validate our gender assignment process (Online Appendix Figure 3a), as is the share of female referees assigned to papers in a field (Online Appendix Figure 3b).

Since our data access agreement precluded the extraction of detailed JEL codes, we developed two measures of the gender-related content of different subfields. The first is the share of female authors of papers with each 2-digit JEL code published in the past 5 years in our EconLit sample of journals. The second is a simple indicator for JEL codes associated with gender-related topics, which we take to be D1 (Household Behavior and Family Economics), J1 (Demographic Economics), K36 (Family and Personal Law), and K38 (Human Rights Law and Gender Law). As Table II shows, all-female teams and mixed-gender teams with a senior female co-author are more likely have JEL codes in “high female share” fields, and to include JEL codes indicating a gender-related topic.

Next we turn to the referee recommendations and characteristics of the referees. The summary recommendations in EE fall into 7 categories, ranging from “Definitely Reject” to “Accept”.¹⁶ Figure I Panel B shows the distributions of summary recommendations by author gender. Most papers receive a “Reject” or “Definitely Reject” recommendation. Again, mixed gender teams are more likely receive a better review, while all-female author teams are less likely.

Finally, turning to the referees, on average about 15% are female. As is the case for authors, female referees tend to have fewer publications than male referees (Figure I Panel D).

3 Assignment of Referees

We begin our analysis by studying the matching process used by editors to assign NDR papers to referees. This tells us how editors treat different types of mixed gender teams, and provides some potential evidence on editors’ concerns about gender-related issues in the review process.

As shown by the first set of vertical bars in Figure II, papers by all-female authors are assigned to female referees at nearly twice the rate (26 percent) as papers by all-male authors (14 percent). Mixed-gender author teams fall in between, with a higher rate for those with a senior female co-author (21 percent) than for the other mixed-gender papers (18 percent).

These simple comparisons do not account for other relevant paper characteristics, such as the field of the paper. We thus estimate linear probability models for the event that a paper is assigned to a female referee with our full set of controls, including journal-by-submission-year effects, the number of authors, the number of publications of the most prolific co-author, indicators for broad field, the share of female authors in the 2-digit JEL code(s) of a paper, and an indicator for gender-related subfields. We also control for the publication record of the referees assigned to each paper. We plot the residualized differences by author gender using the second set of bars in Figure II.¹⁷

¹⁶There are actually 8 categories, but “Conditionally Accept” and “Accept” are so rare we group them in one.

¹⁷Online Appendix Table 3 reports the corresponding coefficients. We also examine, for comparison, the probability of being assigned to a referee with 3 or more recent publications. Controlling for other variables, female-authored papers are significantly less likely to be assigned to such referees, but the effect is small, a 5 percent reduction.

Compared to all-male authored papers, female-authored papers are 7 percentage points (s.e.=1 ppt) more likely to be assigned to a female referee, mixed-gender papers with a senior female co-author are 5 percentage points more likely, and other mixed-gender papers are only 3 percentage points more likely. The differences between the two mixed-gender groups motivate our choice to analyze these groups separately, along the lines described in our analysis plan.¹⁸

One explanation for this gender-matching is that editors are concerned about reviewer biases. Indeed, the editors in our survey think that male reviewers are less likely to give a positive evaluation of female-authored papers than female reviewers (17.5% probability versus 20.7%). They are also more likely to expect gender-matching in the assignment of female authored papers than other groups in our survey. Another possible explanation (for which we have no evidence) is that female authors are more likely to cite other female authors – perhaps because they are working in an area dominated by female authors – and editors tend to select referees from a paper’s bibliography. It is also however possible that the gender difference is due to subfield differences in the share of female authors and referees which our field controls do not capture.

4 Referee Recommendations

4.1 Simple ‘Audit’ Comparison

The next question is whether referee gender actually affects the evaluations of papers. We use two main measures of referee support. The first summarizes the seven referee recommendations into an index based on the predicted $\text{asinh}(GS \text{ citations})$ associated with each category, using the coefficients from the main citation model in CDV (Table 2, Column 4). A second, simpler measure is the share of recommendations that are positive – that is, “Revise and Resubmit” or better.

Figure III Panels A and B compare the mean assessments by referee gender and author gender mix for papers assigned to at least one referee of each gender, with confidence intervals constructed by clustering at the paper level. We weight paper/referee observations by the inverse number of referees for a paper so that each paper receives equal weight. On average, female and male referees give very similar evaluations, tracking each other closely across author groups. There is no evidence of a *relative assessment gap* between male and female referees that depends on author gender.¹⁹

Table III presents a series of models at the referee-paper level, weighing by the inverse number of referees for the paper. With no additional controls (Columns 1 and 5), papers by all-female authors receive significantly lower evaluations, as do those by author teams with undetermined gender. When we add controls for field, number of authors, and previous publications (columns 2 and 6), however, we find no differences in how the referees assess all-female or mixed gender papers with a senior female author relative to all-male papers (the omitted group).

¹⁸In our analysis plan we wrote “We intend to use this pattern of assignment to infer how editors classify mixed-gender papers. Suppose for example that the rate of assigning a female referee is different for mixed gender papers with a senior author who is female [...] than for mixed gender papers with a senior author who is male. This would suggest that it is important to analyze the two mixed-gender author groups separately.”

¹⁹We find a similar pattern if we use the full set of submissions up to 2017 (Online Appendix Figure 4a-b).

To test for any *differential assessment* of male- versus female-authored papers, we include paper fixed effects in columns 3 and 7, identifying the within-paper difference in assessments between male and female referees. By interacting referee gender with the author gender mix we identify the *differences-in-differences* in the evaluation of female versus male referees for papers by a particular author gender group, relative to all-male papers. In Columns 4 and 8 we add additional controls—interaction between referee gender and author publication, referee gender and field of the paper, and author gender mix interacted with referee publications—to ensure that the estimates are not biased by the fact that, say, female referees have fewer publications and referees with fewer publications are more likely to recommend an R&R for a particular author gender team.

Using any of these specifications, we find no large or statistically significant interactions between referee gender and author gender mix. Thus, there is no evidence of any *relative bias* by referees of one gender for or against papers by authors of the other gender.²⁰ This differs from the predictions of our survey respondents, who expected that female referees would be a bit more positive toward female-authored papers. In fact, if anything female referees (controlling for referee publications) are somewhat less likely to recommend positively on papers of all genders.

Our results are consistent, however, with findings in Abrevaya and Hamermesh (2012) who compare the effects of having a female referee for female-authored papers at an anonymous journal with double-blind refereeing, comparing submissions in 2000-08 (when internet searches could presumably reveal the author identities) to those in 1986-94 (when authors were more likely anonymous to the referee). Our estimated interaction effect between all-female authors and female referees in Column 8 (-0.007, s.e.= 0.031) compares to their “triple interaction” estimate of -0.01 (s.e.=0.08).

4.2 Model for Recommendations and Citations

Although the results in Table III rule out any large or statistically significant *relative bias* among male versus female referees, they do not necessarily imply that referees set the same standards for female and male authors. It is possible that referees of both genders are biased for, or against, female-authored papers. To make further progress we need to make comparisons *across* papers by different gender groups, taking into account differences in quality.

We proceed by using realized citations as an indicator of quality. Building on the simple model in CDV, we assume that each paper has true quality q which is only partially observed by editors and referees. At the R&R decision stage, the editor observes the gender composition of the author team, F , a set of other characteristics of the paper and author(s) x_1 , and the referee recommendations R . For expositional purposes we assume that F and R are 1-dimensional, though in our analysis both are multi-dimensional. We assume that expected quality is a simple linear function of $x \equiv (x_1, F, R)$:

$$E[q|x] = \beta_0 + \beta_1 x_1 + \beta_F F + \beta_R R. \quad (1)$$

In addition, we assume that the prediction error $\phi_q \equiv q - E[q|x]$ is normally distributed with mean

²⁰Online Appendix Table 4 shows that these results do not appear to differ for more “senior” female referees (that is, referees with 3+ publications), versus the other female referees.

0 and standard deviation σ_q .²¹ Equation (1) allows for the possibility that the referees set higher or lower thresholds for assigning their recommendations based on the characteristics of the paper or its authors. If, for example, referees tend to give more negative evaluations to female-authored papers conditional on the factors included in x_1 , then $\beta_F > 0$.

The second component of the model is the decision process of editors. We assume that the editor observes a noisy signal s of the error component ϕ_q , with $s = \phi_q + \zeta$, where $\zeta \sim N(0, \sigma_\zeta^2)$. Conditional on s and x , the editor's expectation of the paper's quality is:

$$E[q|s, x] = \beta_0 + \beta_1 x_1 + \beta_F F + \beta_R R + v, \quad (2)$$

where $v \equiv s \times \sigma_q^2 / (\sigma_q^2 + \sigma_\zeta^2)$. With this forecast in hand, the editor then decides whether to reject the paper or invite a resubmission. For simplicity we assume that the editor makes a revise and resubmit decision ($RR = 1$) if $E[q|s, x]$ is above some threshold:

$$RR = 1 \iff \beta_0 + \beta_1 x_1 + \beta_F F + \beta_R R + v > \tau_0 + \tau_1 x_1 + \tau_F F, \quad (3)$$

where τ_1 and τ_F incorporate editorial tastes for papers with different characteristics. In particular, if $\tau_F > 0$ the editor imposes a higher standard for female-authored papers.

Under our assumptions v is normally distributed, and the R&R decision is a probit model:

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

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

The third and final component of the model describes the determination of citations. We assume that cumulative citations c (measured some time after the determination of RR status) reflect a combination of quality and other factors summarized by η : $c = q + \eta$. One important factor included in η is the time from submission to the measurement of citations: we assume that this is incorporated in x_1 . In addition, citations can depend on the gender of the author team, on the editor's R&R decision (which will determine whether the paper is published in the particular journal or not) and other random factors captured in an error component ϕ_η :

$$\eta = \eta_0 + \eta_1 x_1 + \eta_F F + \eta_{RR} RR + \phi_\eta. \quad (5)$$

Combining equations (5) and (1) leads to a simple model for citations:

$$c = \lambda_0 + \lambda_1 x_1 + \lambda_F F + \lambda_R R + \lambda_{RR} RR + \phi \quad (6)$$

²¹A more complete model would start from an assumption about the formation of referee recommendations, conditional on (x_1, F) , and derive the posterior distribution of quality conditional on (x_1, F, R) . We note that if referees submit their posterior estimates of quality given paper characteristics and their own reading of the paper – so $R = E[q|x_1, F, read]$ – then we would expect $\beta_0 = \beta_1 = \beta_F = 0$ and $\beta_R = 1$.

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

A concern for estimation of the λ coefficients in equation (6) is that when editors have private information, RR status will be correlated with ϕ . In this case a simple OLS regression of citations on the variables (x_1, F, R) and RR status can lead to biases. To address this, we assume that different editors have different quality thresholds for reaching an R&R decision (i.e., different values of τ_0) but that the particular editor assigned to a paper has no direct effect on citations. This allows us to use the “leave out” R&R rate for the editor assigned to a paper as an instrumental variable for the editor’s decision. Rather than estimating equation (6) by IV, we adopt a control function approach and include the generalized residual $\hat{r}$ from a probit model for the editor’s decision that includes x and the instrumental variable. Specifically, we estimate the augmented citation model:

$$c = \lambda_0 + \lambda_1 x_1 + \lambda_F F + \lambda_R R + \lambda_{RR} RR + \lambda_r \hat{r} + \phi'. \quad (7)$$

As shown by Wooldridge (2015), the inclusion of $\hat{r}$ absorbs the endogenous component of the R&R decision, yielding consistent estimates of λ_{RR} and the other coefficients of interest.

By comparing the effects of author gender on citations and the R&R decision, we can make an inference about the relative threshold that editors impose on female authors. As a starting point, suppose that $\eta_F = 0$, so the gap between citations and paper quality is unaffected by author gender. In this case the coefficients of R and F in the citation model are $\lambda_R = \beta_R$ and $\lambda_F = \beta_F$, while the corresponding coefficients in the R&R probit model are $\pi_F = (\beta_F - \tau_F)/\sigma_v$, and $\pi_R = \beta_R/\sigma_v$. If editors impose the same bar for male and female authors (i.e., $\tau_F = 0$) we would expect to find

$$\pi_F = \beta_F/\sigma_v = (\lambda_F/\lambda_R) \times \pi_R = \pi_F^{cite-max} \quad (8)$$

in the R&R model. In the absence of editorial preferences for or against female authors, and any gender bias in citations, the effects of the referee recommendations and the female indicator will be *proportional* in the models for citations and the editor’s decision. The expected coefficient in the R&R probit under these assumptions is the same as the one that would emerge if editors selected papers based on the highest expected citations. We therefore refer to equation (8) as the “citation-maximizing” benchmark for the coefficient of F in the R&R probit.

In contrast, if editors impose a *higher* bar on female authors (i.e., $\tau_F > 0$), then we would expect a systematic deviation from proportionality, with

$$\pi_F = (\beta_F - \tau_F)/\sigma_v < \pi_F^{cite-max},$$

i.e., a more negative effect of author gender on R&R rates. Conversely, if they impose a *lower* bar on female authors, then we would expect $\pi_F > \pi_F^{cite-max}$.

Suppose now that $\eta_F < 0$, so female authors get fewer citations than males, controlling for quality. In this case, the coefficient of F in the citation model is $\lambda_F = \beta_F + \eta_F < \beta_F$. If editors impose the same *quality* bar for male and female authors we would expect a coefficient for female

authors in the R&R model that is *more positive* than under the citation-maximizing benchmark: $\pi_F = \beta_F / \sigma_v > \pi_F^{cite-max}$. Thus, if $\eta_F < 0$, the citation-maximizing benchmark is a *lower bound* on the expected coefficient for female-authored papers in the R&R model. In this case, a finding that the estimated coefficient in the R&R model is below the citation-maximizing benchmark can be interpreted as strong evidence that editors are not evaluating papers according to expected quality.

Gender Bias in Citations. To get a sense for the magnitude of any potential gender gap in citations, we asked our survey respondents to quantify their beliefs about the gap: *“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? 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.”*

While this question is admittedly a difficult one to answer, we found agreement on two points. First, the modal response across all groups is that there is no differential citation bias. Second, to the extent that there is a difference, respondents expect that female authors receive fewer citations than male authors (see Table I and Online Appendix Figure 5). The mean elicited citation bias is -6.5 log points: -3.5 to -4.7 log points among male respondents; -10.2 to -11.1 log points among female respondents; and -3.8 log points among editors.

4.3 Citations Conditional on Referee Evaluations

With this background, we consider the evidence on citations, starting with graphical evidence. In Figure IV Panel A, we plot average $\text{asinh}(\text{citations})$ for each recommendation category, separately by author gender mix. We residualize citations with respect to all the controls, including the prior publications of authors.²² At nearly each referee recommendation, female-authored papers have higher citations than male-authored papers, with a 20 log point average difference. This suggests that papers by all-female authors are held to a higher bar by the referees. This pattern is the same for male and female referees (Figure IV Panel B), consistent with the findings in Table III.

To aggregate across the recommendations of the different reviewers, we turn to the regression-based models in columns 1-4 of Table IV.²³ We summarize the referee opinions by the fractions of recommendations in each of the 7 categories allowed in Editorial Express. CDV document that this simple procedure provides a relatively accurate representation of the effect of the recommendations on both citations and the editor’s R&R decision. To account for the facts that citations accumulate gradually over time, and that average citations likely differ for papers submitted to different journals, we include journal $\times$ submission-year effects in all our models.

In a specification that controls only for the referee recommendations and journal-year fixed

²²Online Appendix Figure 6d shows the same graph with confidence interval. Online Appendix Figure 6a shows the evidence for the raw citation variable, with no controls.

²³The specifications in this table follow exactly the format laid out in the analysis plan.

effects (Column 1), female-authored papers receive 7 log points (s.e.=0.05) fewer citations than male-authored papers and mixed-gender papers receive 26-37 log points more citations. Once we add controls for prior publications, the number of co-authors, broad fields, and our two measures of gender-related fields (Column 2), however, the estimated citation premium for all-female teams rises to 24 log points (s.e.=0.06), while the premiums for mixed-gender teams fall and are insignificantly different from 0. These results show the importance of controlling for paper and author characteristics in making inferences about gender-related citation gaps.

We stress the interpretation of the results in Column 2. This specification controls for all observables, but *not* for referee recommendations. Thus, the 24 log point citation premium for all-female papers implies that, among all (non-desk-rejected) submissions, the papers with all-female authors appear to have higher quality, conditional on other observables.

How do these results change once we control for the referee recommendations in Column 3? All-female papers still have a significant citation premium of 22 log points, and mixed-gender papers continue to have small and statistically insignificant premiums relative to all-male papers.

As mentioned above, a confounding issue is that, to the extent that published papers get more citations, the effect of the referee reviews on citations could be biased. In column 4 we therefore add an indicator for a paper’s R&R status and a control function to deal with the endogeneity of R&R status. Specifically, the control function is the estimated generalized residual from the probit model for the editor’s R&R decision shown in column 7 of Table IV (and discussed below) that includes the leave out mean R&R rate of the editor who handles each NDR paper as an instrumental variable for R&R status.

This specification, which we take as our benchmark citation model, yields very similar estimates to the ones in the simpler model in column 3. In particular, the estimated citation premium for female-authored papers remains at 22 log points (s.e.=0.05). We interpret this gap as implying that a paper by an all-female team on average needs to have 25 percent ($\exp(0.22) - 1 = 0.25$) higher citations (relative to a similar paper by an all-male team) to receive comparable recommendations. This gap is equivalent to the difference in citations between a paper that receives two “Weak R&R” recommendations and one that receives one “Weak R&R” and one “R&R” recommendation.

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

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

4.4 Robustness and Heterogeneity

Robustness. Three possible concerns with the estimated citation premiums for female-authored papers in our benchmark model are that (i) they may reflect the effect of some unobservable variable that is correlated with gender, (ii) they may depend on the functional form used for measuring citations, or (iii) they may be due to a peculiar subset of the data. We thus consider a broad spectrum of robustness checks, summarized in Table V. For each specification (shown in a separate row), Columns 1-3 display the coefficients on the three author-gender variables from our citation regression, including the full set of controls as in Column 4 of Table IV. We discuss below the associated coefficients for the R&R decision, reported in Columns 4-6. For several of the robustness specifications, we report additional information in online appendix tables.

As far as unobserved factors, adding controls in Table IV significantly *increased* the estimated citation premium for all-female papers. If other unobservables tend to have the same correlations with female authorship and citations (as formalized in Altonji, Elder, and Taber, 2005), their omission would tend to downward-bias the estimate of the all-female citation premium. To provide more detail, Online Appendix Table 5 displays citation models with subsets of controls (including in all cases the referee recommendations and year-journal fixed effects). When we add only field controls, the results are similar to the specification with no controls. Adding the author publication variables shifts the coefficient on the all-female papers to 0.14 log points (s.e.=0.05), indicating that author publications are the key controls. Further adding controls for the number of authors raises the premium for all-female papers to 0.22, our benchmark specification.

Another possibility is that our controls for the publication record—the maximum number of publications in the 5 years prior to submission across the coauthors—are too crude. When we add controls for the *average* number of publications among the coauthors, for publications in top-5 journals (as opposed to 35 high-quality journals), and for publications 6-10 years prior to submission, the citation premium for all-female papers is unaffected. When we further add measures of the quality of the institution of the co-authors (reported also in row 1 of Table V), the citation premium for all-female authors falls to 0.17, while the premium for mixed-gender teams with a senior female coauthor falls to 0.04. On average, female authors are located at slightly more prestigious institutions, and papers by authors at higher-ranked institutions get more citations, so the addition of these controls lowers the estimated female author premiums by 2-4 log points.

A third possibility is that our field controls are not sufficiently detailed, and that females tend to write papers in particular sub-fields which tend to get more citations. Contrary to this line of reasoning, however, Table IV shows that both the average share of female authors in a sub-field and the share of sub-fields in gender-related areas have small, insignificant effects on citations. Still, to further address this, in the specification reported in row 2 of Table V (also reported in Online Appendix Table 5) we introduce interaction terms that allow the effect of the field variables to differ by year, and the effect of the gender-sub-field variables to differ by field. These additional controls slightly increase the female-author premium to 0.19 log points (s.e.=0.05). Taken as a whole, there is little evidence of an upward bias due to sub-field differences.

Next, in Online Appendix Table 6 we focus on alternative specifications of the citation model: (i) $\ln(1 + citations)$; (ii) a Poisson count model; (iii) the percentile of *citations* within the cohort of submissions to the same journal in the same year; (iv) an indicator for papers in the top x percent of citations, where x corresponds to the R&R rate in that journal-year cell (summarized in row 3 of Table V); (v) an indicator for “superstar” papers in the top 2 percent of citations for the journal-year cell; (vi) a Tobit specification for $asinh(citations)$, top-coding citations at 100. While one cannot directly compare the estimated citation premiums in these various models, we can compare their magnitudes to those of the referee recommendations. In our baseline model, the coefficient for an all-female authored paper is 12% as big as the coefficient for an R&R referee recommendation ($0.22/1.89=0.12$). In the probit model for being in the top x percent of citations, that ratio is similarly 12% and it rises to 22% in the probit model for being in the top 2% of citations, suggesting that female-authored papers have an even higher relative likelihood of achieving “superstar” status.

We then address two additional concerns about our citation measure: the role of papers with zero citations; and our use of Google Scholar rather than Social Science Citations Index (SSCI) citations. About 19% of all NDR papers in our data base have zero GS citations; it is possible that some of these papers may be recorded as zero citations due to, say, a special character in the author name or a typo in the title.²⁴ We deal with this in two ways: re-estimating the model excluding papers with 0 citations; and by explicitly modeling the left censoring of citations at zero using a Tobit model (see columns 7 and 8 of Online Appendix Table 6). In both cases, the relative coefficient of all-female papers is 12% as big as the coefficient for an R&R referee recommendation, as in our benchmark specification.

Finally, we re-estimate the model using SSCI citations, which only include citations to published papers from other published papers. We limit the sample to submissions in the years 2006-2008 and fit a left-censored tobit model, given that, even among 2006-2008 submissions 60.8% have zero SSCI cites. This specification yields a 0.32 citation premium for all-female papers (s.e.=0.15), 16% as large as the estimated effect of an R&R recommendation.²⁵ Thus, our finding of a sizable positive citation premium for female-authored papers is robust to how we measure citations.

Heterogeneity. Table V also reports a variety of estimates fit to different subgroups. We estimate a larger female-author effect for papers from the earlier years (2003-2009, in row 6) than in the later years (2010-2013, in row 7). Citations to older papers are presumably less affected by factors such as conference presentations and prior circulation of working papers, so the finding of a larger female premium for older papers suggests that male authors may have short term advantages that if anything lead us to under-estimate the premium for all-female author teams.

Next, we estimate our baseline model separately for papers with 1 author, 2 authors, and 3+ authors (with the full results in Online Appendix Table 7). We estimate a citation premium for female authors of 17 log point with a single author (row 9) and 34 points with 2 authors (row 10).

²⁴We investigate the possible correlation between author gender and missing citations using a subset of published papers in Section 5.2, and find no evidence of such a correlation.

²⁵See column 10 of Online Appendix Table 6. For reference we also re-estimate our preferred model for $asinh(GS\ citations)$ using a Tobit model for the 2006-2008 period, yielding similar estimates.

For papers with 3+ authors we cannot reliably estimate the impact of an all-female team, given the rarity of such papers, but we estimate a 23 log point premium (s.e.=0.09) for mixed-gender papers with a senior female co-author (row 11).

In rows 12-13 we estimate our model separately for papers with fewer (0-3) versus more (4+) prior author publications. We find a positive citation premium for all-female author teams in both sub-samples, but the result is particularly large (49 log points, s.e.=0.10) for papers written by prolific teams. Unlike in Bohren et al. (forthcoming), we do not find a reversal of the pattern of gender differences for more highly experienced female authors.

Next, in rows 14-16 we estimate the model for sub-fields with above- or below-median shares of female economists. The citation gap for all-female papers is actually ***larger*** in fields with a higher share of female economists (like labor economics) than in fields with a lower share (like theory), though the difference is not statistically significant ($t=1.4$). This pattern is not supportive of the idea that the female-author premium arises from discrimination against female authors that is more prevalent in fields with fewer female authors.

Finally, we consider sub-samples based on referee characteristics. The female-authorship gap is larger for papers with 1 or 2 referees (row 17) than for papers with 3 or more referees (row 18), though the difference is at best only marginally significant ($t=1.7$). Interestingly, the gap is smaller for papers sent to only male referees (row 19) than for papers with at least one female referee (row 20) (though again the difference is not significant, $t=1.3$). Consistent with earlier findings we see no evidence that male and female referees are differentially biased against female authors.

5 Editorial Decisions

So far, we have focused on how referees treat teams of authors with different gender compositions. But referees' opinions are only part of the editorial process: editors make the ultimate decision of whether to reject a paper or invite a revision. Moreover, editors make an initial screening decision on whether to desk-reject a paper or send it to referees. In this section we study the effects of author gender on editors decisions, again following our pre-analysis plan.

5.1 R&R decision

As we saw above, female-authored papers tend to get more citations at each level of the referee's recommendation (Figure IV Panel A). In contrast, the R&R rates are very similar for female- and male-authored papers, conditional on the referee's opinion (Figure IV Panel C). This suggests that editors do not undo the relatively negative referee assessments of female-authored papers.

In Columns 5-7 of Table IV we fit a series of probit models for the R&R decision using the same variables as in our citation models. Our baseline specification in column 7 controls for the recommendations, the full set of controls, and the leave-out mean R&R rate of the editor assigned to the paper, included as an instrumental variable for the control function.

A comparison of the estimates in our baseline R&R model (column 7) and our baseline citation

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

Given that papers by all-female teams receive 0.22 log points more citations (Column 4), under a proportional decision model we would expect a coefficient of $0.55 = 0.22 \times 2.5$ in the R&R probit model, as specified by equation (8). A coefficient of this size would offset the bias in the referee recommendations and ensure that female-authored papers are evaluated in accord with their “quality”, as revealed by citations. The actual coefficient on the all-female papers in the R&R decision in column 7, however, is 0.01 (s.e.= 0.06). Thus, as suggested by the graphical comparisons in Figure IV, editors do not *undo* the referee’s apparent biases. The coefficient is precisely estimated, such that we can confidently reject the hypothesis of a value of 0.55 under the citation maximizing benchmark.

Note that equation (8) is derived under the assumption of equal citations for male and female papers of equal quality. Since most economists in our survey believe that female-authored papers receive if anything fewer citations for their work, equation (8) is actually a lower-bound.

It is useful to draw a parallel to the case of the author publication variables, considered by CDV. Similar to the case of all-female authorship, having a well-published coauthor has a much smaller effect on the editor’s R&R decision than would be expected under citation-maximizing behavior. For example, papers by authors with 6+ publications have an R&R probit weight of 0.41 (Column 7), compared to a predicted weight of $2.48 = 0.99 \times 2.5$. That is, author publications are underweighted by a factor of about 5; we cannot reject that all-female authorship is similarly underweighted by a factor of 5. A key difference, however, is in the interpretation. Authors with more publications could plausibly receive more citations for given quality of their work, reflecting more extensive networks and other factors. In the case of author gender, however, it seems implausible, at least to our survey responders, that female authors receive more citations than males for given quality.

Implications. How large is the impact of the non-proportionality of author gender in our citation and R&R models? The baseline R&R rate for all-female papers is 12.2 percent in the sample of NDR papers, shown in the first bar of Figure V. We now simulate counterfactual R&R rates under two citation-maximizing counterfactuals.

In the first counterfactual, the editor corrects the deviation from citation maximization with respect to the author gender variables, but *not* with respect to the measures of author publications or other variables such as field. This is the relevant counterfactual if the editor believes that these other characteristics potentially affect citations, *conditional* on paper quality — for example, if citations to highly published authors are inflated. It is also the relevant counterfactual if, alternatively, there is no systematic bias in citations relative to quality, but editors set a higher bar for

papers by authors with more publications as part of their editorial policy.

To compute the counterfactual, we start from the predicted R&R probability for the model in Column 7 of Table IV, $\Phi(x\hat{\pi})$, which matches the R&R rate of 12.2% in the subsample of female-authored papers.²⁶ For all-female papers we then add the correction factor $\hat{\delta} = \hat{\pi}_F^{cite-max} - \hat{\pi}_F = 0.22 * 2.5 - 0.01 = 0.54$ and compute $\Phi(x\hat{\pi} + F\hat{\delta})$. We add similar correction factors for the two sets of mixed gender papers. Since the resultant counterfactual R&R rates would lead to a change in the overall number of R&Rs, we adjust the journal×submission year effects in the R&R probit models to match the actual R&R rate (for NDR papers) in each journal-year cell.

As the second set of bars in Figure V shows, the R&R rate for all-female papers would rise from 12.2% to 18.5% – a 50% increase. The counterfactual would have little effect instead for mixed-author papers. Overall, the R&R for an average female economist who submits a paper to the journals in our sample, taking into account that some of her papers will be in the mixed-gender category and some in the all-female category, would increase from 14.3% to 15.9%, an 11% increase. This effect is large enough to potentially matter in career advancement and pay decisions.

In the second counterfactual, captured by the third set of bars in Figure V, we assume that the editor aims to maximize citations not just with respect to the author gender mix, but also with respect to the author publication variables. The R&R rate for all-female papers would remain roughly constant: these papers get a boost in R&R rates because of their gender, but they also suffer a relative reduction because female authors tend to have fewer prior publications, and the increased rate of R&R for highly published authors necessitates a raising of the overall bar for R&R status. Mixed-gender papers would benefit from the author publication correction, such that overall the R&R rate for papers by female economists would rise from 14.3% to 16.0%.

Robustness and Heterogeneity. In Online Appendix Tables 5 and 7-8 and in Table V we consider the robustness and heterogeneity of the R&R probit results, along the same dimensions considered for the citation models. Across virtually all specifications we replicate the key finding that editors fail to significantly upweight female-authored papers, despite the fact that female-authored papers get more citations. To illustrate this, in columns 7 and 8 of Table V we compare the actual R&R rate for female-authored papers in the particular sample or specification, and the counterfactual R&R rate if editors weighted female authorship in accord with its effect on citations. Across all specifications, the counterfactual would increase the R&R rate for female-authored papers, though the increase is larger in some cases (e.g., papers by author teams with at least 3+ prior publications) and smaller in others (e.g., papers by a single author).

5.2 An Analysis of Published Papers

So far, we have shown that all-female papers get more citations than all-male papers, conditional on the referee assessment, and that editors do not correct for this apparent bias in the referee

²⁶In a slight abuse of notation we add the leave-out mean R&R rate of the editor assigned to the paper as a component of the vector x . We also note that a probit model does not necessarily predict the sample mean for a subgroup with an included dummy: in our case however, the deviation is negligible.

assessments. There are two leading explanations for these results. The first is that referees and editors impose a higher bar on papers by female authors, perhaps because of stereotyping biases that lead them to downweight the quality of work by female authors. The second is that female-authored papers have characteristics that lead to higher citations but are not as highly rewarded in the review process. Examples could include working in particular sub-fields, or writing empirical, as opposed to theoretical, papers within a field. Given that we cannot adequately measure these characteristics in our main sample, it is difficult to separate out the two explanations.

To provide some evidence on the second explanation, we collected a secondary sample of all the published articles in the four journals of our sample for the years 2008-15. These 1,719 papers broadly correspond to the accepted papers in our sample, assuming a 2-year delay between submissions and publication. Although we have no information on the referee assessments of these papers, we are able to measure the gender and prior publications of each author (assuming the papers were submitted 2 years prior to publication). We also collected Google Scholar citations as of March 2019 for each papers, using the same procedures as for the main sample.

A key advantage of this sample is that we can measure a variety of features of each paper that are unavailable in our main sample, including detailed JEL codes. We are also able to study the measurement errors in citation counts arising from problems in our scraping procedure (associated with special characters in author’s names and titles) and to estimate the misclassification rates of author gender. The key disadvantage is that it is about one tenth the size of our main sample, and it does not allow us to disentangle the role of referees and editors.

In Table VI, we first estimate our citation model (excluding referee evaluations) on the sample of *accepted* papers from our main EE database. We then repeat the same specification using published papers (column 2). The two samples are not identical (1,713 accepted papers versus 1,719 published paper) and have different citation measures (GS citations as of mid-2015 versus March 2019), but they have similar characteristics (Appendix Table 9). The estimated coefficients in the two samples are also similar. For the accepted papers in our main sample, the estimated female-author effect is 0.24 log points (s.e.=0.13), virtually identical to the estimate obtained in our overall sample when we exclude the referee variables (see column 2 of Table IV), though less precisely estimated. In the sample of published papers the female author citation premium is somewhat smaller at 12 log points (s.e.=0.11), but qualitatively similar. We also find a 24 log points (s.e.=0.13) citation premium for mixed-gender papers with a senior female coauthor in the published paper sample. These estimates replicate the key qualitative patterns in our main findings, and are qualitatively consistent with the findings of Hengel (2019) for a different sample of published papers.

We then add additional controls for paper features that are unavailable in the main sample. First, in Column 3 we include very detailed 2-digit JEL field controls (e.g., for J28), to assess the potential role of sub-field differences beyond our broad field controls. These controls nearly double the R^2 in the citation regression, but barely affect the gender coefficients. This concurs with the evidence above suggesting that sub-field differences are unlikely to explain the findings.

Next, we include two sets of measures of theoretical, as opposed to empirical, content. These are

important additions because the JEL codes do not, by design, measure this dimension. To construct a first set of measures, we search the text of each published paper for words indicating theoretical content (such as “theorem” and “proposition”), empirical content (such as “standard error” and “table”), structural estimation (such as “structural” and “simulation”), or experimental component (such as “RCT” and “laboratory”). In Column 4, we add the asinh of the count of words in each of these four categories. The presence of theoretical material has a substantial negative predictive power on citations, and conversely for the presence of empirical results. Introducing these controls reduces the estimated author gender premia, reflecting the fact that papers by female authors and by mixed gender teams tend to have more empirical content.

The second set of measures are derived from a set of variables coded by a team of research assistants (who were not informed of our interest in gender-related differences). As shown in Column 5, these measures also show a negative citation effect for theoretical content and a positive effect for empirical content, and lead to somewhat smaller author gender premia. Adding both sets of variables (Column 6) leads to a citation premium of 9 log points (s.e.=0.13) for all-female authors, and 13 log points (s.e.=0.15) for mixed-gender papers with senior female authors.

In Column 7, we estimate a Lasso regression (using 10-fold cross validation), motivated by the fact that the specification on Column 6 has hundreds of sub-field controls and a dozen controls for other characteristics. The results are quite similar to the ones in Column 6, alleviating concerns about potential over-fitting of the latter model.

Papers by different author gender teams could also differ in the patterns of accrual of citations. While we cannot conduct such analysis with Google Scholar citations, in Online Appendix Table 10 we provide evidence using SSCI citation measures (which provide a full list of the citing articles). Specifically, we estimate models for the asinh of SSCI citations accumulated 0-1 years after publication, 2-3 years after publication, and 4-5 years after publication. We find no obvious pattern, with similar author-gender premia at 0-1 years out and at 4-5 years out, and smaller effects 2-3 years out. We also estimate models that separate out citations by the impact factor of the citing journals. We find slightly smaller gender premia using citations from high-impact factor journals, very small premia for citations at low-impact-factor journals, and relatively large premia associated with citations from journals with no known impact factor. While these results show no obvious pattern, we emphasize that the estimated gender premia are relatively imprecise.

Overall, the results on empirical versus theoretical content provide suggestive evidence that different characteristics of papers play some role in the observed citation premium, but would probably not explain all the citation premium in our main sample. Nevertheless, this conclusion is tentative, given the difficulty of making inferences from just published papers.

Measurement Error. We hand-checked the citation counts and author gender for all papers in this sample. In terms of citations, we find that our automated scraper gave the correct citation count for 96% of papers (see Online Appendix Table 11). In 31 cases (1.8% of the sample) the title of the published article as recorded by Econlit had special characters, which led the scraper to find zero citations. Similarly, in 27 cases (1.6%) the author name(s) contained special characters (e.g.,

Jordi Galí), leading our automated scraping procedure to find 0 citations. The opposite type of error— finding too many citations—occurs in only 7 cases. Finally, with respect to gender coding, we found only 5 cases of incorrect coding of the author gender mix (0.3% of the sample), with a dozen additional cases in which we were able to assign gender to “undetermined-gender” cases.

The possibly consequential errors are the ones in which a paper is incorrectly recorded with 0 citations. If hypothetically, this error was less common for female economists, it could bias upward the measured citations of all-female papers in the main sample, potentially accounting for our key result. We see no reason *ex ante* why this error would correlate with author gender, given that it depends on special characters in the paper title or last name. Nonetheless, we provide two pieces of evidence. First, in Table V row 4 we run the main specification excluding papers with 0 citations and find similar results. Second, we measure whether the occurrence of papers with 0 citations differs by author gender, compared to the case of positive, but low, citations (defined as papers in the 35-55th percentile of citations in a journal-year). As Online Appendix Figure 7 shows, there is no obvious pattern by gender group and, to the extent that there is one, the all-female group has slightly *more* papers with zero citations, which would bias the citation measure downward.

5.3 Desk-Rejection

One interpretation of our results so far is that editors defer to the referees with respect to their R&R decisions. To provide evidence on editorial preferences with no input from referees, we turn to the desk-rejection decisions. Using the full sample of 29,890 submissions, we compare predictors of citations with predictors of the decision to not desk reject (NDR) in Table VII. At the submission stage, female-authored papers have 24 log points higher citations than submissions by all-male authors, holding constant other paper and author characteristics (Column 1). This mirrors the citation result in the R&R regression in Table IV.

As Column 2 shows, in the NDR decision editors *do* take into account the author gender and are more likely to not desk-reject papers by female authors, holding all else constant. This specification, though, does not tell us whether this is the optimal weight, given that we do not observe a variable like the referee recommendations.

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

responding to the quality difference between female-authored papers and male-authored papers.

5.4 Weight Placed on Referee Recommendations

We now consider how editors use the information provided by male versus female referees. CDV find that the recommendations of more- and less-published referees are equally informative about the quality of papers (as measured by future citations). Yet, editors tend to place more weight on the recommendations of more published referees. Is there a similar difference by gender?

Figure VI Panel A shows the informativeness of referees (i.e. the relationship between referee recommendations and citations), by referee gender and prior publication record. (Referees with 3 or more recent publications are classified as “prominent”). The crucial element is the *slope* of these lines. Male and female referees do not seem to differ in their informativeness and, consistent with results from CDV, more and less prominent referees do not differ in their informativeness either. Figure VI Panel B shows the relative weight given by editors to recommendations from referees who differ in prominence and gender. Consistent with the results in CDV, more prominent referees are valued more, but gender does not seem to have much of an effect.

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

$$Outcome_i = \sum_{j=1}^{N_{referees,i}} (\alpha_0 Female_{ij} + (1 + \alpha_1 Female_{ij}) R_{ij}) / N_{referees,i} + \gamma \mathbf{X}_i + \varepsilon_i$$

where $Outcome_i$ denotes paper i ’s outcome ($asinh(citations)$ or receiving an R&R decision), $Female_{ij}$ is an indicator for the gender of referee j of paper i , $N_{referees,i}$ is the number of referees who evaluate paper i , and R_{ij} denotes an index of recommendations $R_{ij} = \beta_{DefReject} DefReject_{ij} + \dots + \beta_{Accept} Accept_{ij}$, with the same coefficients β_c for each recommendation type, regardless of the gender of the referee (or of the author team). Such a specification adjusts the index R_{ij} for both a gender-specific intercept and slope. If for example, female referees are more positive, we would expect a negative value for α_0 in the citation regression. If the recommendations of female referees are more informative, then we expect $\alpha_1 > 1$ in the citation regression.

Table VIII shows the results. Columns 1-3 have informativeness as measured by citations, while Columns 4-6 report models of the editor’s R&R decision. Consistent with CDV, more prominent referees do not provide more informed recommendations (compared to less prominent referees), but nevertheless, the editors do value them more. Turning to gender, female and male referees do not differ in their informativeness, and editors do not put different value on them either. This gender result is consistent with the expectations of the survey respondents.

6 Delays and Other Outcomes

We now turn to the speed of decision-making. If there are gender differences in the editorial process, they may appear in the form of decision delays.

In Columns 1-2 of Table IX we consider referee-paper observations and estimate the number of days from paper submission to the report completion for reviewers who return a report. With controls for paper and author characteristics (Column 1), the response time of referees is not affected by the gender composition of the authors. In Column 2, we add paper fixed effects, fully controlling for possible differences in papers assigned to male or female referees, which may affect the time to report completion. We find that female referees are neither faster nor slower than male referees, a precise null effect. Further, there is no interaction with the gender composition of the authors.

To integrate the referee delays with the editor delays, in Columns 3-4 we consider decision times at the paper level, with the full set of controls as well as editor fixed effects. We detect no difference based on the author gender mix on the number of days from paper submission to the arrival of the last report (Column 3), or on the full decision time including editor delays (Column 4).

For the papers with an initial R&R that are ultimately accepted, we consider the impact of the author gender mix on the number of rounds of submissions (Column 5) and on the total time from submission to acceptance (Column 6), as well as the time that the authors take to submit the first revision and the number of days from the resubmission until the final acceptance (Online Appendix Table 12). Across all these specifications, we find no difference by the author gender mix.²⁷ Thus, unlike in the analysis of Hengel (2018), referee and editorial delays appear to be gender-neutral, a pattern also visible in Online Appendix Figures 8a-c.

We consider two further outcomes. In Online Appendix Table 13 we consider the propensity of referees to accept a referee invitation. Each observation is a referee request for papers that were not desk-rejected. Once we include controls (Column 2), we estimate no difference in the probability of accepting a referee invitation for all-female-authored papers. With paper fixed effects (Columns 3), we examine if female referee are more likely to provide public goods by agreeing to referee, but we find no such difference. Finally, we find no relative difference depending on whether the reviewer gender matches the author gender (Columns 3 and 4).

Finally, motivated by Hengel (2018)’s analysis, we examine in Online Appendix Table 14 the complexity level of the abstracts. An important caveat is that we only observe the abstract of the most recent version of the paper, and thus we cannot examine, as in Hengel (2018), the change in complexity from the submission to the published version of the paper; also, Hengel (2018) presents a more in-depth analysis of abstract complexity. We simply compare the Gunning Fog (Columns 1 and 3) and Coleman-Liau (Columns 2 and 4) measures of complexity for papers with different author gender teams. We do so separately for the papers that were desk rejected or rejected (Columns 1-2), and the papers that received an R&R decision (Columns 3-4). We find no impact of the author gender mix on the readability of the abstract in either sample.

²⁷We find very similar results if we do not control for the referee recommendation variables. The finding that there is no difference in rounds of revision by author gender also provides suggestive evidence against an alternative interpretation of the citation premium for all-female papers—that male and female papers may be the same quality at the time they are read by referees, but female teams are more responsive to reports and thus their papers may eventually be better. This would presumably be reflected in shorter revisions.

7 Discussion and Conclusion

Are the referees and editors in economics gender neutral? The answer is both “Yes” and “No”.

If we focus on comparisons highlighted in previous research, we obtain relatively precise zero differences between gender groups. Considering referee recommendations, we replicate the findings of Abrevaya and Hamermesh (2012) that there are no difference in how referees of different genders assess papers by female and male authors. Considering delays in publication, as Hengel (2018) does, we find no differences by author gender. Turning to editorial decisions, which have not been previously studied, we find that editors are gender-blind in the sense that they treat female- and male-authored papers the same, conditional on the referee recommendations. Further, editors give about the same weight to recommendations of male and female referees, which is appropriate given that the two groups are equally informative.

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

What accounts for these patterns? We consider two main explanations. The first is that referees hold female authors to a higher bar, perhaps because of stereotype biases. The second is that female-authored papers have characteristics that lead to higher citations but are not as highly rewarded in the review process, such as being in particular sub-fields or having empirical content.

We turn to a set of published papers, broadly corresponding to the accepted papers in our sample, to further investigate this second explanation. The evidence suggests that sub-field differences are unlikely to play a role, but that differences in empirical content could play some role in the gender citation premium, though they would be unlikely to explain the full citation premium in our main sample. Nevertheless, this conclusion is tentative, given the difficulty of making inferences from just published papers. Future research with even more detailed information on paper characteristics for submitted papers will help further separate out these explanations.

Where does that leave us in terms of implications? Our finding that female authors are held to a higher standard is concerning. We estimate that as a result the R&R rate for female-authored papers is about 6 percentage points lower than the rate consistent with a gender-neutral citation-maximizing rule. This gap suggests an important hurdle, aside from the assignment of credit in coauthored work stressed by Sarsons (2018), for junior female economists, as well as a continuing obstacle to career progression for more senior female researchers.

One potential remedy to help female economists – using more female referees – is unlikely to help, given that female referees hold female-authored papers to the same higher bar as male

referees. Recruiting more female referees may only have the unintended consequence of requiring more public good provision by female economists (Babcock et al., 2017).

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

It would be great to revisit our analysis in 3 to 5 years to test whether the gender difference has been corrected. Perhaps, as for NBA referee bias (Price and Wolfers, 2010), publicizing the findings may be enough to correct the pattern (Pope, Price and Wolfers, 2018).

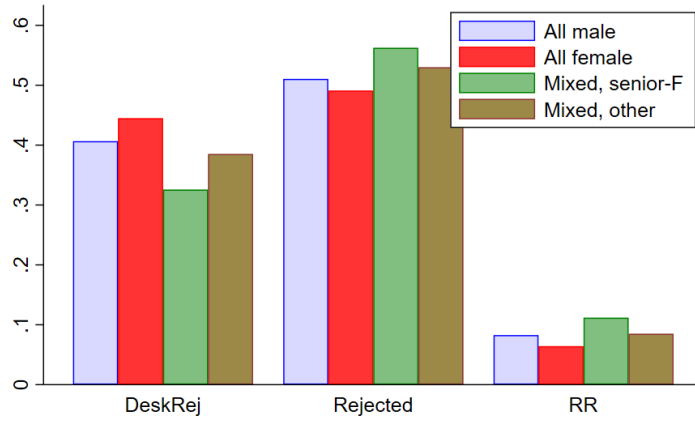
This is just an example of the importance of data transparency in the editorial process, as CDV also stress. Indeed, we are grateful to the four journals in economics which agreed to such data access, something with very few parallels outside economics. The ability to systematically keep track of, and analyze, referee and editorial choices should make it relatively straightforward in the future to check for progress on any form of gender bias, especially if journals were to keep track of the gender of authors and referees in the editorial system. More generally, data transparency and access will help make the editorial process fairer and more efficient.

References

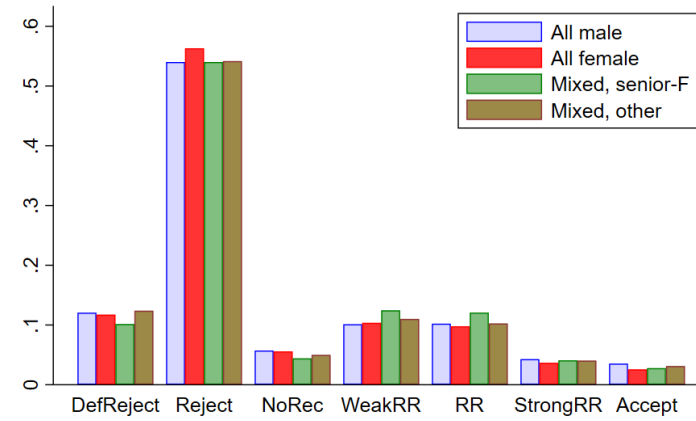
- Abrevaya, Jason and Daniel S. Hamermesh. 2012. "Charity and Favoritism in the Field: Are Female Economists Nicer (To Each Other)?" *Review of Economics and Statistics*, 94(1): 202-207.
- Altonji, Joseph G., Todd E. Elder, and Christopher R. Taber. 2005. "Selection on Observed and Unobserved Variables: Assessing the Effectiveness of Catholic Schools." *Journal of Political Economy*, 113(1), 151-184.
- Azmat, Ghazala, and Rosa Ferrer. "Gender gaps in performance: Evidence from young lawyers." *Journal of Political Economy* 125.5 (2017): 1306-1355.
- Babcock, Linda, Maria Recalde, Lise Vesterlund, and Laurie Weingart. 2017. "Gender Differences in Accepting and Receiving Requests for Tasks with Low Promotability," *American Economic Review*, 107(3): 714-747.
- Bayer, Amanda, and Cecilia Elena Rouse. 2016. "Diversity in the Economics Profession: A New Attack on an Old Problem." *Journal of Economic Perspectives*, 30(4): 221-42.
- Bellemare, Marc F., and Casey J. Wichman. Forthcoming. "Elasticities and the Inverse Hyperbolic Sine Transformation," *Oxford Bulletin of Economics and Statistics*.
- Bertrand, Marianne, and Kevin F. Hallock. 2001. "The gender gap in top corporate jobs." *Industrial and Labor Relations Review* 55.1: 3-21.
- Blank, R.M., 1991. "The effects of double-blind versus single-blind reviewing: Experimental evidence from the American Economic Review". *The American Economic Review*, 1041-1067.
- Bohren, J. Aislinn, Alex Imas, and Michael Rosenberg. Forthcoming. "The Dynamics of Discrimination: Theory and Evidence," *American Economic Review*.
- Bordalo, P., Katherine Coffman, and Nicola Genaioli. Forthcoming. "Beliefs about gender". *American Economic Review*.
- Bransch, F., and M. Kvasnicka, 2017, "Male Gatekeepers Gender Bias in the Publishing Process?". IZA DP No. 11089.
- Broder, I.E., 1993. "Review of NSF economics proposals: Gender and institutional patterns" *The American Economic Review*, 83(4), 964-970.
- Card, David and Stefano DellaVigna. Forthcoming. "What do Editors Maximize? Evidence from Four Economics Journals." *Review of Economics and Statistics*.
- Ceci, Stephen J., Donna K. Ginther, Shulamit Kahn, and Wendy M. Williams. 2014. "Women in Academic Science: A Changing Landscape." *Psychological Science in the Public Interest*, 15(3): 75-141.
- Chari, Anusha and Paul Goldsmith-Pinkham. 2017. "Gender Representation in Economics Across Topics and Time: Evidence from the NBER Summer Institute." NBER Working Paper No. 23953.
- Christensen, Garret S., and Edward Miguel. 2018. "Transparency, Reproducibility, and the Credibility of Economics Research." *Journal of Economic Literature*, Vol. 56, 920-980.

- DellaVigna Stefano and Devin Pope. 2018. "What Motivates Effort? Evidence and Expert Forecasts." *Review of Economic Studies*, 85, 1029–1069.
- Dolado, Juan J., Florentino Felgueroso, and Miguel Almunia. "Are men and women-economists evenly distributed across research fields? Some new empirical evidence." *SERIEs* 3.3 (2012): 367-393.
- Donald, Stephen, G., and Daniel S. Hamermesh. 2006. "What is Discrimination? Gender in the *American Economic Association*, 1935-2004." *American Economic Review*, 96 (4): 1283-1292.
- Ginther, Donna, K., and Shulamit Kahn. 2004. "Women in Economics: Moving Up or Falling Off the Academic Career Ladder?" *Journal of Economic Perspectives*, 18 (3): 193-214.
- Goldin, Claudia. "A grand gender convergence: Its last chapter." *American Economic Review* 104.4 (2014): 1091-1119.
- Hengel, Erin. 2018. "Publishing while Female. Are Women Held to Higher Standards? Evidence from Peer Review," Working paper.
- Hengel, Erin. 2019. "Gender Differences in Citations at Top Economics Journals," Working paper.
- Lundberg, Shelly. "Report: Committee on the Status of Women in the Economics Profession (CSWEP)." *American Economic Review* 107.5 (2017): 759-76.
- Niederle, Muriel, and Lise Vesterlund. 2011. "Gender and competition." *Annual Review of Economics*, 3.1: 601-630.
- Pope, Devin, Joseph Price, and Justin Wolfers. 2018. "Awareness Reduces Racial Bias", *Management Science*, 64(11): 4967-5460.
- Price, Joseph and Justin Wolfers. 2010. "Racial Discrimination Among NBA Referees," *Quarterly Journal of Economics*, 125(4):1859-1887.
- Reuben, Ernesto, Paola Sapienza, and Luigi Zingales. 2014. "How stereotypes impair women's careers in science." *Proceedings of the National Academy of Sciences*: 201314788.
- Reuben, Ernesto, Paola Sapienza, and Luigi Zingales. 2015. "Taste for competition and the gender gap among young business professionals." No. w21695. National Bureau of Economic Research.
- Sarsons, Heather. 2018. "Gender Differences in Recognition for Group Work." Working Paper.
- Wu, Alice. Forthcoming. "Gender Stereotyping in Academia: Evidence from Economics Job Market Rumors Forum." *Review of Economics and Statistics*.

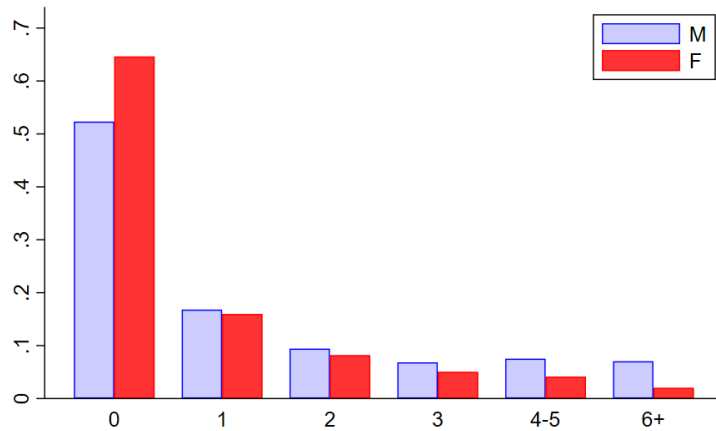
Panel A. Distribution of Editorial Decisions



Panel B. Distribution of Ref. Recommendations



Panel C. Distribution of Author Publications



Panel D. Distribution of Referee Publications

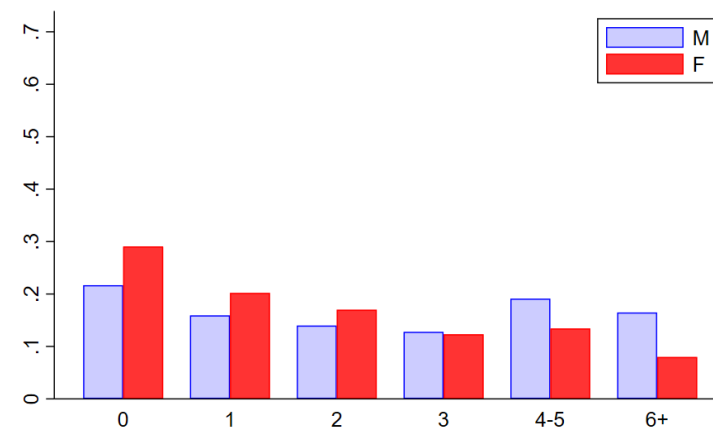


Figure I
Summary Statistics by Gender

Notes. This figure displays a few key summary statistics by gender. Panel A plots the distribution of the editor's decision and Panel B shows the distribution of referee recommendations. Panel C 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. Panel D reports the distribution of publications among referees by gender.

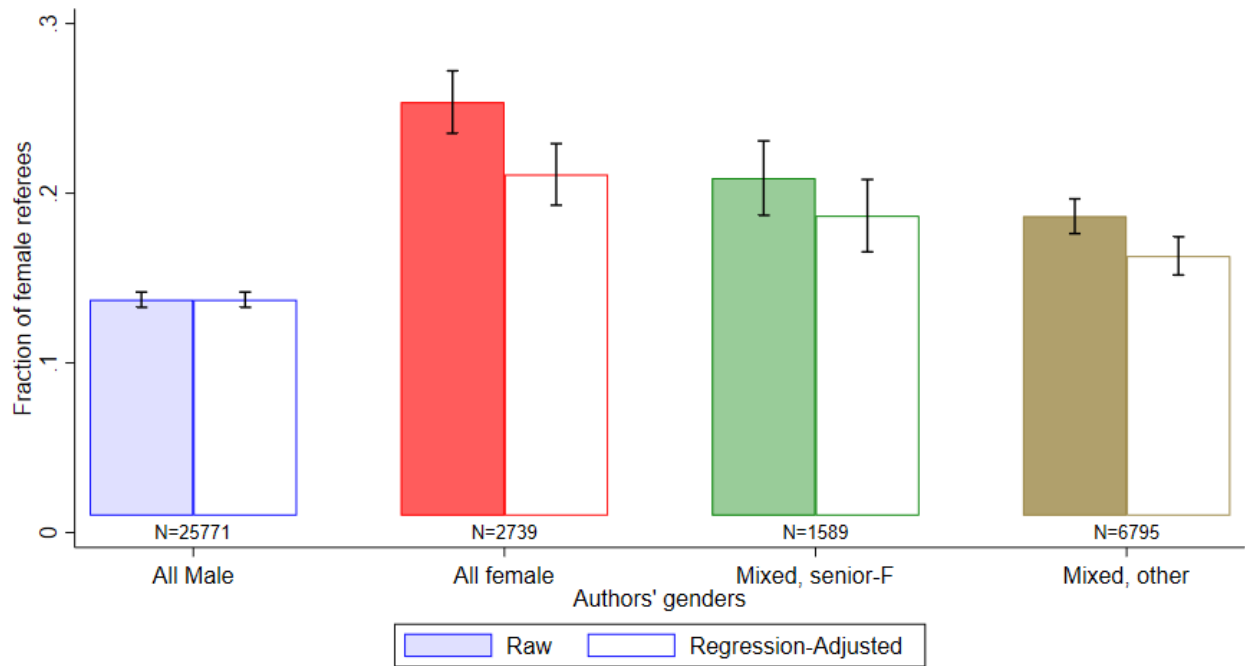
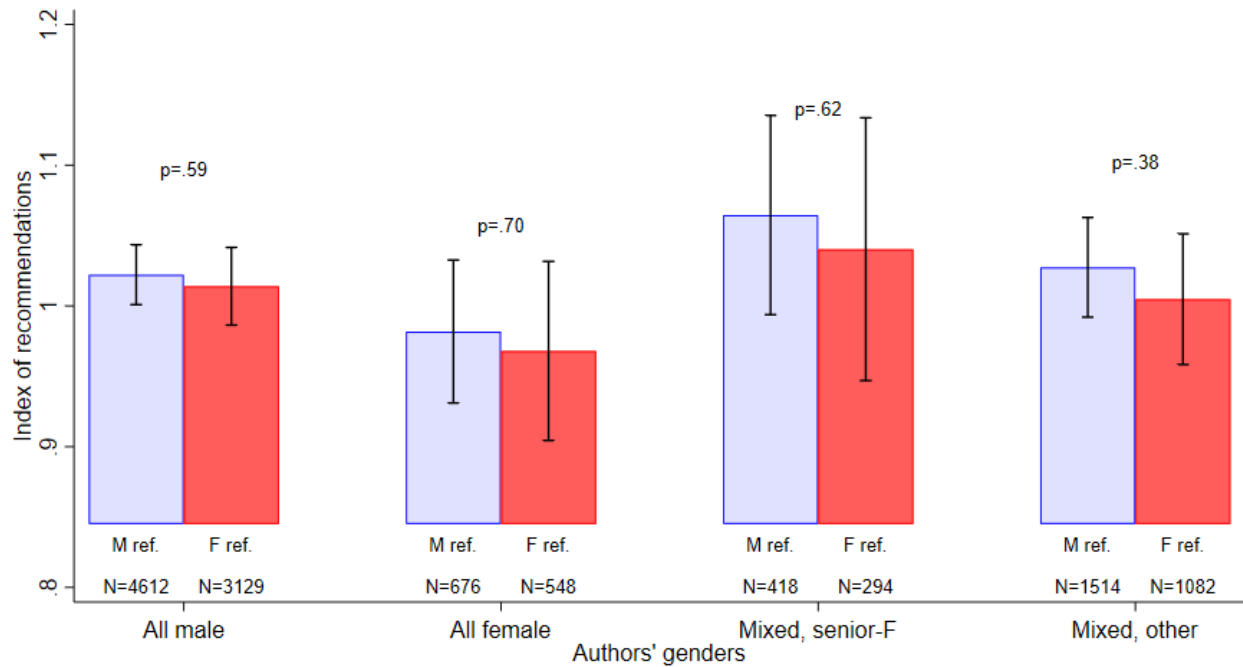


Figure II.

Referee Assignment: Referee Gender as Function of Author Gender

Notes. Observations are at the referee-paper level. In the regression-adjusted bars, we adjust using the results in Online Appendix Table 3, column 3.

Panel A. Index of Referee Recommendations



Panel B. Share of Positive Referee Recommendations

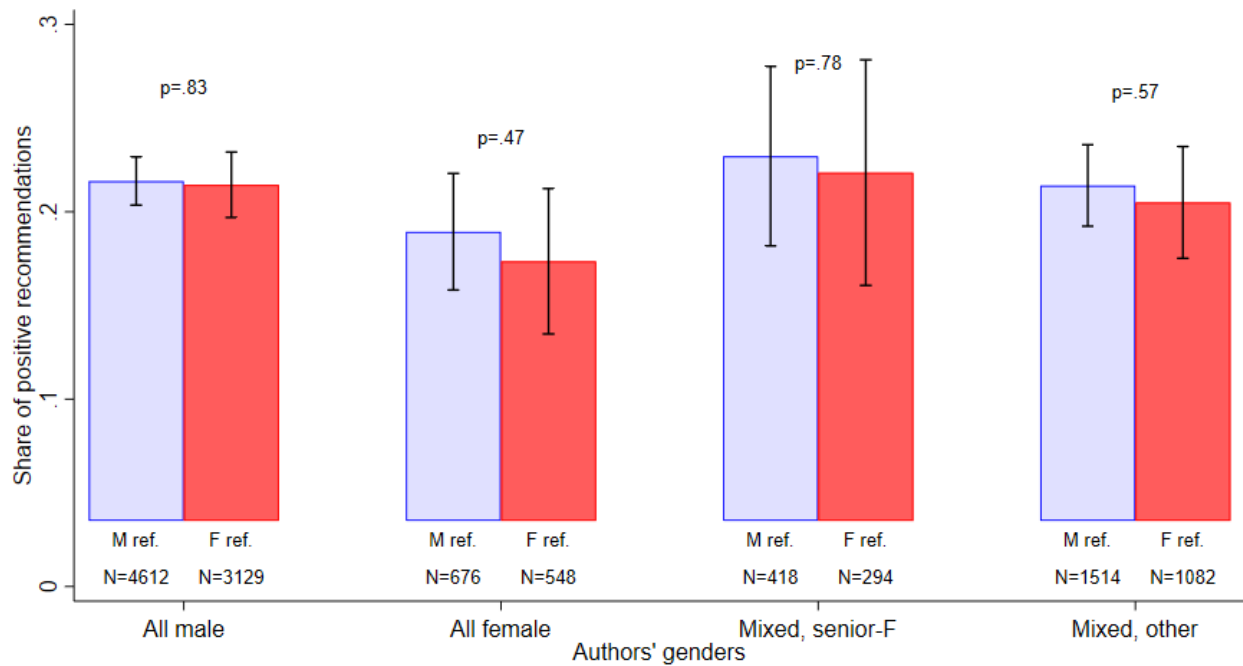
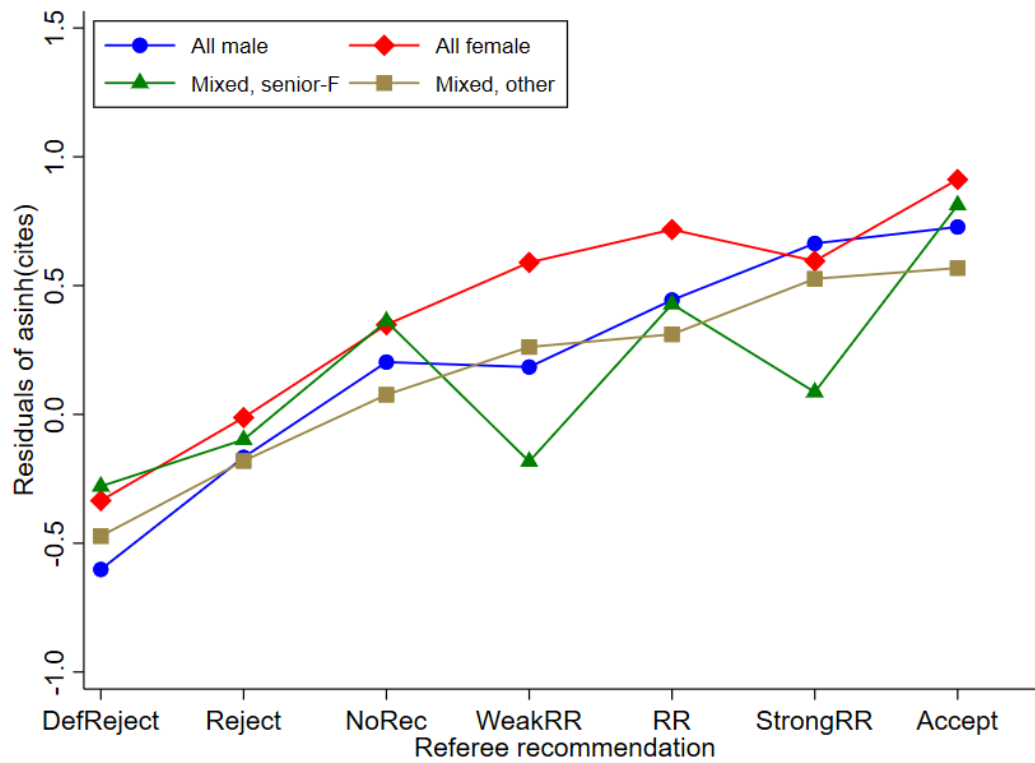


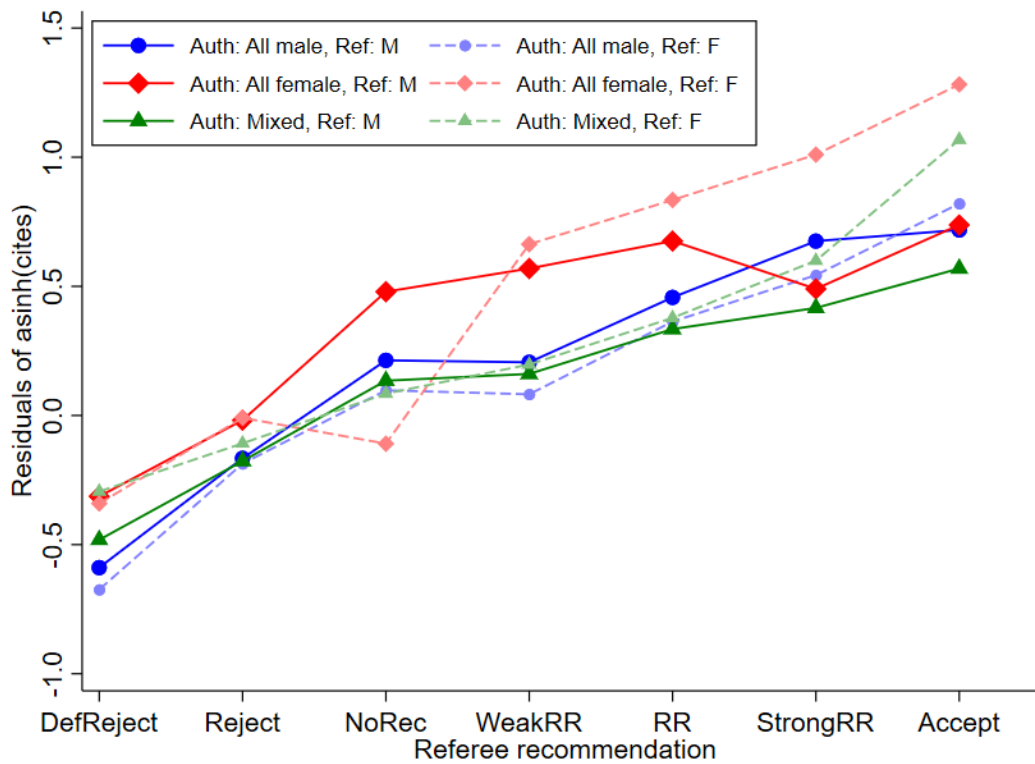
Figure III
Referee Evaluation by Author Gender and Referee Gender

Notes. Panel A 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 (forthcoming). From Definitely Reject to Accept, the values are 0, 0.67, 1.01, 1.47, 1.92, 2.27, 2.33. The bands show 2 standard error intervals, clustered at the paper level. Includes only 4,341 papers with both male and female referees. Panel B shows the share of positive recommendations, defined as RR-Accept. In both panels, female referees are weighted at the paper level by $N_{\text{male}} / N_{\text{female}}$.

Panel A. Referee Recommendations and Citations



Panel B. Referee Recommendations and Citations, by Author Gender and Referee Gender



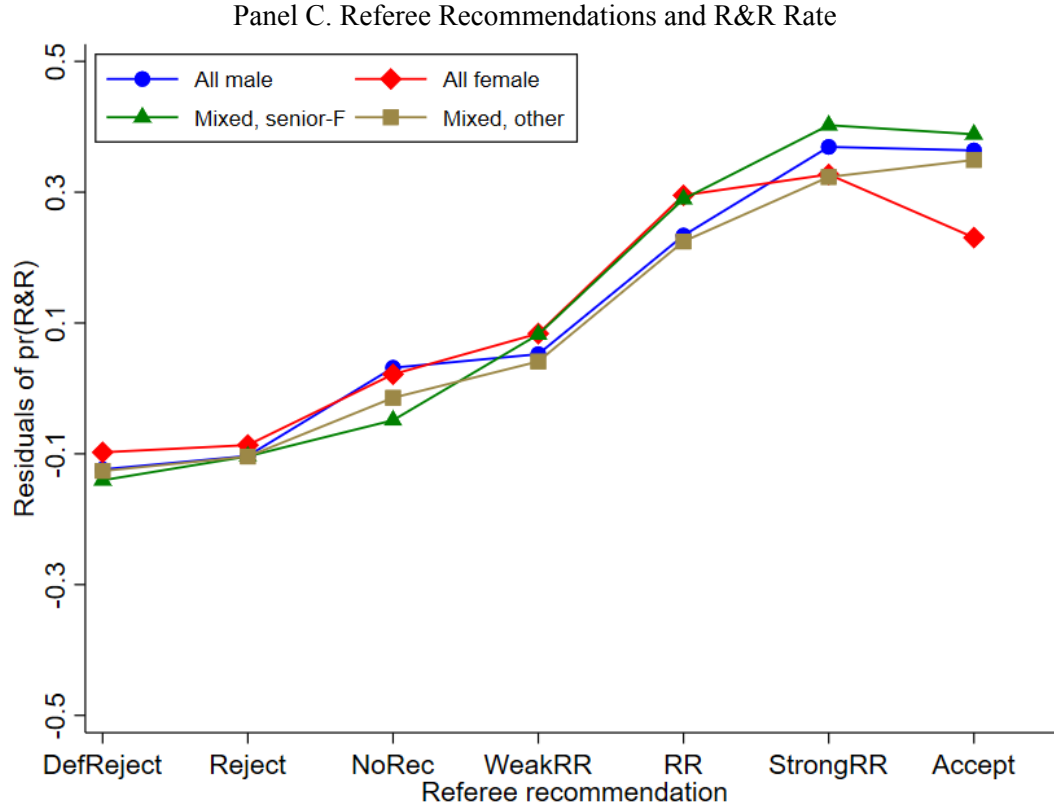


Figure IV

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

Notes. Panels A and B show the weighted residuals of *asinh* (citations) for a paper receiving a given recommendation, while Panel C shows the residuals of $pr(R\&R)$ for a paper receiving a given recommendation. In both models, residuals are calculated by regression onto author publications, the number of authors, and fields. Panels A and C show the results separately by the gender break down of the author team. Panel B 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. Panel B omits confidence intervals for legibility.

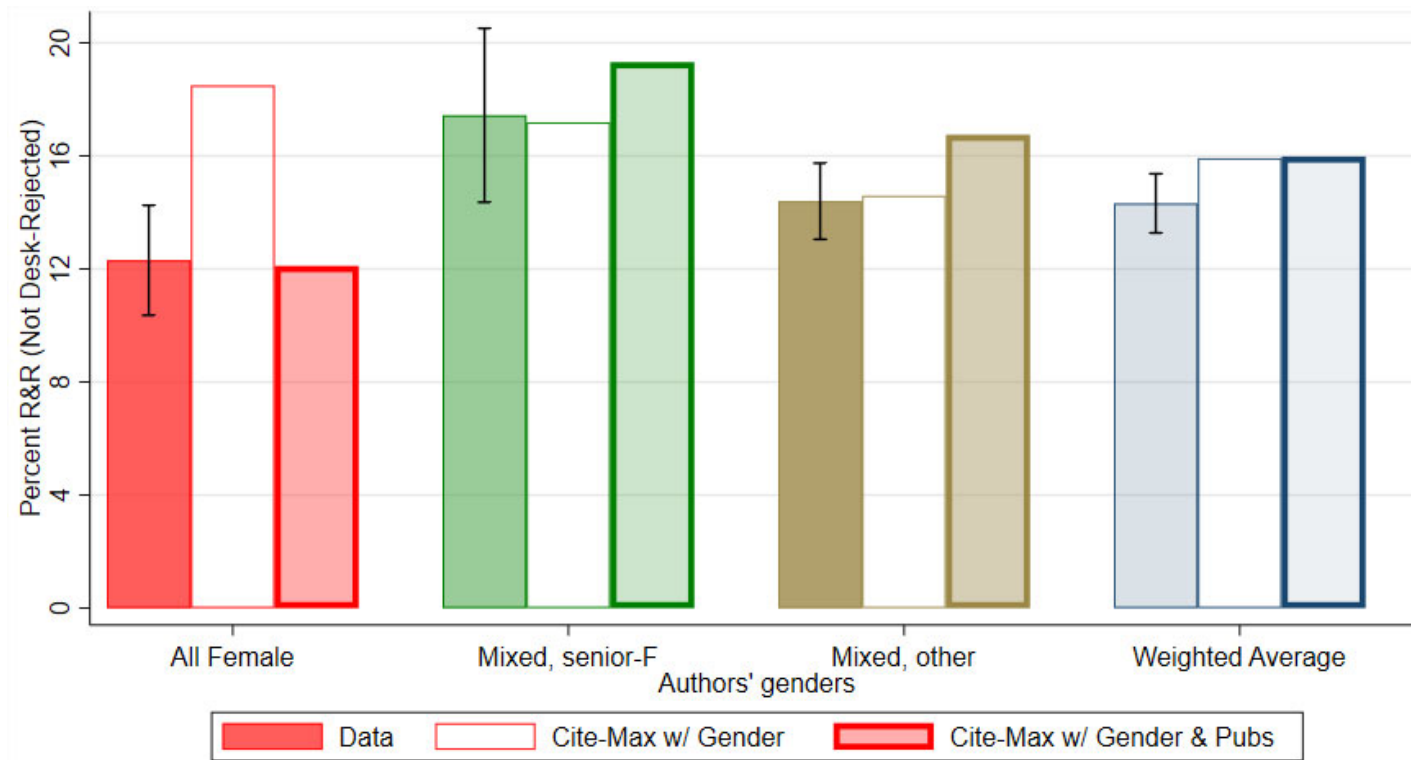


Figure V
Implication of Findings for R&R Rate for Female-Authored Papers

Notes. This figure presents the implications of the results in Table IV for the actual and counterfactual R&R rate for teams of authors. The figure breaks down papers by all-female authors, by mixed-gender teams with a senior female author and other mixed-gender teams. Within each group, the first bar plots the observed R&R rate for that group, conditional on a paper not being desk-rejected. The second bar reports the counterfactual R&R rate for that category of papers that we would expect to observe if the editorial process aimed to put the weight on the gender-author mix associated with citation maximization with respect to the author-gender mix variables (the details are in the text), keeping the R&R rate constant in each journal-year cell. The third bar reports the counterfactual R&R rate if the editorial process put the citation-maximizing weight not only on the gender-author mix variables, but also on the author past publication variables. The final set of columns presents an average over the first 3 sets of columns, weighting the different groups by the number of papers in each category.

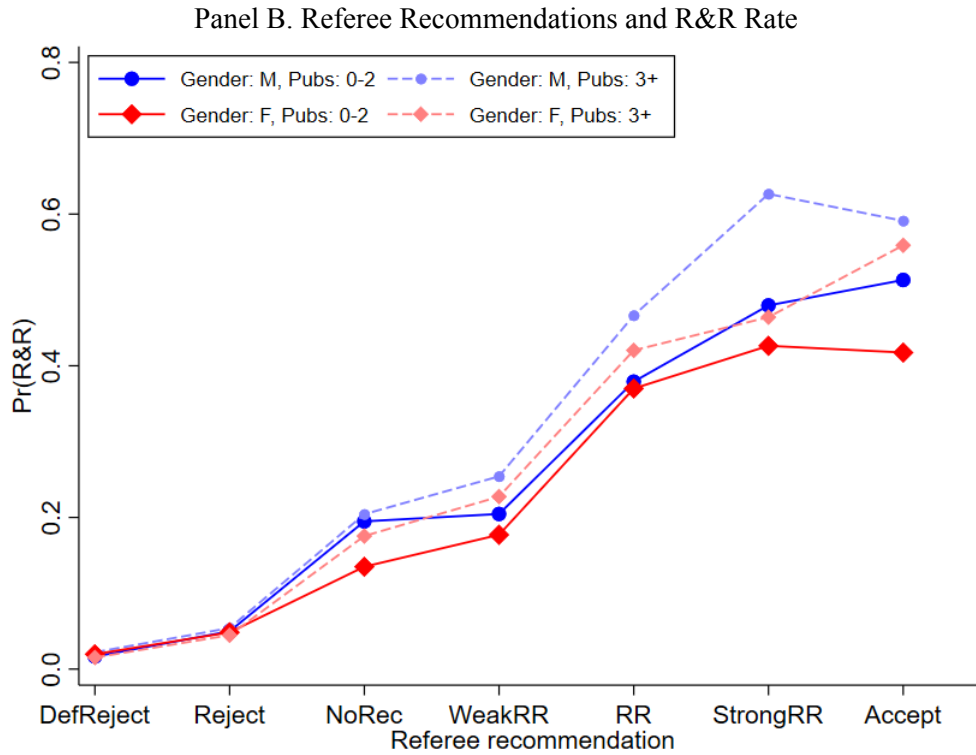
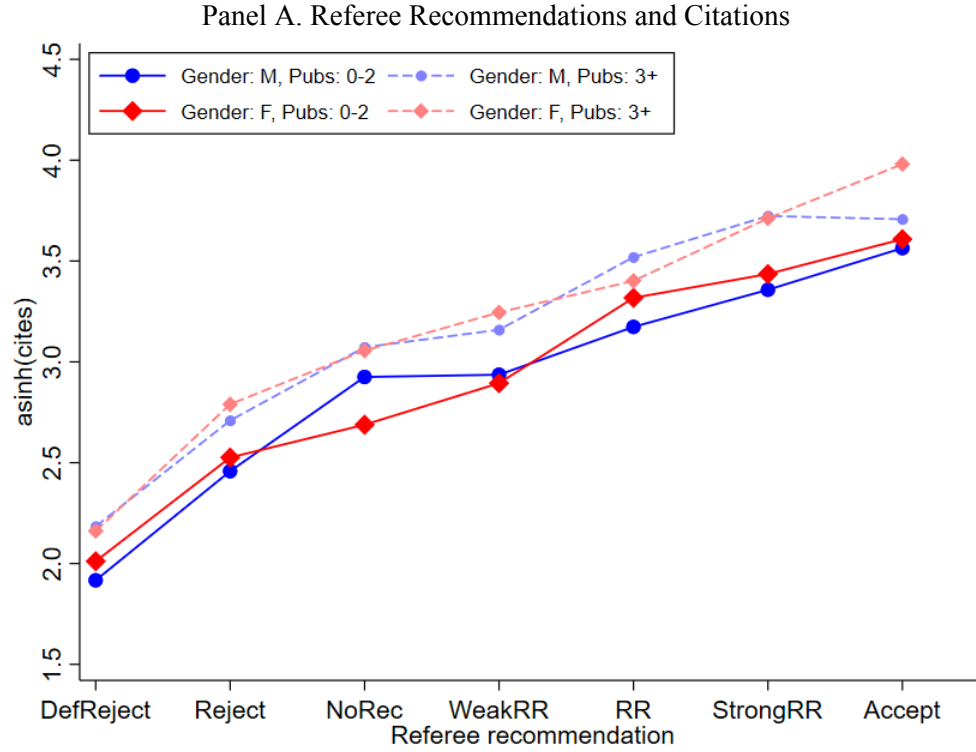


Figure VI
Referee Informativeness, by Referee Gender and Publications

Notes. Panel A shows the weighted *asinh* (citations) for a paper receiving a given recommendation. Panel B shows the R&R rate for a paper receiving a given recommendation. Both show the results separately by referee gender.

TABLE I
SURVEY OF ECONOMISTS ABOUT ROLE OF AUTHOR AND REFEREE GENDER

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

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

TABLE II
SUMMARY STATISTICS FOR NON-DESK-REJECTED PAPERS

Sample:	Non-Desk-Rejected Papers					
	All male	All female	Mix., F-led	Mix., other	Undet.	All
	(1)	(2)	(3)	(4)	(5)	(6)
<i>Google Scholar Citations</i>						
Asinh Citations	2.72 (1.84)	2.61 (1.85)	3.09 (1.81)	2.92 (1.81)	2.15 (1.77)	2.74 (1.84)
<i>Editorial Decisions</i>						
Received R&R Decision	0.15	0.12	0.17	0.14	0.12	0.15
<i>Author Publications in 35 high-impact journals</i>						
Publications: 0	0.32	0.59	0.00	0.22	0.53	0.32
Publications: 1	0.17	0.18	0.23	0.15	0.11	0.17
Publications: 2	0.12	0.09	0.22	0.13	0.11	0.12
Publications: 3	0.11	0.05	0.19	0.12	0.08	0.11
Publications: 4-5	0.13	0.07	0.20	0.16	0.08	0.13
Publications: 6+	0.16	0.02	0.17	0.22	0.09	0.15
<i>Number of Authors</i>						
1 author	0.36	0.72	0.00	0.00	0.26	0.31
2 authors	0.43	0.25	0.53	0.46	0.39	0.42
3 authors	0.18	0.03	0.37	0.40	0.27	0.22
4+ authors	0.03	0.00	0.10	0.14	0.08	0.05
<i>Field of Paper</i>						
Development	0.04	0.07	0.05	0.05	0.04	0.05
Econometrics	0.06	0.04	0.06	0.06	0.09	0.06
Finance	0.06	0.04	0.06	0.05	0.09	0.06
Health, Urban, Law	0.04	0.07	0.07	0.06	0.04	0.05
History	0.01	0.02	0.01	0.01	0.01	0.01
International	0.06	0.06	0.10	0.06	0.05	0.06
Industrial Organization	0.05	0.05	0.04	0.05	0.05	0.05
Lab/Experiments	0.02	0.02	0.04	0.04	0.02	0.03
Labor	0.11	0.20	0.14	0.15	0.10	0.12
Macro	0.11	0.06	0.08	0.08	0.10	0.10
Micro	0.11	0.08	0.09	0.11	0.09	0.11
Public	0.05	0.05	0.03	0.05	0.04	0.05
Theory	0.11	0.08	0.09	0.08	0.11	0.10
Unclassified	0.05	0.05	0.05	0.06	0.04	0.05
Missing Field	0.11	0.12	0.10	0.11	0.13	0.11
<i>Gender-Field Variables</i>						
Share female in fields	0.15	0.18	0.17	0.17	0.15	0.16
Gender-topic fields	0.03	0.08	0.06	0.06	0.04	0.04
<i>Referee Recommendations</i>						
Fraction Definitely Reject	0.12	0.12	0.10	0.12	0.19	0.12
Fraction Reject	0.54	0.56	0.54	0.54	0.51	0.54
Fraction with No Rec'n	0.06	0.06	0.04	0.05	0.06	0.06
Fraction Weak R&R	0.10	0.10	0.12	0.11	0.10	0.10
Fraction R&R	0.10	0.10	0.12	0.10	0.09	0.10
Fraction Strong R&R	0.04	0.04	0.04	0.04	0.03	0.04
Fraction Accept	0.04	0.03	0.03	0.03	0.02	0.03
<i>Referee Publications in 35 high-impact journals</i>						
Share of refs w/ 3+ publications	0.46	0.40	0.47	0.45	0.45	0.45
<i>Referee genders (share per paper)</i>						
Male	0.85	0.75	0.79	0.81	0.84	0.83
Female	0.14	0.24	0.20	0.18	0.14	0.15
Ambiguous	0.01	0.01	0.01	0.01	0.02	0.01
Number of Observations	10,199	1,097	585	2,612	654	15,147
Share of Papers	0.67	0.07	0.04	0.17	0.04	1.00

Notes. Sample of submissions, excluding papers with no referees or only 1 referee assigned. Author publications are counts of publications in 35 high-impact journals (see 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 submission. Indicators of fields for a paper with N codes are set to 1/N.

TABLE III
REFEREE RECOMMENDATIONS, IMPACT OF AUTHOR TEAM GENDER

Specification:	OLS Models for Index of Referee Recommendations				Linear Probability Models for R&R Recommendation or Better			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
<i>Authors' Genders (Omitted: All Male Authors)</i>								
All Female Authors	-0.043 (0.014)	0.018 (0.014)			-0.029 (0.008)	0.001 (0.008)		
Mixed-Gender Author Team senior author female	0.045 (0.020)	0.009 (0.020)			0.014 (0.012)	-0.002 (0.012)		
Mixed-Gender Author Team other	-0.015 (0.010)	-0.030 (0.011)			-0.011 (0.006)	-0.018 (0.006)		
Undetermined Gender Team	-0.121 (0.019)	-0.073 (0.019)			-0.049 (0.010)	-0.026 (0.010)		
<i>Referee Gender (Omitted: Male Referee)</i>								
Female Referee			-0.025 (0.018)	-			-0.012 (0.012)	-
<i>Gender Interactions</i>								
All Female Auth. × Female Ref.			0.009 (0.047)	0.004 (0.049)			-0.002 (0.030)	-0.007 (0.031)
Mixed Auth. (F-senior) × Female Ref.			-0.027 (0.062)	-0.024 (0.063)			-0.009 (0.041)	-0.007 (0.042)
Mixed Auth. (other) × Female Ref.			-0.009 (0.036)	-0.001 (0.037)			-0.003 (0.023)	0.001 (0.024)
Undetermined Auth. × Female Ref.			0.055 (0.070)	0.048 (0.073)			0.004 (0.042)	-0.001 (0.044)
Paper Fixed Effects	No	No	Yes	Yes	No	No	Yes	Yes
Controls for Author Pub., No. of Authors, Field, Gender-Field, and Journal-Year	No	Yes	-	-	No	Yes	-	-
Control for Referee Pub.	No	No	Yes	Yes	No	No	Yes	Yes
Control for Referee Pub. × Author Gender	No	No	No	Yes	No	No	No	Yes
Control for Referee Gender × Author Pub	No	No	No	Yes	No	No	No	Yes
Control for Referee Gender × Field	No	No	No	Yes	No	No	No	Yes
N	38,840	38,840	38,840	38,840	38,840	38,840	38,840	38,840
R-squared	0.016	0.044	0.504	0.504	0.012	0.030	0.452	0.453

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

TABLE IV
CITATIONS AND EDITOR DECISIONS, IMPACT OF AUTHOR TEAM GENDER

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

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

TABLE V
CITATIONS AND EDITOR DECISION, ROBUSTNESS

	Coefficients in Citation Model			Coefficients in Probit of R&R Decision			R&R Rate for All-Female Papers	
	All-Female Authors	Gender, Senior Female	Mixed-Gender, Other	All-Female Authors	Mixed-Gender, Senior Female	Mixed-Gender, Other	Data	Cite-Max
Robustness Dimension:	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
<i>Benchmark</i>	0.22 (0.05)	0.06 (0.07)	0.01 (0.04)	0.01 (0.06)	0.10 (0.07)	-0.02 (0.05)	0.122 (0.01)	0.185
<i>Varying Controls</i>								
1. Extra Controls for Author and Institutional Prominence	0.17 (0.05)	0.04 (0.07)	0.00 (0.04)	-0.01 (0.07)	0.13 (0.07)	-0.01 (0.05)	0.122	0.183
2. Extra Controls for Auth. Inst. Prominence, plus interactions of Field with Year and Gender-Field	0.19 (0.05)	0.04 (0.07)	-0.01 (0.04)	0.00 (0.07)	0.12 (0.07)	0.00 (0.05)	0.122	0.189
<i>Alternative Citation Measures</i>								
3. Probit Model for Top-Cited Paper	0.13 (0.05)	0.10 (0.06)	0.00 (0.03)	0.01 (0.06)	0.10 (0.07)	-0.02 (0.05)	0.122	0.190
4. Dropping observations with 0 citations (N=12306)	0.14 (0.04)	0.10 (0.06)	0.01 (0.03)	0.04 (0.07)	0.11 (0.06)	-0.01 (0.05)	0.146	0.222
5. Tobit Model of asinh (SSCI Cites) (2006-2008) (N=4507)	0.32 (0.15)	-0.04 (0.27)	-0.02 (0.11)	-0.01 (0.11)	0.41 (0.11)	0.10 (0.06)	0.124	0.215
<i>By Year of Submissions</i>								
6. Papers in Years 2003-2009 (N=7207)	0.31 (0.08)	0.10 (0.13)	0.02 (0.06)	0.00 (0.08)	0.36 (0.14)	-0.01 (0.07)	0.147	0.258
7. Papers in Years 2010-2013 (N=7940)	0.13 (0.07)	0.02 (0.08)	0.00 (0.06)	0.05 (0.09)	-0.13 (0.09)	-0.02 (0.06)	0.097	0.129
8. Papers in Years 2014-2017 (N=6544)	-	-	-	0.09 (0.17)	0.06 (0.17)	-0.05 (0.10)	-	-
<i>By Number of Authors</i>								
9. Papers with 1 Author (N=4639)	0.17 (0.07)	-	-	0.11 (0.08)	-	-	0.122	0.150
10. Papers with 2 Authors (N=6406)	0.34 (0.08)	-0.08 (0.10)	0.01 (0.06)	-0.23 (0.12)	0.12 (0.11)	-0.13 (0.08)	0.128	0.308
11. Papers with 3 Authors (N=4102)	0.06 (0.23)	0.23 (0.09)	0.03 (0.05)	-0.15 (0.49)	0.10 (0.10)	0.05 (0.07)	0.079	0.115
<i>By Author Publications</i>								
12. Papers by Authors with 0-3 Previous Pubs. (N=10771)	0.18 (0.05)	0.06 (0.09)	0.06 (0.04)	0.04 (0.07)	0.08 (0.10)	-0.03 (0.07)	0.113	0.167
13. Papers by Authors with 4+ Previous Pubs. (N=4376)	0.49 (0.10)	0.05 (0.10)	-0.05 (0.07)	-0.14 (0.22)	0.15 (0.13)	-0.01 (0.07)	0.219	0.533
<i>By Share of Women in the Field</i>								
14. Fields with Lower Share of Women (N=6753)	0.08 (0.09)	0.08 (0.11)	0.04 (0.05)	0.06 (0.15)	0.19 (0.11)	-0.14 (0.08)	0.112	0.128
15. Fields with Higher Share of Women (N=6754)	0.24 (0.07)	0.00 (0.08)	0.02 (0.06)	-0.02 (0.12)	0.11 (0.11)	0.07 (0.07)	0.105	0.178
16. Field Variable Missing (N=1640)	0.47 (0.15)	0.22 (0.28)	-0.28 (0.11)	0.00 (0.25)	-0.02 (0.25)	-0.01 (0.17)	0.203	0.341
<i>By Number of Referees</i>								
17. Papers with 1-2 Referees (N=7940)	0.29 (0.07)	0.07 (0.12)	0.03 (0.05)	0.07 (0.08)	0.14 (0.15)	0.03 (0.08)	0.105	0.203
18. Papers with 3+ Referees (N=7207)	0.12 (0.07)	0.04 (0.08)	-0.02 (0.06)	-0.04 (0.09)	0.08 (0.08)	-0.05 (0.08)	0.142	0.198
<i>By Share of Women Referees</i>								
19. Papers with All-Male Referees (N=10195)	0.17 (0.06)	-0.05 (0.11)	-0.03 (0.06)	0.02 (0.09)	0.03 (0.08)	-0.08 (0.07)	0.133	0.183
20. Papers with some Female Referees (N=4952)	0.29 (0.07)	0.19 (0.09)	0.08 (0.05)	0.09 (0.11)	0.22 (0.11)	0.08 (0.11)	0.109	0.200

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

TABLE VI
CITATIONS, RESULTS ACCEPTED PAPERS, AND PUBLISHED PAPERS

Data Set:	Editorial Express	Published Papers from Econlit					
Sample of Papers:	Accepted Papers	Publications in our 4 Journals, 2008-2015					
Asinh of GS Citations Measured in:	2015	2019					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
<i>Authors' Genders</i>							Lasso
All Female	0.24 (0.13)	0.12 (0.11)	0.12 (0.13)	0.08 (0.13)	0.10 (0.13)	0.09 (0.13)	0.08
Mixed-Gender Author Team senior author female	-0.11 (0.16)	0.24 (0.13)	0.20 (0.16)	0.13 (0.15)	0.15 (0.15)	0.13 (0.15)	0.19
Mixed, other	0.09 (0.11)	-0.00 (0.07)	0.04 (0.08)	0.02 (0.08)	0.03 (0.08)	0.03 (0.08)	-0.02
Undetermined	-0.40 (0.19)	-0.63 (0.29)	-0.75 (0.31)	-0.75 (0.32)	-0.78 (0.34)	-0.78 (0.33)	-0.67
<i>Asinh(sum of counts of words)</i>							
Theorem, Proposition, etc. (6 words)				-0.12 (0.04)		-0.07 (0.04)	-0.05
Empirical, Table, etc. (11 words)				0.20 (0.04)		0.15 (0.04)	0.17
Structural, Simulation, etc. (8 words)				0.00 (0.02)		-0.01 (0.02)	
RCT, Laboratory, etc. (5 words)				0.04 (0.02)		0.03 (0.03)	0.00
<i>Coding of Paper Content</i>							
Role of Model (0-1 scale)					-0.15 (0.15)	-0.12 (0.15)	
Role of empirical analysis (0-1 scale)					0.41 (0.16)	0.15 (0.17)	0.12
Role of Policy (0-1 scale)					0.20 (0.12)	0.20 (0.12)	0.15
# Propositions/theorems < 75th percentile among positives					-0.24 (0.11)	-0.20 (0.11)	-0.08
# Propositions/theorems > 75th percentile among positives					-0.41 (0.12)	-0.29 (0.12)	-0.05
# Modeling equations < 75th percentile among positives					0.04 (0.08)	0.04 (0.09)	
# Modeling equations > 75th percentile among positives					0.24 (0.14)	0.25 (0.14)	
# Estimating Equations < 75th percentile among positives					0.07 (0.08)	0.02 (0.08)	
# Estimating Equations > 75th percentile among positives					0.01 (0.09)	-0.01 (0.09)	
Controls for Fraction Referee Recs	Yes	No	No	No	No	No	No
Controls for Author Pub., No. of Authors, Field, Gender-Field, and Journal-Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls for 2-Digit JEL	No	No	Yes	Yes	Yes	Yes	Yes
N	1,713	1,719	1,719	1,719	1,719	1,719	1,719
R ²	0.30	0.28	0.54	0.56	0.56	0.57	-

Notes. The sample in Column 1 is accepted papers in main sample. The sample in Columns 2-7 is published articles from Econlit, in the 4 journals in our sample in 2008-2015, corresponding approximately to submissions in our sample. For this data set, we measure Google Scholar citations in March 2019 (as opposed to mid 2015 for Column 1). For the sum of word-counts, we create four groups of words out of 29 words. The qualitative coding of the role of model, empirical analysis, and policy is on a 0-10 scale, which we rescale to a 0-1 scale. Additionally, we counted the number of propositions and theorems, model equations, and estimating equations. For each of these three categories, we indicate (if the paper has a positive number) whether or not the paper exceeds the 75th percentile for that variable. In column 7, we perform a lasso variable selection, forcing the inclusion of the author-gender variables.

TABLE VII
DESK REJECTION, IMPACT OF AUTHOR TEAM GENDER

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

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

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

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

TABLE IX
REFEREE DELAYS, DECISION TIME, AND DURATION OF REVISIONS

	OLS Regression of Number of Days from Submission to Referee Report		Number of Days		No. of Rounds (for R&Rs)	Total Days from First Sub. to Accept
	(1)	(2)	Sub. To Last Report Received	Submission To Editor Dec.		
<i>Authors' Genders</i>						
All Female	-1.36 (0.97)		-1.10 (1.40)	-1.85 (1.49)	0.03 (0.09)	3.59 (30.47)
Mixed-Gender Author Team senior author female	-0.24 (1.93)		0.60 (2.03)	0.71 (3.18)	-0.06 (0.08)	-8.39 (31.89)
Mixed-Gender, other	0.17 (0.72)		1.18 (0.84)	1.82 (1.60)	0.01 (0.05)	19.41 (22.27)
Undetermined Gender Mix	-1.81 (1.45)		-0.32 (1.97)	-0.71 (2.50)	0.08 (0.11)	38.45 (52.18)
<i>Referee Gender (Omitted: Male Referee)</i>						
Female Referee		0.44 (0.98)				
<i>Gender Interactions</i>						
All Female Auth. X Female Ref.		1.32 (2.52)				
Mixed Auth. (senior-F) X Female Ref.		-2.91 (3.39)				
Mixed Auth. (other) X Female Ref.		1.11 (1.92)				
Undetermined Auth. X Female Ref.		0.87 (4.71)				
R&R Indicator			1.71 (1.62)	31.67 (9.26)		
Sample	Non-Desk-Rejected Papers		Non-Desk-Rejected Papers		R&R Papers That Are Ultimately Accepted	
Paper Fixed Effects	No	Yes	No	No	No	No
Controls for Referee Publications	No	Yes	No	No	No	No
Controls for Referee Recommendations	No	No	Yes	Yes	Yes	Yes
Controls for Author Pub., No. of Authors, Field & Gender-Field, Journal-Year f.e	Yes	-	Yes	Yes	Yes	Yes
Editor Fixed Effects	No	-	Yes	Yes	Yes	Yes
Mean of Dependent Variable:	73.6	73.6	94.9	116.9	2.79	634.1
N	38,825	38,825	15,147	15,147	1,713	1,667
R-squared	0.31	0.71	0.43	0.44	0.39	0.40

Notes. The sample in Columns 1-2 is referee-paper observations, including any referee who returned a review for that paper. Report time is calculated as the number of days from paper submission to referee report submission, rounded to the nearest 10. The sample in Columns 3-6 is papers observations, including only R&R papers that are ultimately published in Columns 5-6. Decision time is calculated as the number of days from paper submission to referee report submission, rounded to the nearest 10. Clustered standard errors by editor in parentheses.

A Online Appendix

A.1 Data

A.1.1 Google Scholar Citations

We extract Google Scholar citations using an automated web scraper. For every paper, we search the title of the paper in Google Scholar with “allintitle:” (e.g. “allintitle:Behavioral Welfare Economics”). This ensures that every result contains every word of our search criterion. Then, we extract the results of the first page of the search, storing the paper title, list of authors, and number of citations. We then compare the last names of the authors from Google Scholar compared to our data base and keep every entry with at least one matching last name. Finally, we sum the citations of the remaining entries to take into account earlier versions of the paper with many citations.

A.1.2 Published Papers

Sample. For the sample of published papers, we analyze papers in the *Journal of the European Economic Association*, the *Quarterly Journal of Economics*, the *Review of Economic Studies*, and the *Review of Economics and Statistics* from 2008 to 2015. To ensure a complete dataset, we extract the list of articles directly from the journal websites, for a total of 1838 articles. (We found that Econlit does not always provide the full list of papers published.) We then exclude articles that fall under the following categories: Papers and Proceedings, comments, errata, corrigenda, notes, and editorial announcements and letters, yielding a final sample of 1719 papers. For each article, we download the PDF.

In his sample, we run the same code used to generate the measures for author publications, author gender, number of authors, broad field, and the two measures of gender-field composition, as for our main sample. We also use the same procedure to extract the Google Scholar citations in March of 2019. We also extract the Web of Science (SSCI) citations, in this case downloading the micro data on all the citing papers, allowing us to perform the more detailed analysis of citations in Online Appendix Table 10.

Unlike in our main sample, we are able to check directly the accuracy of the key variables above. In particular, we hand check the Google Scholar citations and the gender coding of the authors for all papers in the sample. Online Appendix Table 11 presents the resulting evidence on the degree of measurement error in the data.

Additional Controls. For this sample of published papers, we build additional measures of characteristics of papers which we do not have for the main sample. The first one is a precise measure of sub-field of the paper. Specifically, for each paper, we obtain all the two-digit JEL code (i.e., G21) listed in EconLit. We create indicators for each individual JEL code, for a total of over 500 sub-field controls. We classify a paper with J JEL codes as being assigned with weight $1/J$ to each of the listed fields. For example, for a paper with 3 two-digit JELs, each of the three JELs are assigned $\frac{1}{3}$ and the remaining two-digit JELs are assigned 0.

The second set of controls aims to measure differences even within a sub-field in the content of a paper using word counts. We search within the full text of the paper, using an R program to search PDFs, for occurrence of a set of words. We count as associated with a theoretical profile the words “proposition, theorem, lemma, proof, model, theory.” Similarly, we count as associated with an empirical profile for the paper the words “empirical, data, standard error, table, regression, diff-in-diff, natural experiment, IV, RDD, impact, research design.” We also count words indicating structural estimates (“structural, logit, BLP, maximum likelihood, mixture, simulation, policy simulation, calibration”) and words associated with experimental results (“field experiment,

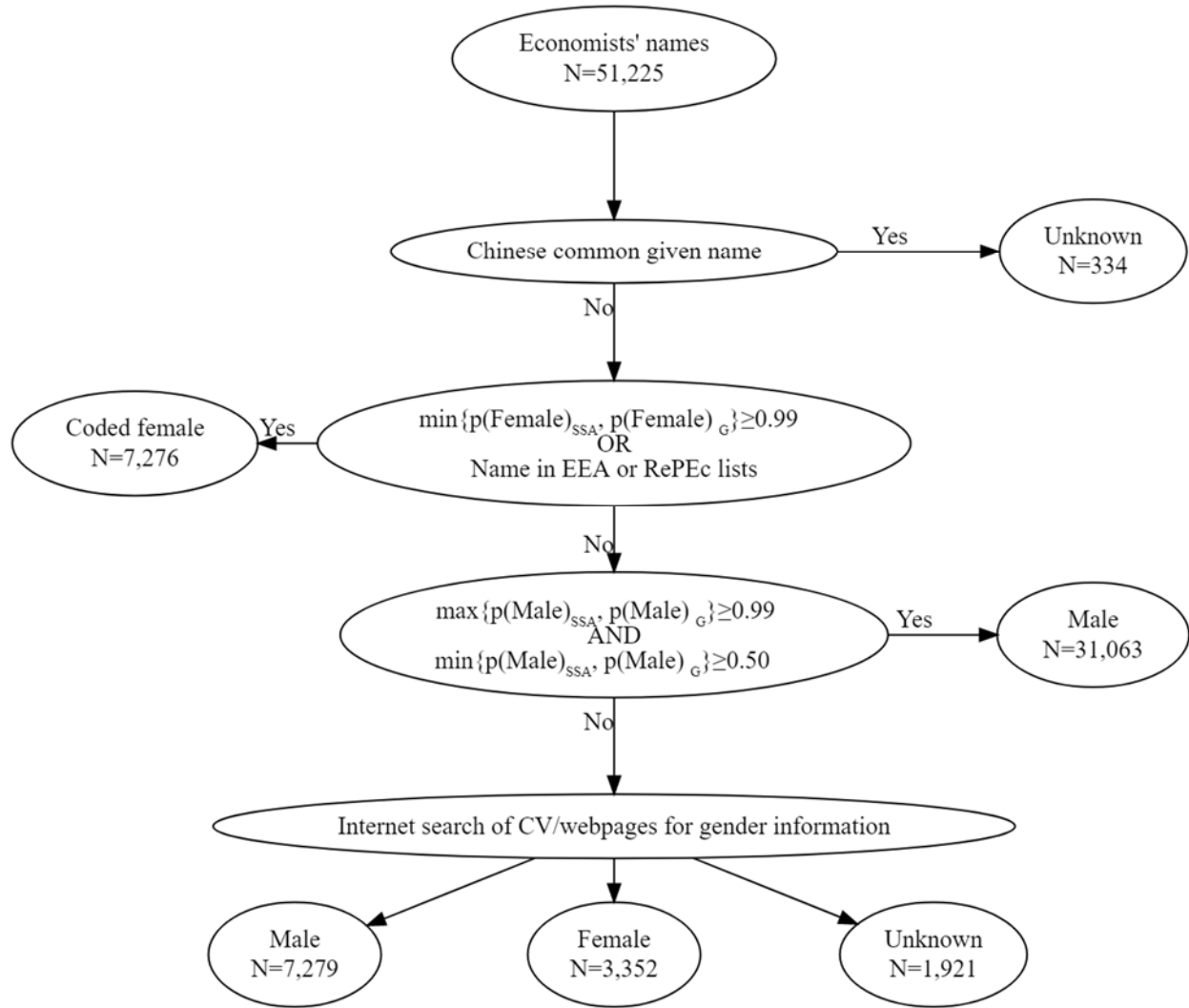
RCT, laboratory, subjects, survey”). We then use the asinh of the sum of counts of words in a category as control for the that feature of the paper in Table VI.

The third set of controls also measures differences along the empirical/theoretical line, but using qualitative assessment by a team of undergraduate research assistants. For the coding of paper content we assigned each paper randomly to 1 of 12 undergraduates. The undergraduates were then asked to rate the roles of empirical analysis, modeling, and policy of the papers on a scale of 1-10. Additionally, they were tasked with counting the number of propositions and theorems, modeling equations, and estimating equations.

The instructions for the task are the following: *“For NoPropsTheorems, count the number of Propositions, Theorems, Lemmas, and Corollaries, including the Appendices if a part of the published paper. Do not count Claims or Definitions. If the Appendix includes a proof of a Proposition stated within the text, do not count this twice. Include Corollaries not in the text but present in any such proofs. For NoMathEquations, count the number of equations related to a “theoretical model,” such as utility maximization, market equilibrium, etc. Count also equations that derive a theoretical econometric model. Include displayed equations but not those in footnotes. As in NoPropsTheorems, include equations in the published Appendix. Equations that span multiple lines count as one equation. For NoEstimatingEquations, count the number of equations related to empirical econometrics or estimation, such as OLS or IV specifications. If the equation is a derivation of a theoretical property of an econometric model, then include the equation in NoProsTheorems instead. For PagesofModel, count approximately the number of pages dedicated to modeling and deriving. Use integer numbers. For PagesofEmpirical, count approximately the number of pages dedicated to data or empirical model. Include any “Data Sections.” Use integer numbers. For RoleofModel, provide a qualitative assessment of the role of model and theory in the paper on an integer scale from 0 (none) to 10 (absolutely central). Ask yourself, “How central is the modeling contribution in the paper?” For RoleofEmpirical, provide a qualitative assessment of the role of data and empirical methods in the paper, on an integer scale from 0 (none) to 10 (absolutely central). Ask yourself, “How central is the empirical contribution of the paper?” For RoleofPolicy, provide a qualitative assessment of the role and influence of policy in the paper, on an integer scale from 0 (none) to 10 (absolutely central). Ask yourself, “How relevant is this paper to policy?”*

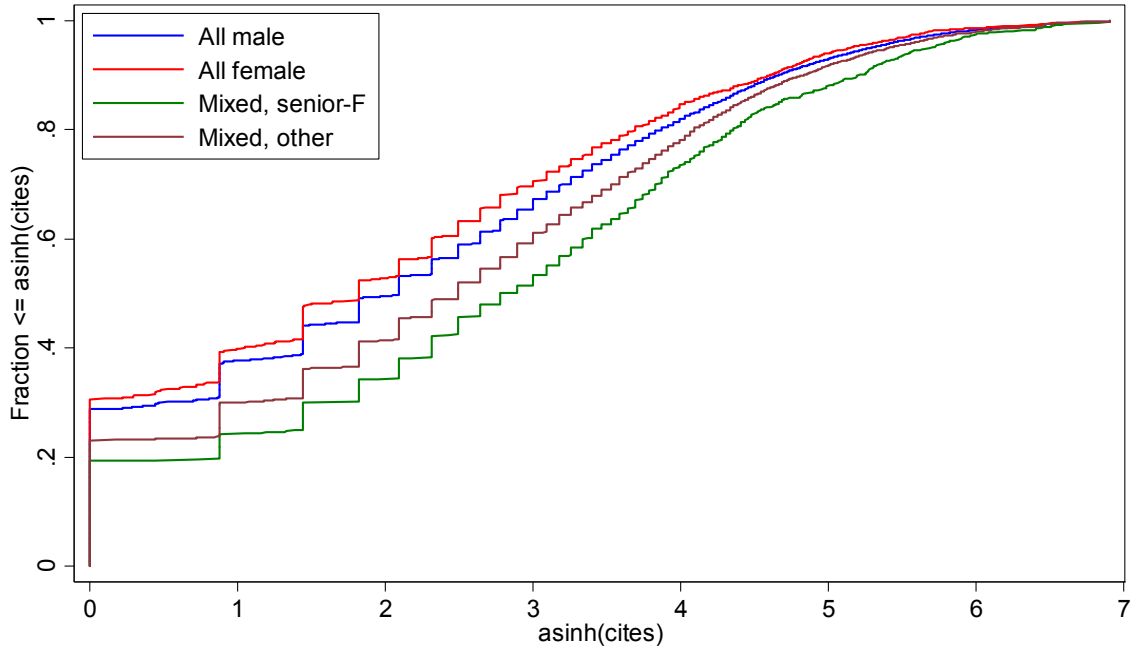
For the published papers, we also extract SSCI cites. For papers that cite the published papers, we attempt to assign 5-year impact factors for the journal in which it is published. In Online Appendix 10 Column 5, the sample is papers published in journals with 5-year impact factor 5 or greater. In Column 6, we have papers published with 5-year impact factor less than 3. Finally, in Column 7, we have papers published in journals that we could not match to an impact factor. Notably, some journals are outside of economics.

Online Appendix Figure 1. Coding Gender for Names

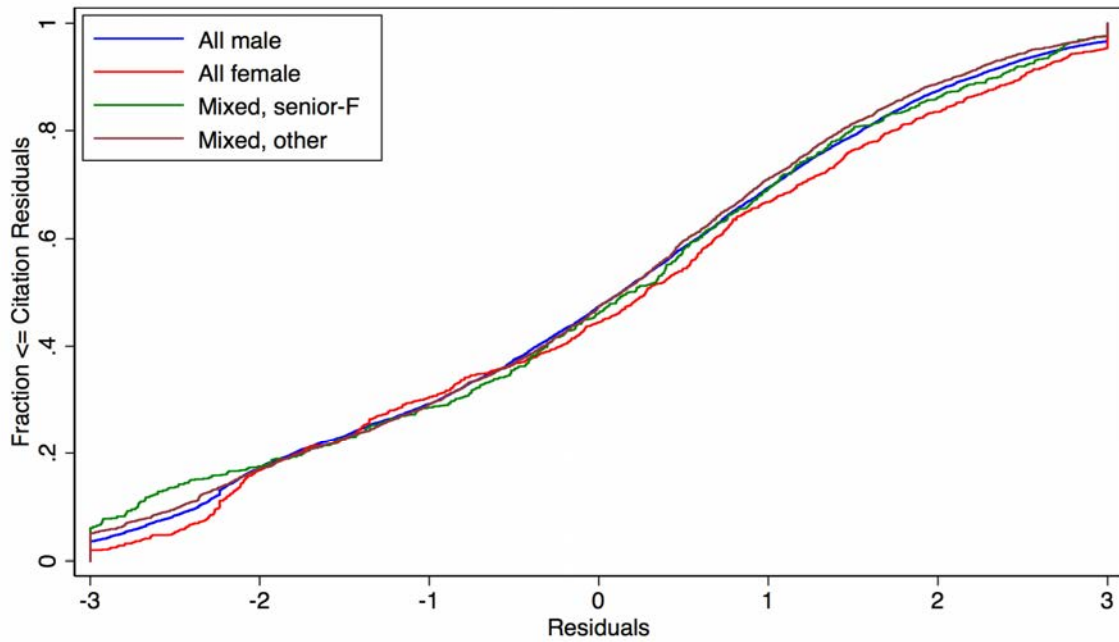


Notes. Graph shows the process by which gender is assigned to names.

Online Appendix Figure 2a. Paper Citations by Gender

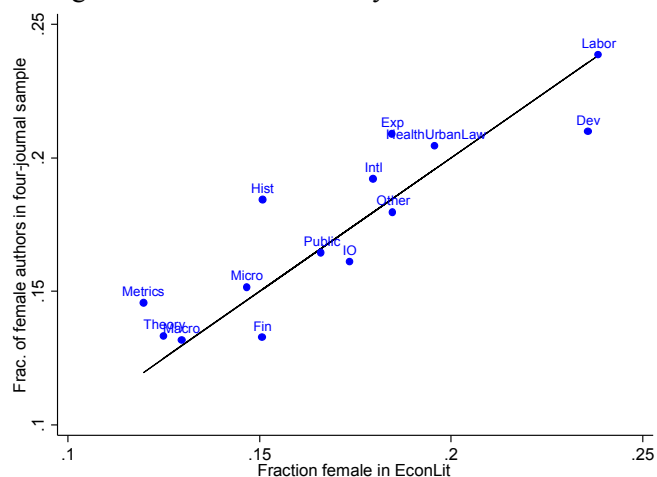


Online Appendix Figure 2b. Paper Citations by Gender, Citations Residualized

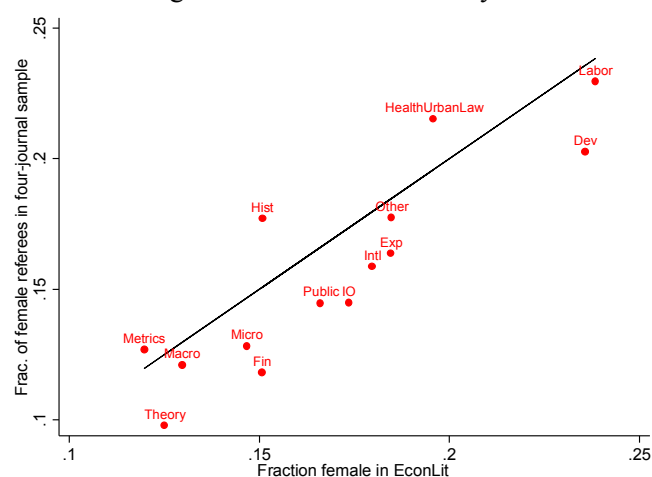


Notes. The Figure displays a few key summary statistics by gender. Figure 2a displays the CDF of the (asinh of) paper citations. Figure 2b displays the same citation variable, but after partialling out the key controls for journal-year fixed effects, fields, number of authors, and number of author publications (as in Table 5, Column 3).

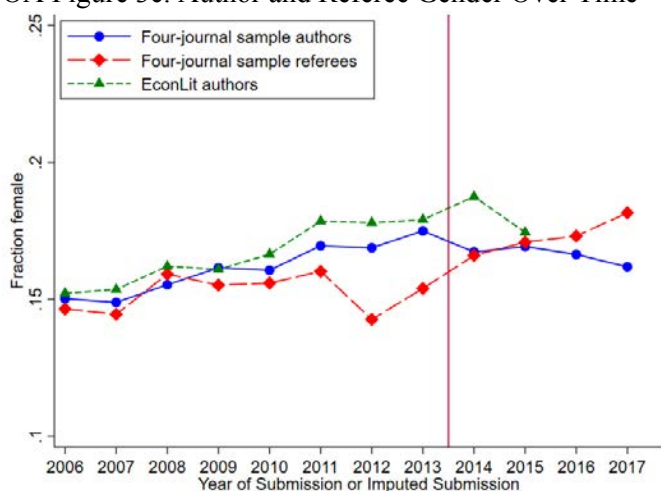
Online Appendix Figure 3. Share of Female Authors and Referees, by Field
OA Figure 3a. Author Gender by Field



OA Figure 3b. Referee Gender by Field



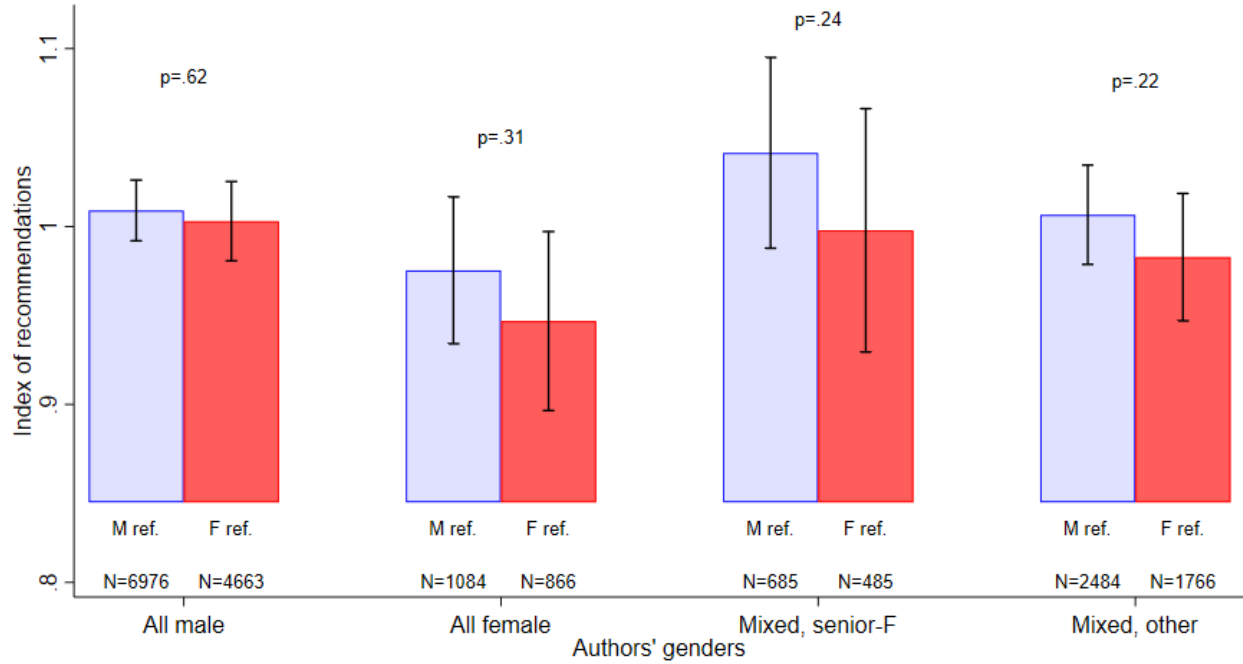
OA Figure 3c. Author and Referee Gender Over Time



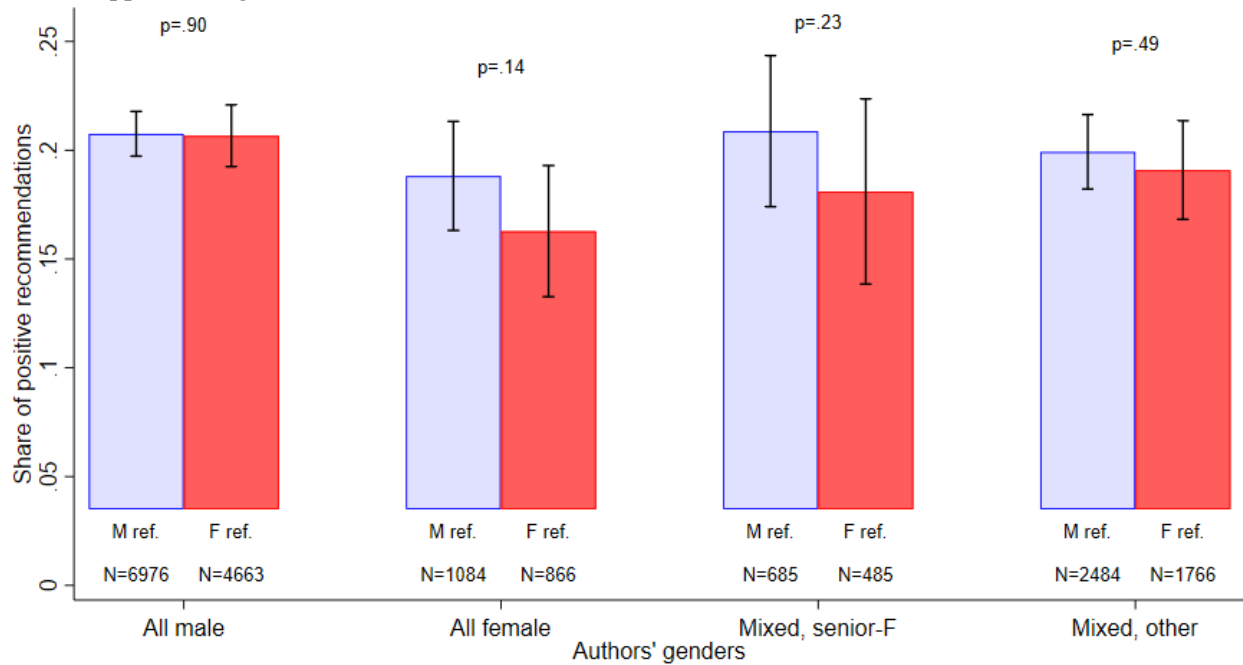
Notes. In this figure, we compare the fraction of female referees and female authors in our four journal sample as well as in a sample of authors with publications in 63 high-impact journal listed in Online Appendix Table 1, Panel B (which we label “EconLit authors”). In 3a and 3b, we break down the distributions by field, giving equal weight to each paper (ensuring papers with multiple JELs do not receive too much weight), with a 45 degree line for reference. In 3c, we compare fraction female across time. Each observation is either an author-paper or referee-paper. For the Econlit sample, we obtain the approximate year of submission as the year of publication minus 2. A vertical line has been added halfway through 2013, marking the end of our main sample.

Online Appendix Figure 4. Referee Evaluation by Author Gender and Referee Gender, Extended Sample (Up to 2017)

Online Appendix Figure 4a. Index of Referee Recommendations

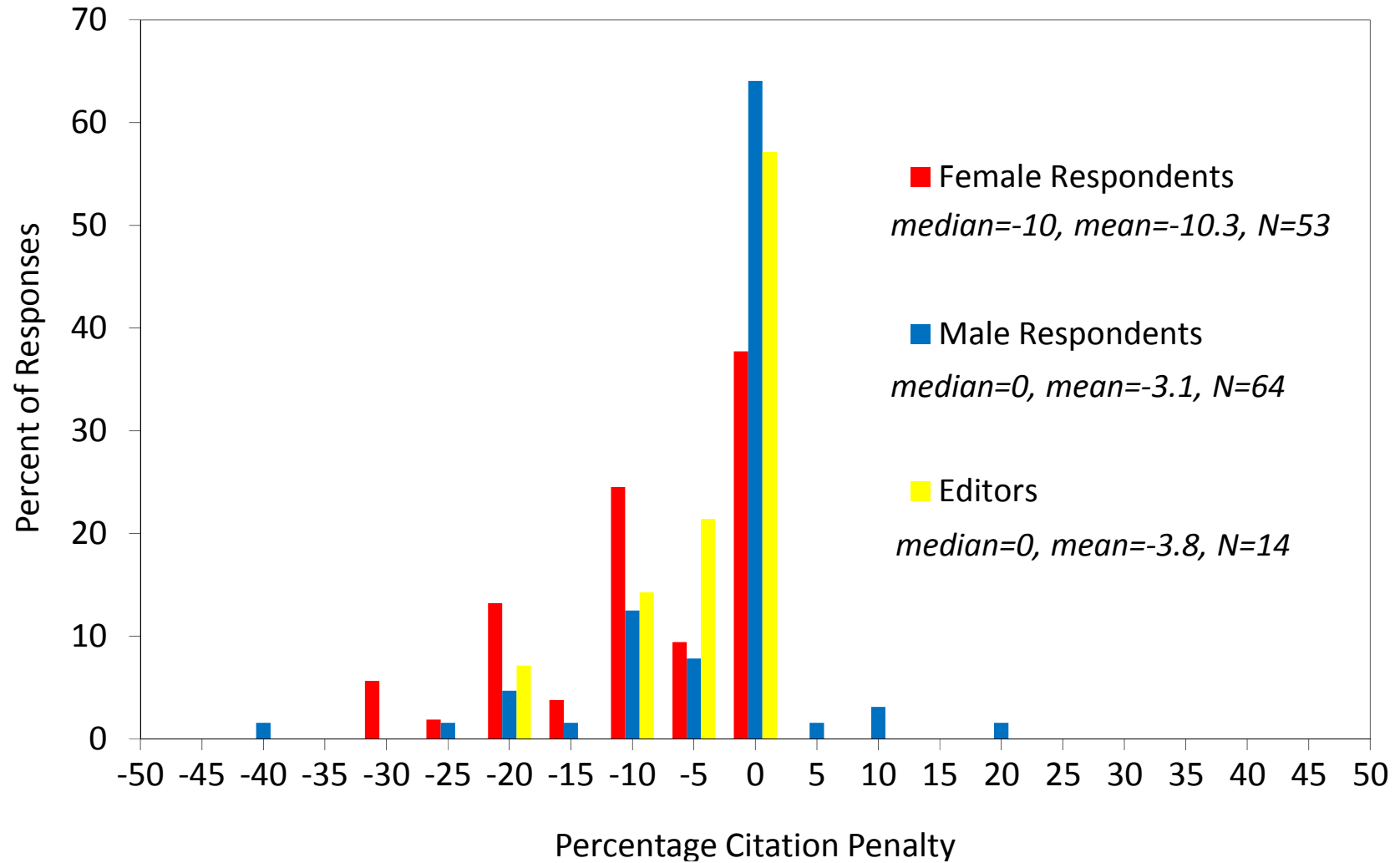


Online Appendix Figure 4b. Share of Positive Referee Recommendations



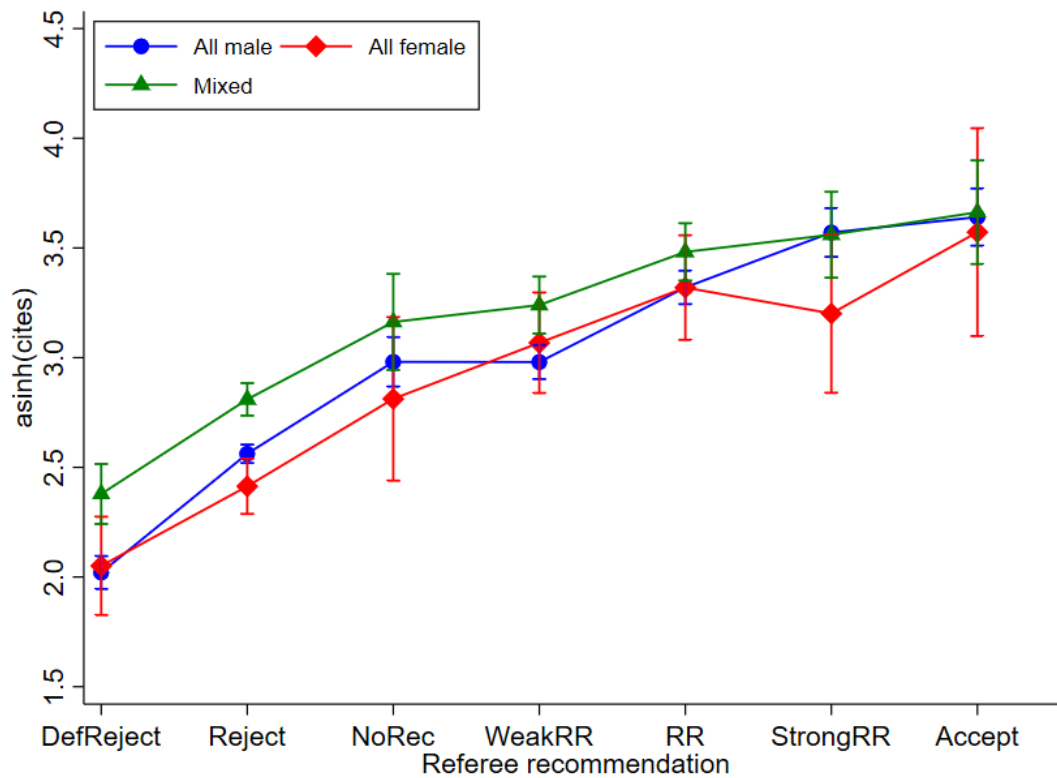
Notes. Online Appendix Figure 4a displays the mean recommendation given by referees based on gender. The index of referee recommendations is constructed using the coefficients in the citation model in Card and DellaVigna (forthcoming). From Definitely Reject to Accept, the values are 0, 0.67, 1.01, 1.47, 1.92, 2.27, 2.33. The bands show 2 standard error intervals, clustered at the paper level. The figure includes only 12,273 paper-referees that satisfy having both male and female referees. Additionally, we have dropped unknown gender referees and authors. Figure 4b shows the share of positive recommendations, defined as RR-Accept. In both panels, female referees are weighted at the paper level by $N_{\text{male}} / N_{\text{female}}$.

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

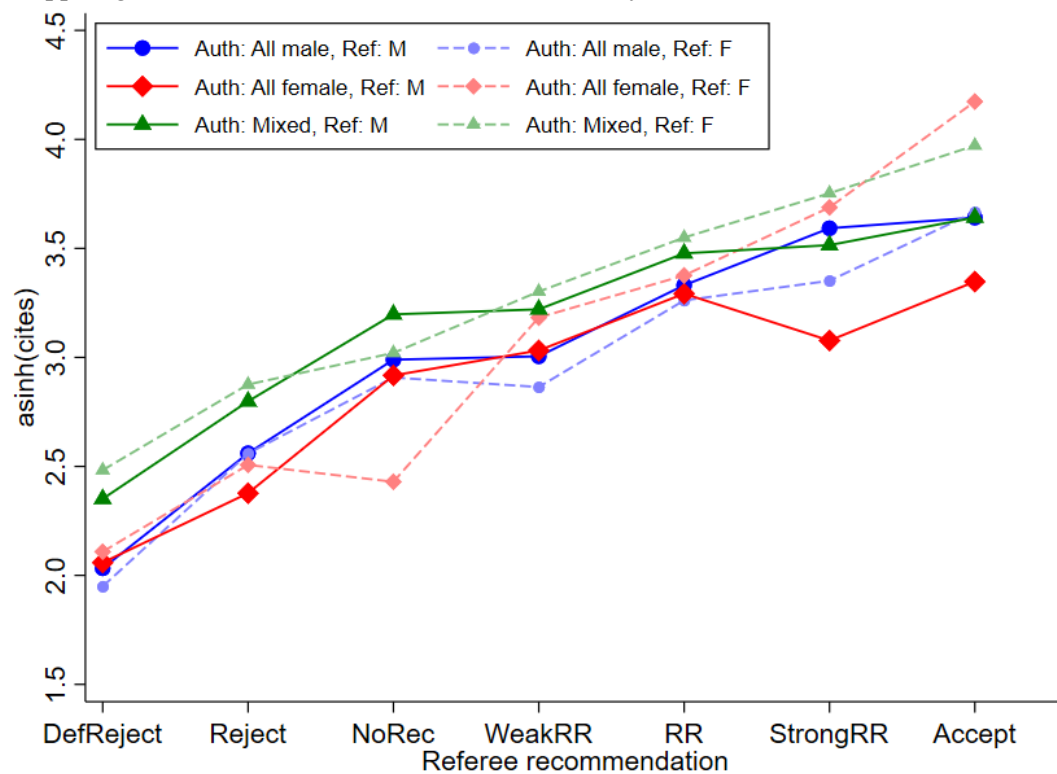


Notes. Tabulation of the response to question Q8 in the survey (Table I). The number of observations differs from the one in Table I because some of the survey respondents did not answer question Q8.

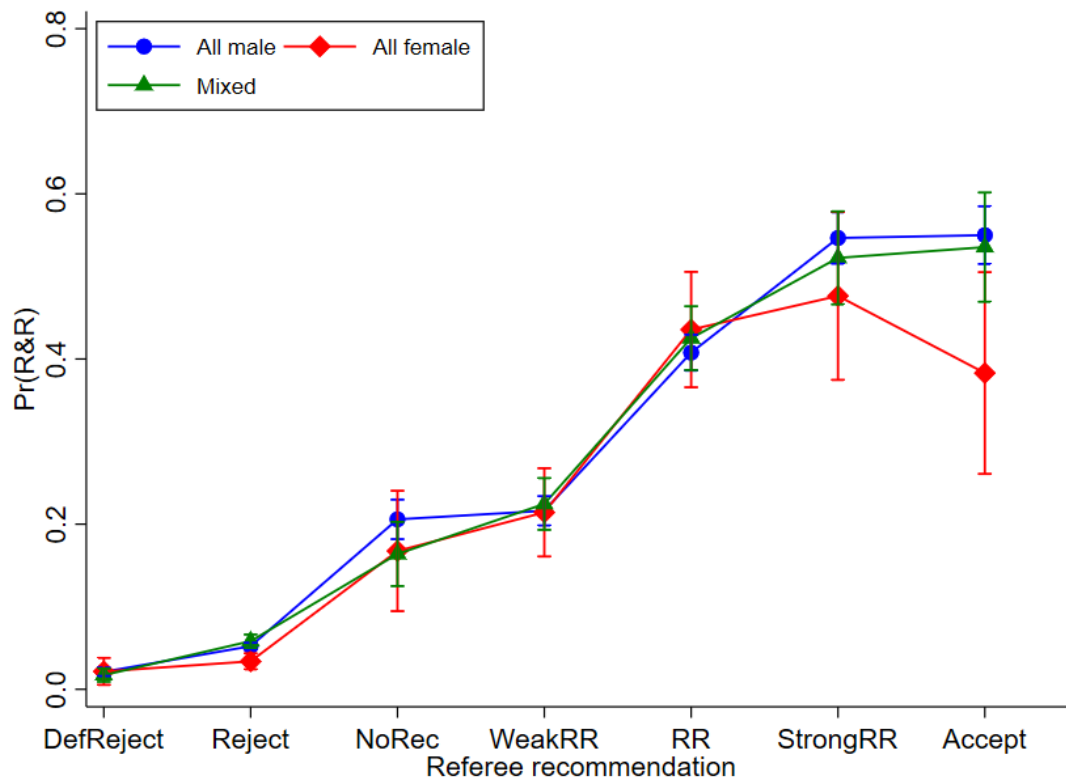
Appendix Figure 6. Differences in Citations and R&R Rate, by Author Gender, Additional Material
Online Appendix Figure 6a. Referee Recommendations and Citations, No Controls



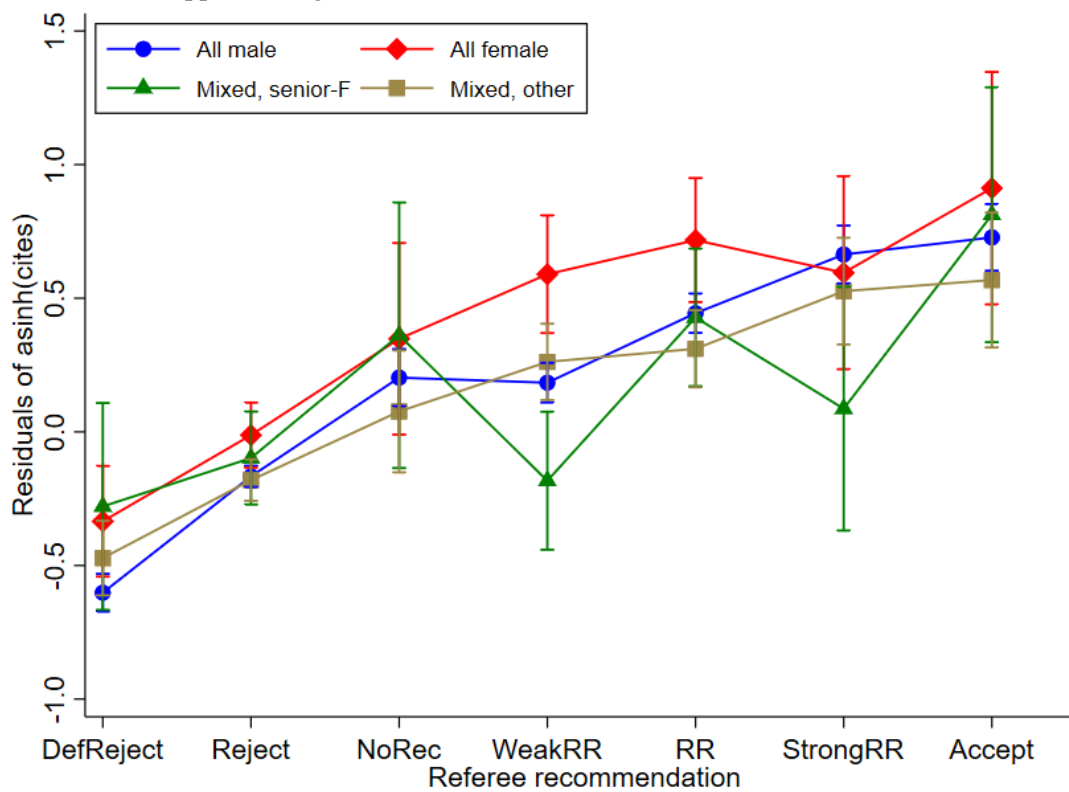
Online App. Figure 6b. Recommendations and Citations, by Author and Referee Gender, No Controls



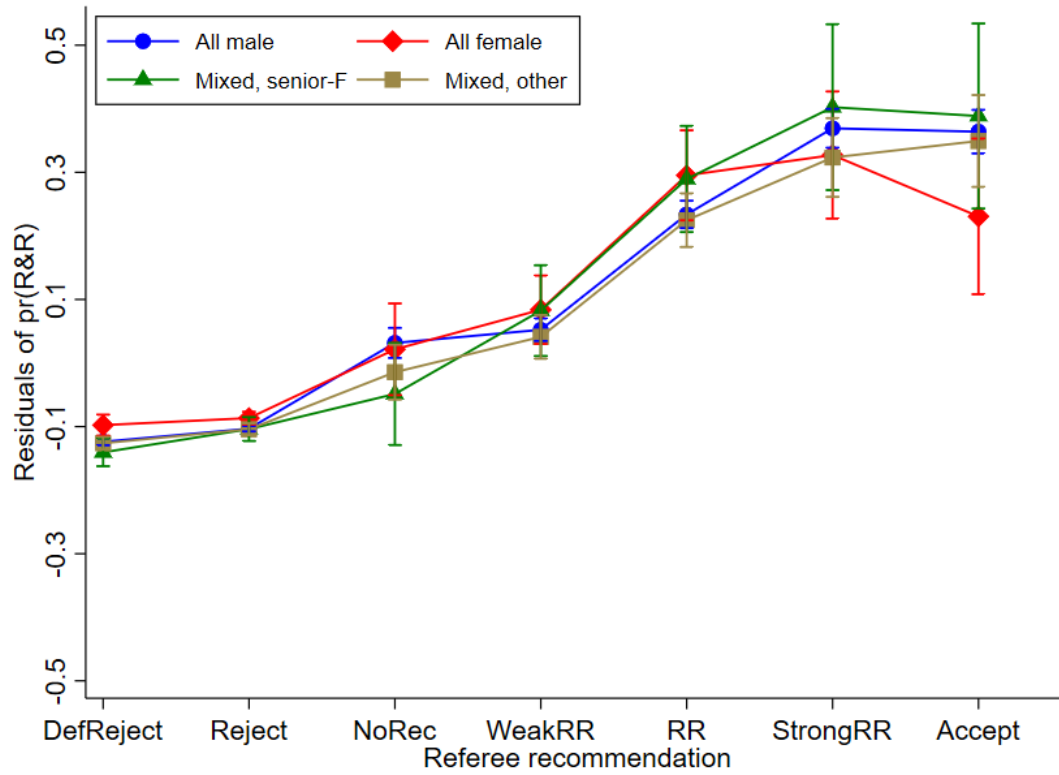
Online Appendix Figure 6c. Referee Recommendations and R&R Rate, No Controls



Online Appendix Figure 6d. Referee Recommendations and Citations, Controls

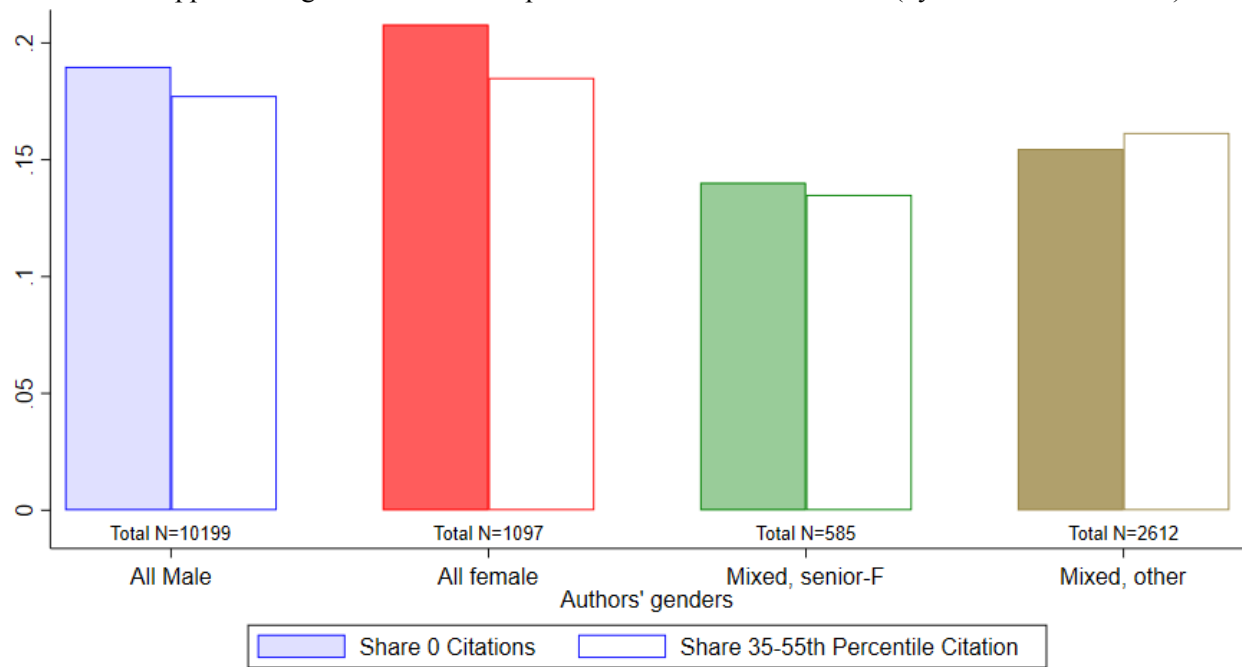


Online Appendix Figure 6e. Referee Recommendations and R&R Rate, Controls



Notes. Online Appendix Figure 6b shows the weighted *asinh* (citations) for a paper receiving a given recommendation, while OA Figure 6c shows the R&R rate for a paper receiving a given recommendation. Figures 6a and 6c show the results separately by author gender. Figure 6b splits these two categories further into referees' gender. The unit of observation is a referee report, and observations are weighted by the number of referee reports for the paper to ensure that each paper receives equal weight. Standard errors are clustered at the paper level. Figure 6b omits confidence intervals for legibility. Figures 6d and 6e are versions of, respectively, Figure IV Panel A and Figure IV Panel B with confidence intervals.

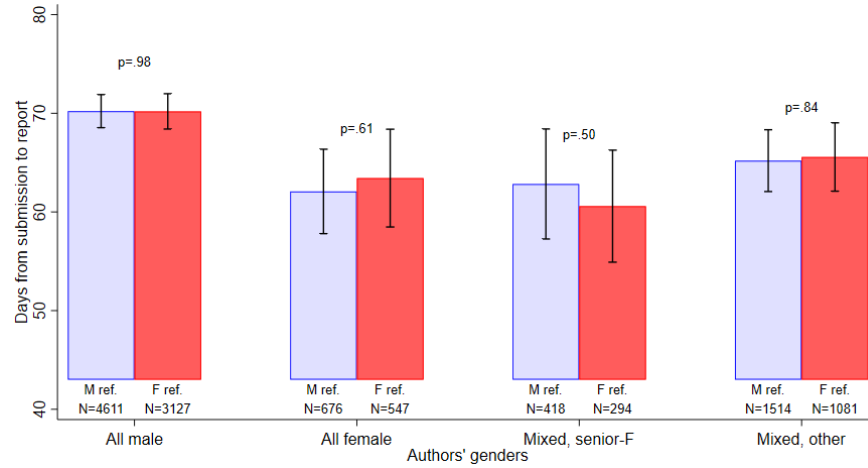
Online Appendix Figure 7. Share of Papers With 0 and Low Citations (by Author Gender Mix)



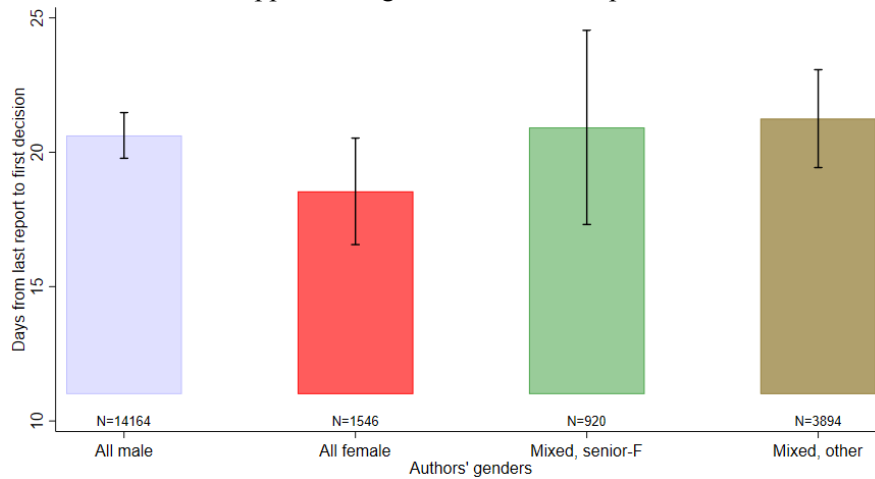
Notes. This figure presents, within each gender group, the share of papers with 0 citations and papers from the 35th up to the 55th percentile in citations. The percentile of citations is calculated within a journal-year cell.

Online Appendix Figure 8. Other Editorial Outcomes: Referee and Editorial Delay

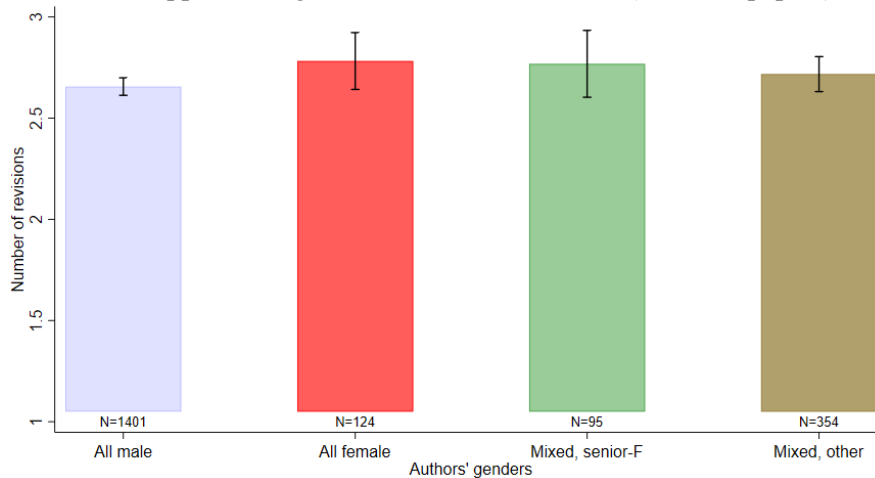
Online Appendix Figure 8a. Referee Response Time



Online Appendix Figure 8b. Editor Response Time



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



Notes. Figure 8a includes 12,273 paper-referees that satisfy having both male and female referees. Additionally, we have dropped unknown gender referees and authors. In 8b and 8c, each observation is a paper, submitted 2003–2013, for which we have the appropriate variable. In Panel A, female referees are weighted at the paper level by $N_{\text{male}} / N_{\text{female}}$. Panel B omits papers when the editor decides before the last report arrives.

Onl. App. Table 1. Journals Used for Publication Counts and Gender Coding

Panel A. List of Journals Used in Publication Counts

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

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

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

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

Online Appendix Table 2. Summary Statistics for all Submissions

Sample:	All Papers					
	All male	All female	Mix., F-led	Mix., other	Undet.	All
	(1)	(2)	(3)	(4)	(5)	(6)
<i>Google Scholar Citations</i>						
Asinh Citations	2.11 (1.83)	1.97 (1.80)	2.72 (1.85)	2.41 (1.82)	1.27 (1.57)	2.11 (1.83)
<i>Editorial Decisions</i>						
Not Desk-Rejected	0.59	0.56	0.67	0.61	0.41	0.58
Received R&R Decision	0.08	0.06	0.11	0.08	0.04	0.08
<i>Author Publications in 35 high-impact journals</i>						
Publications: 0	0.46	0.69	0.00	0.33	0.70	0.46
Publications: 1	0.17	0.15	0.28	0.17	0.11	0.17
Publications: 2	0.10	0.07	0.24	0.13	0.07	0.10
Publications: 3	0.08	0.04	0.18	0.10	0.04	0.08
Publications: 4-5	0.09	0.04	0.17	0.12	0.04	0.09
Publications: 6+	0.10	0.01	0.13	0.14	0.04	0.10
<i>Number of Authors</i>						
1 author	0.44	0.76	0.00	0.00	0.34	0.37
2 authors	0.39	0.21	0.52	0.48	0.36	0.39
3 authors	0.15	0.03	0.39	0.39	0.23	0.19
4+ authors	0.03	0.00	0.09	0.13	0.07	0.05
<i>Field of Paper</i>						
Development	0.04	0.06	0.05	0.05	0.05	0.05
Econometrics	0.07	0.05	0.06	0.06	0.09	0.07
Finance	0.07	0.05	0.07	0.06	0.10	0.07
Health, Urban, Law	0.05	0.06	0.06	0.06	0.04	0.05
History	0.01	0.01	0.01	0.01	0.01	0.01
International	0.06	0.07	0.09	0.06	0.06	0.06
Industrial Organization	0.05	0.05	0.04	0.05	0.05	0.05
Lab/Experiments	0.02	0.02	0.04	0.04	0.01	0.02
Labor	0.10	0.17	0.13	0.13	0.07	0.11
Macro	0.11	0.08	0.08	0.08	0.13	0.10
Micro	0.11	0.08	0.10	0.11	0.08	0.11
Public	0.05	0.04	0.04	0.05	0.03	0.05
Theory	0.10	0.07	0.08	0.07	0.08	0.09
Unclassified	0.06	0.06	0.06	0.06	0.07	0.06
Missing Field	0.11	0.12	0.10	0.11	0.14	0.11
<i>Gender-Field Variables</i>						
Share female in fields	0.15	0.18	0.17	0.17	0.15	0.16
Gender-topic fields	0.03	0.08	0.05	0.05	0.03	0.04
Number of Observations	19,814	2,273	921	4,723	2,159	29,890
Share of Papers	0.66	0.08	0.03	0.16	0.07	1.00

Notes. Table presents information on mean characteristics of all submitted papers. Author publications are based on publications in 35 high-impact journals (Online Appendix Table 1) in the 5 years prior to submission. In the case of multiple authors, the measure is the maximum over all coauthors. Field is based on JEL codes at paper submission. Indicators of fields for a paper that lists N codes are set to 1/N.

Online Appendix Table 3. Referee Assignment

<u>Dependent Variable:</u>	Linear Probability Models			
	Indicator for Female Referee			Referee with 3+ Pub.
	(1)	(2)	(3)	(4)
<i>Authors' Genders (Omitted: All Male Authors)</i>				
All Female Authors	0.111 (0.009)	0.103 (0.010)	0.074 (0.009)	-0.025 (0.010)
Mixed-Gender Author Team	0.064 (0.011)	0.062 (0.011)	0.049 (0.011)	-0.006 (0.013)
senior author female				
Mixed-Gender Author Team	0.043 (0.006)	0.040 (0.006)	0.026 (0.006)	-0.010 (0.007)
other				
Undetermined Gender Team	0.014 (0.009)	0.007 (0.009)	0.004 (0.009)	-0.001 (0.014)
<i>Gender-field controls</i>				
Share female in sub-fields			0.297 (0.043)	-0.036 (0.049)
Fraction of gender-topic sub-fields			0.198 (0.019)	-0.034 (0.021)
Mean of the Dependent Variable:	0.157	0.157	0.157	0.461
Controls for Author Publications	No	Yes	Yes	Yes
Controls for Referee Publications	No	Yes	Yes	No
Controls for No. of Authors	No	Yes	Yes	Yes
Controls for Field	No	No	Yes	Yes
Indicators for Journal-Year	Yes	Yes	Yes	Yes
N	38,438	38,438	38,438	38,438
R-squared	0.015	0.029	0.048	0.023

Notes. The sample is paper-referee observations for 15,147 papers with at least two referees assigned, excluding unknown gendered referees. The dependent variable in Columns 1-3 is an indicator for the referee being female, while the dependent variable in Column 4 is an indicator for the referee having at least 3 publications in the 35 publications in the previous 5 years. Standard errors clustered by paper in parentheses.

Online Appendix Table 4. Referee Recommendations, by Referee Publications

Specification:	OLS Models for Index of Referee Recommendations		Linear Probability Models for Receiving an R&R Recommendation or Better	
	(1)	(2)	(3)	(4)
<i>Referee Gender (Omitted: Male Referee)</i>				
Female Referee	0.066 (0.141)	-	-0.005 (0.086)	-
Female Referee $\times$ 3+Publications for Ref.	-0.020 (0.036)	-0.015 (0.036)	-0.023 (0.023)	-0.019 (0.023)
<i>Gender Interactions</i>				
All Female Auth. $\times$ Female Ref.	-0.008 (0.057)	-0.002 (0.058)	-0.033 (0.037)	-0.035 (0.038)
Mixed Auth. (F-senior) $\times$ Female Ref.	-0.006 (0.080)	-0.023 (0.080)	-0.004 (0.054)	-0.014 (0.054)
Mixed Auth. (other) $\times$ Female Ref.	0.035 (0.045)	0.023 (0.047)	0.021 (0.029)	0.009 (0.030)
Undetermined Auth. $\times$ Female Ref.	0.073 (0.092)	0.055 (0.093)	-0.019 (0.060)	-0.029 (0.060)
<i>Gender Interactions $\times$ Publication</i>				
All Female Auth. $\times$ Female Ref. $\times$ 3+ Publications for Ref.	0.039 (0.095)	0.034 (0.095)	0.097 (0.060)	0.094 (0.060)
Mixed Auth. (F-senior) $\times$ Female Ref. $\times$ 3+ Publications for Ref.	-0.032 (0.115)	-0.026 (0.114)	0.003 (0.074)	0.004 (0.074)
Mixed Auth. (other) $\times$ Female Ref. $\times$ 3+ Publications for Ref.	-0.098 (0.070)	-0.102 (0.070)	-0.047 (0.044)	-0.047 (0.044)
Undetermined Auth. $\times$ Female Ref. $\times$ 3+ Publications for Ref.	-0.070 (0.146)	-0.042 (0.148)	0.048 (0.088)	0.059 (0.089)
Paper Fixed Effects	Yes	Yes	Yes	Yes
Control for Referee Pub.	Yes	Yes	Yes	Yes
Control for Referee Pub.*Author Gender	No	Yes	No	Yes
Control for Referee Gender * Author Pub	No	Yes	No	Yes
Control for Referee Gender * Field	No	Yes	No	Yes
N	38,840	38,840	38,840	38,840
R-squared	0.51	0.51	0.45	0.45

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

Online Appendix Table 5. Citations and Editor Decision, Impact of Controls

	OLS Models for Asinh of GS Citations					Probit Models for Receiving R&R Dec.				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
<i>Authors' Genders</i>										
All Female	-0.11 (0.05)	0.14 (0.05)	0.22 (0.05)	0.17 (0.05)	0.19 (0.05)	-0.08 (0.06)	0.02 (0.06)	0.01 (0.07)	-0.01 (0.07)	0.01 (0.07)
Mixed-Gender Author Team senior author female	0.33 (0.07)	0.14 (0.07)	0.08 (0.07)	0.04 (0.07)	0.04 (0.07)	0.15 (0.07)	0.11 (0.07)	0.14 (0.07)	0.13 (0.07)	0.13 (0.07)
Mixed, other	0.24 (0.04)	0.13 (0.04)	0.01 (0.04)	0.00 (0.04)	-0.00 (0.04)	0.04 (0.05)	-0.00 (0.05)	-0.01 (0.05)	-0.01 (0.05)	0.02 (0.05)
Undetermined	-0.37 (0.05)	-0.24 (0.05)	-0.30 (0.05)	-0.27 (0.05)	-0.27 (0.05)	-0.08 (0.08)	-0.04 (0.08)	-0.03 (0.08)	-0.03 (0.09)	0.04 (0.08)
<i>Author Publications in 35 High-Impact Journals (Max across Authors)</i>										
1 Publication		0.37 (0.04)	0.22 (0.05)	0.22 (0.05)	0.23 (0.05)		0.04 (0.05)	-0.06 (0.06)	-0.05 (0.06)	-0.05 (0.06)
2 Publications		0.59 (0.04)	0.33 (0.05)	0.32 (0.05)	0.33 (0.05)		0.19 (0.06)	-0.00 (0.07)	-0.01 (0.07)	-0.00 (0.08)
3 Publications		0.70 (0.03)	0.33 (0.06)	0.33 (0.06)	0.34 (0.06)		0.17 (0.06)	-0.11 (0.09)	-0.10 (0.09)	-0.12 (0.09)
4-5 Publications		0.95 (0.05)	0.43 (0.09)	0.42 (0.09)	0.43 (0.09)		0.33 (0.05)	-0.05 (0.11)	-0.05 (0.11)	-0.03 (0.11)
6+ Publications		1.15 (0.05)	0.36 (0.12)	0.33 (0.11)	0.34 (0.12)		0.42 (0.07)	-0.14 (0.13)	-0.15 (0.14)	-0.13 (0.14)
<i>Author Publications in 35 High-Impact Journals, Mean across Authors</i>										
Average Publications Across Coauthors			0.06 (0.02)	0.06 (0.02)	0.06 (0.02)			0.04 (0.03)	0.04 (0.03)	0.03 (0.03)
<i>Author Publications in Top 5 Journals (Max Across Authors)</i>										
1 Publication			0.29 (0.04)	0.23 (0.04)	0.23 (0.04)			0.22 (0.05)	0.19 (0.05)	0.20 (0.05)
2 Publications			0.41 (0.04)	0.30 (0.04)	0.30 (0.04)			0.27 (0.08)	0.22 (0.08)	0.23 (0.08)
3+ Publications			0.53 (0.07)	0.34 (0.06)	0.34 (0.06)			0.45 (0.08)	0.37 (0.07)	0.39 (0.07)
<i>Author Publications in 35 High-Impact Journals, 6-10 years ago (Max Across Authors)</i>										
1-3 Publications			-0.10 (0.03)	-0.07 (0.03)	-0.07 (0.03)			0.15 (0.05)	0.17 (0.05)	0.17 (0.05)
4+ Publications			0.03 (0.05)	0.04 (0.04)	0.05 (0.04)			0.11 (0.06)	0.12 (0.06)	0.12 (0.06)
<i>Rank of Authors' Institution</i>										
US: 1-10				0.43 (0.04)	0.43 (0.04)				0.21 (0.05)	0.24 (0.06)
US: 11-20				0.29 (0.05)	0.29 (0.05)				0.18 (0.05)	0.19 (0.05)
Europe: 1-10				0.32 (0.04)	0.31 (0.04)				0.10 (0.06)	0.08 (0.06)
Rest of World: 1-5				-0.16 (0.09)	-0.17 (0.08)				0.10 (0.09)	0.09 (0.09)
R&R Indicator (Mechanical Publ. Effect)	-0.02 (0.14)	0.05 (0.14)	0.03 (0.13)	0.05 (0.14)	0.15 (0.14)					
Control Function for Selection (Value Added of the Editor)	0.42 (0.08)	0.33 (0.08)	0.33 (0.08)	0.30 (0.08)	0.25 (0.08)					
Editor Leave-out-Mean R&R Rate						3.38 (0.71)	3.41 (0.73)	3.39 (0.73)	3.42 (0.72)	3.44 (0.71)
Controls for Referee Rec., # Authors, Field & Gender-Field, and Journal-Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls for No. of Authors	No	No	Yes	Yes	Yes	No	No	Yes	Yes	Yes
Interactions of Field with Year and Gender-Field Variables	No	No	No	No	Yes	No	No	No	No	Yes
R ² / pseudo R ²	0.22	0.27	0.28	0.29	0.30	0.48	0.49	0.50	0.50	0.52

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

Online Appendix Table 6. Models of Alternative Measures of Citations

	OLS Model for Log(1+GS Citations)	Poisson Model for Number of GS Citations	OLS Model for GS Citation Percentile	Probit Model for Top Group of GS Citations	Probit Model for Top 2% of GS Citations	Tobit Model for asinh(GS Citations) Right- Censored at 100 Citations	OLS Model for Asinh Citations for Papers With Nonzero Cites	Tobit Model for asinh(GS Citations) All Years	Tobit Model for asinh(GS Citations) 2006-2008	Tobit Model for asinh(SSCI Citations) 2006-2008
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
<i>Authors' Genders</i>										
All Female	0.19 (0.05)	0.19 (0.06)	3.60 (0.85)	0.13 (0.05)	0.17 (0.09)	0.22 (0.05)	0.14 (0.04)	0.27 (0.07)	0.31 (0.13)	0.32 (0.15)
Mixed-Gender Author Team senior author female	0.05 (0.06)	0.07 (0.06)	0.51 (1.10)	0.10 (0.06)	0.07 (0.11)	0.07 (0.07)	0.10 (0.06)	0.05 (0.08)	0.14 (0.14)	-0.04 (0.27)
Mixed, other	0.01 (0.03)	0.03 (0.05)	0.07 (0.64)	-0.00 (0.03)	0.00 (0.05)	0.01 (0.04)	0.01 (0.03)	0.01 (0.05)	0.12 (0.10)	-0.02 (0.11)
Undetermined	-0.27 (0.04)	-0.32 (0.07)	-5.05 (0.81)	-0.19 (0.07)	-0.18 (0.11)	-0.31 (0.05)	-0.19 (0.07)	-0.41 (0.07)	-0.66 (0.17)	-0.61 (0.27)
<i>Fractions of Referee Recommendations</i>										
Reject	0.54 (0.05)	0.48 (0.09)	10.55 (0.95)	0.30 (0.08)	0.30 (0.13)	0.65 (0.06)	0.43 (0.06)	0.86 (0.08)	0.60 (0.18)	0.80 (0.21)
No Recommendation	0.84 (0.09)	0.84 (0.12)	16.04 (1.52)	0.55 (0.11)	0.51 (0.21)	1.00 (0.11)	0.79 (0.11)	1.26 (0.12)	1.04 (0.20)	1.58 (0.26)
Weak R&R	1.24 (0.09)	1.14 (0.11)	23.32 (1.48)	0.79 (0.11)	0.72 (0.18)	1.46 (0.11)	0.98 (0.10)	1.84 (0.11)	1.58 (0.17)	1.58 (0.28)
R&R	1.61 (0.12)	1.39 (0.13)	30.57 (1.96)	1.09 (0.14)	0.76 (0.21)	1.93 (0.13)	1.23 (0.15)	2.39 (0.14)	2.08 (0.25)	1.90 (0.37)
Strong R&R	1.94 (0.20)	1.73 (0.18)	36.48 (3.10)	1.21 (0.21)	0.94 (0.26)	2.30 (0.24)	1.52 (0.22)	2.83 (0.23)	2.46 (0.36)	2.48 (0.51)
Accept	1.99 (0.18)	1.84 (0.17)	36.46 (2.47)	1.34 (0.20)	1.19 (0.26)	2.37 (0.20)	1.52 (0.22)	2.87 (0.19)	2.81 (0.28)	3.17 (0.39)
<i>Author Publications in 35 High-Impact Journals</i>										
1 Publication	0.25 (0.04)	0.21 (0.05)	4.52 (0.70)	0.19 (0.05)	0.20 (0.11)	0.29 (0.04)	0.15 (0.03)	0.38 (0.05)	0.35 (0.12)	0.46 (0.13)
2 Publications	0.42 (0.03)	0.39 (0.04)	7.52 (0.58)	0.32 (0.05)	0.35 (0.07)	0.50 (0.04)	0.27 (0.03)	0.62 (0.05)	0.72 (0.10)	0.86 (0.18)
3 Publications	0.50 (0.03)	0.43 (0.03)	9.08 (0.55)	0.32 (0.05)	0.39 (0.08)	0.59 (0.04)	0.32 (0.04)	0.72 (0.05)	0.78 (0.10)	0.76 (0.16)
4-5 Publications	0.70 (0.05)	0.64 (0.07)	12.11 (0.80)	0.50 (0.05)	0.57 (0.08)	0.82 (0.06)	0.50 (0.04)	0.97 (0.07)	1.02 (0.11)	1.24 (0.14)
6+ Publications	0.86 (0.04)	0.79 (0.05)	14.82 (0.76)	0.67 (0.05)	0.78 (0.07)	1.02 (0.05)	0.70 (0.04)	1.15 (0.06)	1.14 (0.10)	1.23 (0.19)
R&R Indicator (Mechanical Publ. Effect)	0.11 (0.13)	0.09 (0.11)	-0.69 (2.24)	0.21 (0.13)	0.33 (0.18)	0.11 (0.15)	0.21 (0.12)	-0.10 (0.16)	0.12 (0.26)	1.63 (0.40)
Control Function for Selection (Value Added of the Editor)	0.27 (0.08)	0.27 (0.06)	5.48 (1.28)	0.17 (0.08)	0.11 (0.10)	0.32 (0.09)	0.19 (0.07)	0.44 (0.09)	0.39 (0.15)	0.07 (0.23)
Controls for No. of Authors, Field, Gender- Field, Journal-Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	15,147	15,147	15,147	15,147	15,147	15,147	15,147	15,147	4,507	4,507
R ² / pseudo R ²	0.28	0.352	0.20	0.15	0.16	0.49	0.28	0.49	0.49	0.49

Notes. In columns 1-7, the sample is 15,147 non-desk-rejected papers with at least two referees assigned. Columns 8 and 9 restricted to years 2006-08. Standard errors clustered by editor in parentheses.

Online Appendix Table 7. Citations and Editor Decision, Results Split by Number of Authors

Number of Authors:	OLS Models for Asinh of Google Scholar Citations			Probit Models for Receiving Revise-and-Resubmit		
	1 Author	2 Authors	3+ Authors	1 Author	2 Authors	3+ Authors
	(1)	(2)	(3)	(4)	(5)	(6)
<i>Authors' Genders</i>						
All Female	0.17 (0.07)	0.34 (0.08)	0.06 (0.23)	0.11 (0.08)	-0.23 (0.12)	-0.15 (0.49)
Mixed-Gender Author Team senior author female		-0.08 (0.10)	0.23 (0.09)		0.12 (0.11)	0.10 (0.10)
Mixed, other		0.01 (0.06)	0.03 (0.05)		-0.13 (0.08)	0.05 (0.07)
Undetermined	-0.22 (0.12)	-0.28 (0.08)	-0.37 (0.09)	-0.04 (0.20)	-0.10 (0.14)	-0.00 (0.15)
<i>Fractions of Referee Recommendations</i>						
Reject	0.79 (0.08)	0.48 (0.10)	0.67 (0.14)	0.75 (0.29)	0.83 (0.27)	1.01 (0.24)
No Recommendation	1.07 (0.14)	0.77 (0.17)	1.20 (0.24)	2.60 (0.34)	2.86 (0.26)	2.94 (0.23)
Weak R&R	1.52 (0.12)	1.36 (0.15)	1.40 (0.19)	3.14 (0.33)	3.18 (0.25)	3.33 (0.24)
R&R	2.37 (0.22)	1.57 (0.18)	1.75 (0.18)	4.85 (0.40)	4.69 (0.31)	4.62 (0.24)
Strong R&R	2.99 (0.31)	1.82 (0.28)	2.02 (0.31)	5.88 (0.48)	5.71 (0.33)	5.43 (0.32)
Accept	2.55 (0.28)	1.99 (0.29)	2.35 (0.36)	5.35 (0.37)	5.52 (0.33)	5.53 (0.30)
<i>Author Publications in 35 High-Impact Journals (Max across Authors)</i>						
1 Publication	0.36 (0.07)	0.28 (0.06)	0.21 (0.10)	0.03 (0.09)	0.12 (0.09)	-0.09 (0.14)
2 Publications	0.46 (0.09)	0.49 (0.06)	0.51 (0.09)	0.19 (0.14)	0.06 (0.10)	0.36 (0.15)
3 Publications	0.40 (0.13)	0.60 (0.06)	0.65 (0.08)	-0.00 (0.15)	0.26 (0.09)	0.09 (0.14)
4-5 Publications	0.83 (0.10)	0.76 (0.07)	0.88 (0.11)	0.41 (0.14)	0.39 (0.10)	0.22 (0.14)
6+ Publications	0.68 (0.19)	0.98 (0.08)	1.12 (0.10)	0.26 (0.21)	0.45 (0.10)	0.39 (0.12)
R&R Indicator (Mechanical Publ. Effect)	0.15 (0.25)	0.19 (0.21)	-0.10 (0.23)			
Control Function for Selection (Value Added of the Editor)	0.37 (0.15)	0.22 (0.13)	0.38 (0.14)			
Editor Leave-out-Mean R&R Rate				3.27 (1.20)	4.09 (1.00)	2.80 (0.95)
Controls for Field & Gender-Field Ctrls	Yes	Yes	Yes	Yes	Yes	Yes
Indicators for Journal-Year	Yes	Yes	Yes	Yes	Yes	Yes
Indicator for 4+ Authors	-	-	Yes	-	-	Yes
N	4,639	6,406	4,102	4,639	6,406	4,102
R ² / pseudo R ²	0.26	0.25	0.27	0.53	0.50	0.48

Notes. Dependent variable for OLS models in Columns 1-3 is asinh of Google Scholar citations. Dependent variable in probit models in Columns 4-6 is indicator for receiving revise and resubmit decision. The control function for selection in Columns 1-3 is calculated using predicted probabilities based on Columns 4-6. Standard errors clustered by editor in parentheses.

Online Appendix Table 8. Citations and Editor Decision, Heterogeneity

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

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

Online Appendix Table 9. Summary Statistics for Published Papers and Accepted Papers

Sample:	Accepted Papers in Editorial Express						Publications in our 4 Journals, 2008-2015					
	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)
<i>Google Scholar Citations</i>												
Asinh Citations	4.09	4.16	4.15	4.19	3.44	4.09	5.26	5.18	5.81	5.40	4.90	5.30
	(1.68)	(1.66)	(1.73)	(1.69)	(1.85)	1.69	(1.20)	(1.04)	(1.19)	(1.17)	(1.19)	1.19
<i>Author Publications in 35 high-impact journals</i>												
Publications: 0	0.19	0.50	0.00	0.11	0.48	0.20	0.14	0.47	0.00	0.07	0.10	0.14
Publications: 1	0.14	0.21	0.08	0.11	0.15	0.14	0.15	0.21	0.08	0.11	0.00	0.14
Publications: 2	0.11	0.10	0.19	0.14	0.02	0.11	0.12	0.13	0.18	0.15	0.10	0.13
Publications: 3	0.12	0.04	0.19	0.13	0.10	0.12	0.12	0.09	0.19	0.14	0.20	0.13
Publications: 4-5	0.19	0.07	0.16	0.20	0.10	0.18	0.20	0.07	0.24	0.19	0.20	0.19
Publications: 6+	0.24	0.07	0.38	0.32	0.15	0.25	0.26	0.02	0.31	0.34	0.40	0.27
<i>Number of Authors</i>												
1 author	0.31	0.71	0.00	0.00	0.29	0.26	0.25	0.64	0.00	0.00	0.00	0.21
2 authors	0.45	0.26	0.43	0.40	0.31	0.42	0.49	0.34	0.47	0.40	0.50	0.46
3 authors	0.19	0.03	0.39	0.41	0.33	0.23	0.21	0.02	0.36	0.38	0.40	0.24
4+ authors	0.05	0.00	0.18	0.19	0.08	0.08	0.06	0.00	0.16	0.22	0.10	0.09
<i>Field of Paper</i>												
Development	0.05	0.10	0.07	0.05	0.04	0.05	0.05	0.12	0.08	0.07	0.07	0.06
Econometrics	0.05	0.05	0.06	0.06	0.13	0.06	0.05	0.01	0.06	0.04	0.13	0.05
Finance	0.06	0.04	0.06	0.04	0.05	0.05	0.06	0.04	0.09	0.03	0.09	0.06
Health, Urban, Law	0.03	0.10	0.04	0.06	0.01	0.04	0.06	0.09	0.06	0.08	0.01	0.07
History	0.02	0.02	0.02	0.02	0.01	0.02	0.03	0.03	0.01	0.02	0.00	0.02
International	0.06	0.06	0.14	0.05	0.07	0.06	0.05	0.05	0.12	0.06	0.03	0.06
Industrial Organization	0.05	0.07	0.02	0.06	0.06	0.05	0.08	0.11	0.07	0.08	0.01	0.08
Lab/Experiments	0.02	0.02	0.05	0.03	0.01	0.02	0.00	0.00	0.00	0.01	0.00	0.00
Labor	0.10	0.12	0.14	0.13	0.13	0.11	0.16	0.25	0.18	0.22	0.23	0.18
Macro	0.10	0.08	0.07	0.06	0.11	0.09	0.10	0.05	0.05	0.07	0.12	0.09
Micro	0.11	0.04	0.09	0.09	0.12	0.10	0.12	0.08	0.10	0.11	0.06	0.11
Public	0.06	0.06	0.05	0.03	0.02	0.05	0.05	0.06	0.04	0.03	0.00	0.04
Theory	0.10	0.05	0.05	0.08	0.04	0.09	0.13	0.05	0.08	0.08	0.17	0.11
Unclassified	0.05	0.03	0.04	0.04	0.04	0.04	0.07	0.06	0.07	0.11	0.08	0.08
Missing Field	0.17	0.20	0.13	0.23	0.15	0.18	-	-	-	-	-	-
<i>Gender-Field Variables</i>												
Share female in fields	0.15	0.18	0.16	0.16	0.14	0.15	0.17	0.21	0.20	0.20	0.18	0.18
Gender-topic fields	0.03	0.06	0.07	0.05	0.05	0.04	0.08	0.10	0.11	0.11	0.13	0.09
Number of Observations	1175	107	77	302	52	1713	1190	97	74	348	10	1719
Share of Papers	0.69	0.06	0.04	0.18	0.03	1.00	0.69	0.06	0.04	0.20	0.01	1.00

Notes. Table presents information on mean characteristics of published papers from EconLit as well as accepted papers from Editorial Express. Author publications are based on publications in 35 high-impact journals (Online Appendix Table 1) in the 5 years prior to submission. In the case of multiple authors, the measure is the maximum over all coauthors. Field is based on JEL codes at paper submission. Indicators of fields for a paper that lists N codes are set to 1/N. The JELs from EconLit are those assigned by the journals whereas the JELs from Editorial Express are self-reported by the authors. To make the Gender-Field Variables comparable between the two groups, for just these variables, we have excluded papers that are missing JELs.

Online Appendix Table 10. Citations and Published Papers, Robustness

Dependent Variable:	Asinh of Web of Science (SSCI) Citations						
	All	0-1 Years	2-3 Years	4-5 Years	In Top	In Lower-	With
	Citations	After	After	After	Journals by	Ranked Journals	Unknown
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
<i>Authors' Genders</i>							
All Female	0.11 (0.13)	0.08 (0.13)	-0.01 (0.14)	0.09 (0.19)	0.06 (0.14)	0.01 (0.13)	0.17 (0.15)
Mixed-Gender Author Team senior author female	0.19 (0.16)	0.21 (0.17)	0.20 (0.15)	0.22 (0.20)	0.15 (0.17)	0.14 (0.17)	0.29 (0.17)
Mixed, other	0.05 (0.08)	-0.04 (0.08)	0.01 (0.08)	0.02 (0.11)	0.03 (0.08)	0.01 (0.08)	0.04 (0.09)
Undetermined	-0.58 (0.27)	-0.04 (0.27)	-0.62 (0.29)	-0.38 (0.36)	-0.11 (0.40)	-0.64 (0.28)	-0.99 (0.36)
<i>Author Publications in 35 High-Impact Journals</i>							
1 Publication	0.05 (0.11)	0.15 (0.11)	0.18 (0.11)	-0.12 (0.15)	0.13 (0.12)	0.01 (0.11)	0.10 (0.12)
2 Publications	0.15 (0.12)	0.26 (0.12)	0.29 (0.12)	-0.10 (0.16)	0.26 (0.13)	0.10 (0.12)	0.10 (0.14)
3 Publications	0.23 (0.12)	0.19 (0.12)	0.39 (0.12)	0.14 (0.16)	0.29 (0.13)	0.19 (0.12)	0.31 (0.14)
4-5 Publications	0.22 (0.11)	0.29 (0.11)	0.29 (0.12)	0.08 (0.15)	0.26 (0.12)	0.13 (0.12)	0.20 (0.13)
6+ Publications	0.45 (0.11)	0.40 (0.11)	0.54 (0.11)	0.36 (0.14)	0.46 (0.12)	0.36 (0.11)	0.56 (0.13)
Controls No. of Authors, for 2-Digit JEL, Gender-Field and Journal-Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Mean of Dependent Variable:	3.84	1.47	2.55	2.39	1.81	3.09	2.13
N	1,719	1,719	1,719	1,300	1,719	1,719	1,719
R ²	0.550	0.401	0.477	0.523	0.541	0.524	0.518

Notes. The sample is published papers from Econlit, as in Table VI. In Column 1, the dependent variable is asinh of web of science citations extracted in March 2019. In Columns 2-4, the dependent variables are asinh of web of science citations 0-1, 2-3, and 4-5 years out of publication. For 4-5 years out, we exclude papers from 2014 and 2015. In Columns 5-7, the dependent variables are asinh of web of science citations by papers published in journals with 5-year impact factors greater than 5 (Column 5), less than 3 (Column 6), and those with unknown impact factor (Column 7).

Online App. Table 11. Measurement Error in Citation and Gender Coding, Published Papers

Sample:	Published Papers, Econlit Sample				
	Missing Cites				
	No Meas. Error in Cites	Because of Title with Special Characters	Too Many Cites Because of Title Issue	Missing Cites Because of Author Name	
Panel A. Meas. Error in Cites	(1)	(2)	(3)	(4)	
All papers (N=1719)	N=1654 96.2%	N=31 1.8%	N=7 0.4%	N=27 1.6%	
By Authors' Genders					
All Male	N=1142	N=24	N=3	N=21	
All Female	N=92	N=3	N=0	N=2	
Mixed-Gender Author Team senior author female	N=72	N=1	N=1	N=0	
Mixed, other	N=338	N=3	N=3	N=4	
Undetermined	N=10	N=0	N=0	N=0	
Citations					
Asinh(Citation) as Scraped	5.26	0.00	6.92	0.68	
Asinh(Citation) as Corrected		4.94	5.74	5.00	
Example(s):		"&" in title, e.g., "Hybrid R&D" "" inside title, e.g., "Jobs, Jobs, Jobs: A "New" Perspective on Protectionism"	Multiple papers by Douglas Bernheim with words in title "Behavioral Welfare Economics"	Special characters in author last name, e.g., Jordi Galí vs Jordi Gali	
Correct Author Gender is					
	All Male	All Female	Mixed Gender, Senior Female	Mixed Gender, Undetermi ned	
Panel B. Meas. Error in Gender	(1)	(2)	(3)	(4)	(4)
Initial Author Gender Coding					
All Male	1173	0	0	1	0
All Female	0	94	0	0	0
Mixed-Gender, senior author female	1	0	71	0	0
Mixed Gender, other	3	0	0	342	0
Undetermined	13	3	3	5	10

Notes. In Panel A, we present the gender breakdown of cases of measurement error in the Google Scholar citations, and reason that the citation changed. We also present the average asinh of citations before and after correcting the citations. In Panel B, we present the cases of measurement error in gender, with a tabulation of the initial gender coding and corrected gender coding.

Online Appendix Table 12. Editorial Delays, Robustness

	OLS Regression	
	Days Before Resub.	Days from Resub.
	(R&Rs)	to Accept
	(1)	(2)
<i>Authors' Genders</i>		
All Female	-10.65 (18.76)	16.17 (23.27)
Mixed-Gender Author Team senior author female	-11.32 (16.62)	-2.83 (20.03)
Mixed-Gender, other	1.36 (12.42)	13.24 (12.27)
Undetermined Gender Mix	-29.72 (24.52)	49.23 (38.79)
<i>Sample</i>		
Controls for Referee Recommendations	Yes	Yes
Controls for Author Pub., No. of Authors, Field & Gender-Field, Journal-Year f.e	Yes	Yes
Editor Fixed Effects	Yes	Yes
Mean of Dependent Variable:	245.1	229.1
N	1,668	1,673
R-squared	0.24	0.35

Notes. The sample is papers observations, including only R&R papers that are ultimately published. Clustered standard errors by editor in parentheses.

Online Appendix Table 13. Referee Acceptance, by Author and Referee Gender

Specification:	Linear Probability Model for Referee Accepting a Report Request			
	(1)	(2)	(3)	(4)
<i>Authors' Genders (Omitted: All Male Authors)</i>				
All Female Authors	0.009 (0.008)	0.009 (0.008)		
Mixed-Gender Author Team senior author female	0.033 (0.011)	0.029 (0.011)		
Mixed-Gender Author Team other	-0.000 (0.006)	0.000 (0.006)		
Undetermined Gender Team	-0.015 (0.010)	-0.007 (0.010)		
<i>Referee Gender (Omitted: Male Referee)</i>				
Female Referee			-0.000 (0.010)	-
<i>Gender Interactions</i>				
All Female Auth. X Female Ref.			0.010 (0.026)	-0.002 (0.028)
Mixed Auth. (senior-F) X Female Ref.			-0.008 (0.033)	-0.006 (0.034)
Mixed Auth. (other) X Female Ref.			-0.025 (0.019)	-0.020 (0.020)
Undetermined Auth. X Female Ref.			0.011 (0.038)	0.001 (0.039)
Paper Fixed Effects	No	No	Yes	Yes
Controls for Referee Publications	No	No	Yes	Yes
Controls for Author Publications, Number of Authors, Field, Gender-Field, and Journal-Year	Yes	Yes	-	-
Control for Referee Pub.*Author Gender	No	No	No	Yes
Control for Referee Gender * Author Pub	No	No	No	Yes
Control for Referee Gender * Field	No	No	No	Yes
N	60,445	60,445	60,445	60,445
R-squared	0.013	0.018	0.249	0.251

Notes. Standard errors clustered by paper in parentheses. The sample in each column is a referee-paper observation, including any referee invited to review a paper.

Online App. Table 14. Abstract Complexity, Impact of Author Team Gender

	Measure of Complexity of Abstract			
	Gunning Fog	Coleman- Liau	Gunning Fog	Coleman- Liau
	(1)	(2)	(3)	(4)
<i>Authors' Genders</i>				
All Female	-0.05 (0.07)	0.11 (0.05)	0.50 (0.30)	0.18 (0.19)
Mixed-Gender Author Team	0.29 (0.11)	0.12 (0.07)	0.07 (0.31)	0.09 (0.22)
senior author female	0.12 (0.06)	0.04 (0.04)	-0.13 (0.20)	0.11 (0.13)
Mixed, other	0.32 (0.08)	-0.02 (0.05)	-0.51 (0.33)	0.16 (0.25)
Undetermined				
<i>Author Publications in 35 High-Impact Journals (Max across Authors)</i>				
1 Publication	-0.10 (0.05)	0.05 (0.04)	-0.24 (0.23)	-0.19 (0.15)
2 Publications	-0.10 (0.07)	0.11 (0.04)	-0.11 (0.25)	-0.14 (0.26)
3 Publications	-0.21 (0.08)	-0.03 (0.05)	-0.43 (0.26)	-0.34 (0.20)
4-5 Publications	-0.23 (0.07)	-0.03 (0.05)	0.19 (0.24)	-0.04 (0.16)
6+ Publications	-0.18 (0.07)	-0.06 (0.05)	-0.11 (0.23)	-0.33 (0.15)
Sample	Rejected and Desk- Rejected Papers		R&R Papers Only	
Controls for Author Publications, Number of Authors, Field, Gender-Field, and Journal-Year	Yes	Yes	Yes	Yes
Mean of the Dependent Variable:	19.4	15.3	19.3	15.4
N	27,545	27,545	2,366	2,366
R-squared	0.02	0.02	0.04	0.04

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



The Equity in STEM “First Look”

Produced by the ARC Network, the Equity in STEM “First Look” is a quarterly digital publication that disseminates evidence-based systemic initiatives through multimedia elements, including presentations, promising practices, white papers, videos, podcasts and webinars. Each curated issue allows peers from the ARC Network Community to share content at all stages of development, facilitating early adoption and implementation and broadening our impact.



Funded by the National Science Foundation ADVANCE Program, Award HRD-1740860, the ADVANCE Resource and Coordination (ARC) Network seeks to achieve gender equity for faculty in higher education science, technology, engineering, and mathematics (STEM) disciplines. As the STEM equity brain trust, the ARC Network recognizes the achievements made so far while producing new perspectives, methods and interventions with an intersectional, intentional and inclusive lens. The leading advocate for women in STEM the Association for Women in Science (AWIS) serves as the backbone organization of the ARC Network. Learn more about the ARC Network at www.EquityInSTEM.org.

UNH UNBIASED: Leadership Development and Policy Change to Promote Institutional Transformation

Abstract at Time of Award

The University of New Hampshire (UNH) ADVANCE Institutional Transformation program, in keeping with the institution's mission of inclusive excellence, has as its overall goal to increase the number, retention and success of women faculty in the STEM disciplines. Based upon careful analysis of a University climate survey, several strategies are employed that include, but are not limited to strategic changes in recruitment and retention policies and practices; department chair professional development; formal mentoring policies; a wage equity analysis; and, flexible workplace policies. Coupled with these strategies is a rigorous social science study aimed at investigating the impact of gender equity institutional transformation on the representation and job satisfaction of women faculty at UNH. It is expected that results from this project and outcomes of the research study will seminally inform the current gender equity literature. Thus, this project has the potential to serve as a model for other institutions who identify comparable challenges in hiring, promoting, and, in general, advancing women STEM faculty.

Project Outcomes Report

The University of New Hampshire (UNH) is the state's flagship, mid-size, public, research-intensive university. In 2012, UNH received funding from the National Science Foundation's ADVANCE Institutional Transformation (IT) program to initiate sustainable institutional transformation in order to increase the number, retention, and success of women STEM faculty.

To increase the number of women STEM faculty through hiring, UNH ADVANCE developed GEAR UP (Gender Equity and Recruitment of Underrepresented People), a workshop whose purpose is to increase awareness of, and learn ways to overcome, unconscious bias in faculty searches. The program consists of an interactive theater session depicting a faculty search process followed by facilitated discussion. To-date, 319 faculty and staff involved in faculty hiring have attended GEAR UP. In workshop evaluations, 94% of participants indicate that they would recommend GEAR UP to colleagues. Analysis of annual faculty climate survey data over time reveal that GEAR UP significantly improved faculty perceptions of the search committee process as equitable. Importantly, women faculty perceptions of the fairness of the search process improved over time such that significant gender differences in 2013 were no longer significant in 2017. Interviews conducted by the External Evaluator show that, after having attended GEAR UP, search committee members reported giving extra attention to using explicit, objective criteria in assessing applicants and that GEAR UP provided participants with a vocabulary for addressing inequities that occurred in the search process.

The impact of GEAR UP on faculty hiring is also evident in changes in faculty composition from 2012 to 2018. In 2012, 19.5% of the 251 STEM faculty members were women. By Academic

Year 2018-2019, the percent of women STEM faculty had increased by 46.1% to 28.5% of STEM faculty. An increase of 110% in the number of STEM women at the assistant professor rank from 10 in 2012 to 21 in 2018-2019 can be attributed, at least in part, to a change in hiring practices.

In 2012, only 12 out of 44, or 27.3% of STEM faculty women were full professors rank, while 103 out of 171 (or 60.2%) of STEM faculty men were full professors. In contrast, in 2018-2019, 20 out of 61 (or 32.8%) of STEM faculty women are at the full professor rank and 76 out of 153 (or 49.7%) of STEM faculty men are full professors. This change in the make-up of faculty at the top level of their profession, aided by an increase in the number of promotions for women and retirements for men, is likely to have a sustained effect through improved gender balance among faculty in leadership positions.

To improve support and department-level climate for STEM faculty women, UNH ADVANCE developed and implemented a department chair professional development program in the spring of 2014. Participation was organized by cohort in order to facilitate evaluation of any resulting change in workplace climate. Promising early faculty climate survey results show that in the two years following the training, women faculty in the first cohort reported significant increases over time in job satisfaction and perceived fairness of department processes. In addition, women faculty perceptions of the fairness of department processes also improved over time such that significant gender differences in 2013 are no longer significant in 2017. These results offer a preliminary indication of reduced gender bias and improved department climate during the course of the grant.

UNH ADVANCE developed and implemented “Pathways to Tenure” a three-session mentoring workshop for assistant professors. Workshop goals include understanding promotion and tenure requirements, finding advice from department- and college-level committees and navigating work-life balance. Since its inception in 2014, 87 assistant professors attended the workshop. Pre- and post-workshop surveys revealed significant improvements in attendee knowledge of the college level promotion and tenure processes and of UNH’s tenure clock extension and family leave policies. Overall, 93% of post-test respondents felt better prepared for the promotion and tenure process and agreed or strongly agreed that they would recommend the program to their colleagues.

UNH ADVANCE benchmarked other universities’ policies and worked closely with the UNH Chapter of the American Association of University Professors during negotiations in 2015 to incorporate contract provisions to facilitate faculty work-family balance. For example, in the current collective bargaining agreement, (1) a tenure clock extension for parental leave now automatically applies unless a qualifying faculty member explicitly undertakes to opt-out of the benefit; (2) 12 weeks parental leave is now provided to both parents even if they both work at UNH; and (3) a new modified duties option is provided for faculty members dealing with qualifying events under the Family and Medical Leave Act.

Overall, ADVANCE has been a transformative force across the UNH campus and will be sustained through institutional support of the UNH ADVANCE office within the Provost's Office.

Women in Economics: Stalled Progress

Shelly Lundberg and Jenna Stearns

Although women are still a minority in the economics profession, female representation in the discipline has increased slowly over the past century. By the mid-2000s, just under 35 percent of PhD students and 30 percent of assistant professors were female, but these numbers have remained roughly constant ever since. This is not the first time progress on the path to gender equality in economics has stalled: women were more prominent as researchers in the early years of the 20th century than they were mid-century. Listings of dissertations in progress in the *American Economic Review* show that women were writing 6 percent of US PhD dissertations in 1912, rising to a peak of nearly 20 percent in 1920 but then falling back to 7 percent by 1940 (Forget 2011). Forget (2011) links the decline in female representation in academic economics to the emergence of home economics and social work as academic fields, the expansion of employment opportunities in government, and increased hostility and overt discrimination in economics departments. Cherrier (2017b) draws a parallel between these trends in economics and the defeminization of computer science as this field became increasingly professionalized, “scientized,” and lucrative after the mid-1980s.

■ Shelly Lundberg is the Leonard Broom Professor of Demography, University of California Santa Barbara, Santa Barbara, California. She is also a Research Fellow, Institute for the Study of Labor (IZA), Bonn, Germany, and Adjunct Professor of Economics, University of Bergen, Bergen, Norway. Jenna Stearns is an Assistant Professor of Economics, University of California Davis, Davis, California. Lundberg is the corresponding author at slundberg@ucsb.edu.

† For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at <https://doi.org/10.1257/jep.33.1.3>

doi=10.1257/jep.33.1.3

Common explanations for women's underrepresentation in economics in the mid-20th century included comparative advantage and diverging preferences by gender. By the early 1970s, however, overt discrimination was blamed for "the bizarre and irrational underrepresentation of women in the economics profession" (CSWEP 1973). The contested establishment of the Committee on the Status of Women in the Economics Profession (CSWEP) at the 1971 American Economic Association (AEA) business meeting took place in the wake of public discussion and government action on discrimination, actions by other professional associations to increase the representation of women, and growing interest in discrimination as an economic phenomenon with the early work by Becker and Arrow (Cherrier, Chassonnery-Zaigouche, and Singleton 2018). A Caucus of Women Economists drafted resolutions requiring the AEA to adopt "a positive program to eliminate sex discrimination." The resolutions were presented at the business meeting, where they provoked heated debate and several speeches in opposition, but were approved by a vote of the attending membership. The room had been packed by progressive economists prior to the vote, according to a first-hand account by Strober (2016, chap. 6). Though the beginning statement "Resolved that the American Economic Association declares that economics is not a man's field," was amended to insert "not exclusively," the resolutions were adopted in full, including the establishment of CSWEP (Cherrier 2017b). Feminist activism scored similar successes throughout the academy during this period: women's committees were established in the American Sociological Association in 1970, and in the American Physical Society in 1972.

Despite large gains in female representation in economics in the 1970s and 1980s, reactions to women's progress were mixed. In the Fall 1998 issue of this journal, the 25th anniversary of CSWEP was commemorated with a symposium that reflected contrasting views of efforts to diversify the profession. The various contributions reviewed women's progress in economics favorably but expressed concern about the implications of low representation of women among economics undergraduates (Bartlett 1998); critiqued CSWEP's nonmilitancy and offered unfavorable comparisons with more activist women's committees in other fields (Bergmann 1998); and asserted that, partly as a result of CSWEP's activities, the "pendulum has probably swung too far so that men are the ones currently being discriminated against" (Friedman 1998).

Since then, women's progress in academic economics has slowed, with virtually no improvement in the female share of junior faculty or graduate students in decades. Little consensus has emerged as to why, though there has been a renewal of widespread interest in the status and future of women in economics and of the barriers they face to professional success. In this paper, we first document trends in the gender composition of academic economists over the past 25 years, the extent to which these trends encompass the most elite departments, and how women's representation across fields of study within economics has changed. We then review the recent literature on other dimensions of women's relative position in the discipline, including research productivity and income, and assess evidence on the

barriers that female economists face in publishing, promotion, and tenure. While differences in preferences and constraints may directly affect the relative productivity of men and women, productivity gaps do not fully explain the gender disparity in promotion rates in economics. Furthermore, the progress of women has stalled relative to that in other disciplines in the past two decades. We propose that differential assessment of men and women is one important factor in explaining this stalled progress, reflected in gendered institutional policies and apparent implicit bias in promotion and tenure processes.

Women in PhD-granting Economics Departments, 1972–2017

In 1972 and 1973, the Committee on the Status of Women in the Economics Profession conducted surveys of economics departments “to remedy the total lack of information on how many women economists exist or are currently being trained” (Bell 1973). This task was assumed by the American Economic Association in 1974, and questions about faculty and graduate student gender were combined with other data requests to form the Universal Academic Questionnaire (UAQ) sent to academic departments. The UAQ provided the data for CSWEP’s reports on the status of women in economics until 1993, when the committee resumed their separate survey to improve response rates. This survey gathers information each fall on the gender composition of new and graduating PhD students; faculty at the assistant, associate, and full professor levels; nontenure track faculty; and senior undergraduate majors. Most of the data presented in this section combines the CSWEP and UAQ data for PhD-granting departments from 1993 to 2017.¹

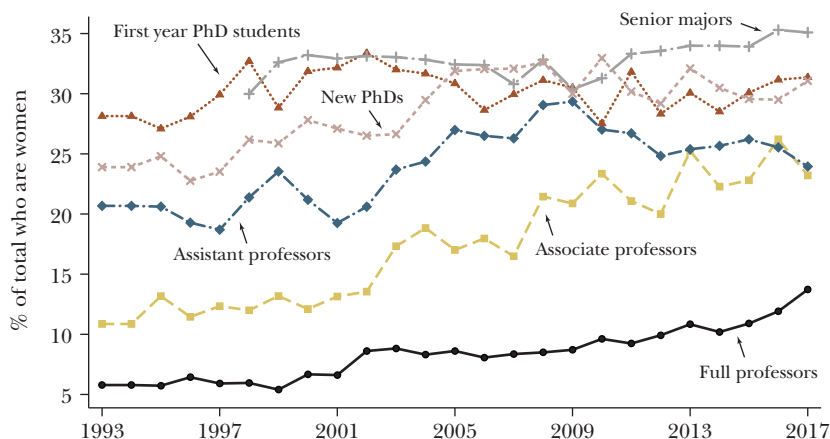
We can provide a longer-term perspective for one important set of departments. The microdata from the first CSWEP survey has been lost, but the 1972 CSWEP Annual Report, published in the *AER Papers and Proceedings* issue, includes aggregate results for one identifiable group of 43 departments—the Chairman’s Group. This group was known informally as “the cartel,” because the chairs met every year for breakfast at the ASSA meetings and discussed planned salary offers for new assistant professors. These departments are listed in the report, and with the exception of the University of Rochester, all of them responded to the initial survey.² The departments in the Chairman’s Group granted about two-thirds of US economics PhDs in the early 1970s, and we can track the faculty and graduate student gender composition in this set of highly ranked departments over a 45-year period.

¹ Response rates to the CSWEP survey of PhD-granting departments have been 100 percent in recent years, but below that prior to 2015—nonresponses are replaced by UAQ data when possible. The data for 2000 has been lost. The cleaned data were produced by the Inter-university Consortium for Political and Social Research (ICPSR) under the direction of Margaret Levenstein, and is available to researchers through ICPSR. About 4 percent of the observations are imputed.

² The report also includes aggregate gender ratios for “all departments” based on 397 questionnaires returned out of 1364 questionnaires sent (Bell 1973).

Figure 1

Representation of Women among First-Year PhD Students, New PhDs, and Faculty by Rank for the Chairman's Group of Departments, 1993–2017



Source: Authors, using data from CSWEP and from the UAQ for PhD-granting departments from 1993 to 2017.

Note: The Chairman's Group consists of Brown University, University of California—Berkeley, University of California—Davis, University of California—Los Angeles, Carnegie Mellon University, University of Chicago, University of Colorado, Columbia University, Cornell University, Duke University, University of Florida, Harvard University, University of Illinois, Indiana University, Iowa State University, Johns Hopkins University, University of Maryland, Massachusetts Institute of Technology, University of Michigan, Michigan State University, University of Minnesota, New York University, State University of New York—Buffalo, University of North Carolina—Chapel Hill, Northwestern University, Ohio State University, University of Pennsylvania, University of Pittsburgh, Princeton University, Purdue University, University of Rochester, University of Southern California, Stanford University, Texas A & M University, University of Texas—Austin, Vanderbilt University, University of Virginia, University of Washington—Seattle, Washington State University, Washington University in St. Louis, Wayne State University, University of Wisconsin, and Yale University.

Substantial progress was made during the 1970s and 1980s in the representation of female faculty within the Chairman's Group departments. In 1972, women accounted for only 2 percent of full professors, 4 percent of associate professors, and 9 percent of assistant professors. By the time the CSWEP survey was resumed in 1993, the fraction of full professors who were female had tripled to 6 percent, 11 percent of associate professors were women, and the female share of assistant professors had more than doubled to 21 percent.

Figure 1 shows that the proportion of senior female faculty in the Chairman's Group continued to grow slowly from 1993 to the present. Among full professors, the female share increased from 6 percent to more than 13 percent, and among associate professors, from 11 to 23 percent. For assistant professors, however, the pattern is somewhat different: the share of women increased from 20 percent in 1993 to 29 percent in 2009, and then decreased over the past decade to 24 percent, leaving

little net growth at junior ranks over the past 24 years.³ Over the same period, there has also been little improvement in female representation among first-year PhD students, from 28 percent in 1993 to an average of 30 percent in the past five years. (During the 1990s, there was a consistent gap of a couple of percentage points between the female share of first-year graduate students and exiting PhDs five years later that seems to indicate higher attrition for female graduate students, but this gap disappeared by the entering class of 2000.) This stasis extends to undergraduate study of economics as well: the female share of senior economics majors has remained between 30 and 35 percent since the data series began in 1998. Progress towards gender equality at the intake levels of the profession appears to have ceased (with some deterioration for junior faculty), while women's representation at senior levels continues to rise, fueled for now by the entry of women into academic economics in past decades.

Although the Chairman's Group does not provide a complete picture of PhD departments, there are reasons to be particularly interested in the progress that women have made in elite departments. Economics is a very hierarchical social science (as discussed in this journal by Fourcade, Ollion, and Algan 2015), and a high fraction of both the articles published in top journals and the faculty who train PhD students come from the most highly-ranked departments. Figure 2 shows the 1993–2017 faculty and student data for the departments rated in the top 20 by US News and World Report. The data is a bit noisier for these smaller samples than for the Chairman's Group, but some trends are clear.

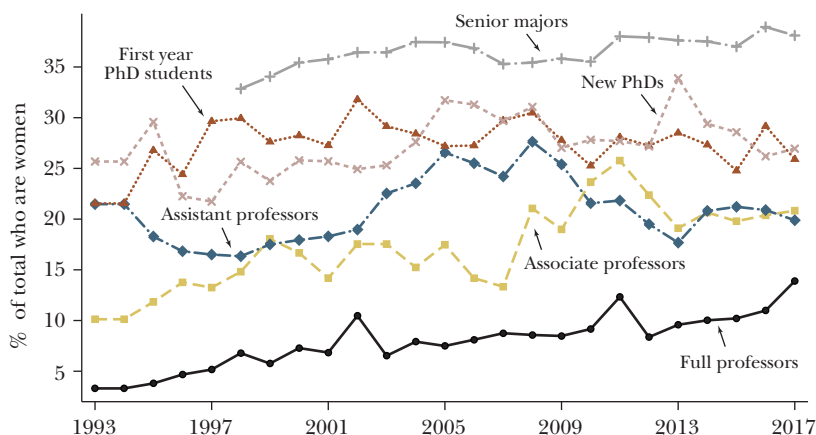
In top 20 programs, the representation of women among full professors was only 3 percent in 1993, grew slowly to 10 percent in recent years, and then rose to nearly 14 percent in 2017. The female fraction of associate professors (which grew steadily throughout this period in the Chairman's Group), increased from 10 percent to as high as 26 percent in 2011, but has declined in recent years to about 20 percent. Female representation among assistant professors stood at about 21 percent in 1993, reached a peak of 27.6 percent in 2008, and has since fallen back to 20 percent, meaning that no net progress has been made at the junior faculty level in top 20 departments over the past 24 years. These patterns are qualitatively similar if we look only at the top 10 programs in the US News and World Report Rankings as well.

To compare women's progress in economics to other academic disciplines, we have combined the data from the Committee on the Status of Women in the Economics Profession on the Chairman's Group with data on the share of female faculty by rank in top-50 departments for several science and social science disciplines. These data, for 2002, 2005, 2007, and 2012, come from the Nelson (2004) Diversity Surveys of department chairs, collected under the auspices of the University

³This decrease is not apparent in the data on all PhD-granting departments that is presented in the CSWEP annual report (CSWEP 2018), where the assistant professor gender ratio appears flat for the 2005–2017 period. A separate analysis confirms that, for the non-Chairman's Group departments (which tend to be lower-ranked than the Chairman's Group), female representation among assistant professors has continued to grow slowly.

Figure 2

Representation of Women among First-Year PhD Students, New PhDs, and Faculty by Rank: Top 20 Economics Departments, 1993–2017



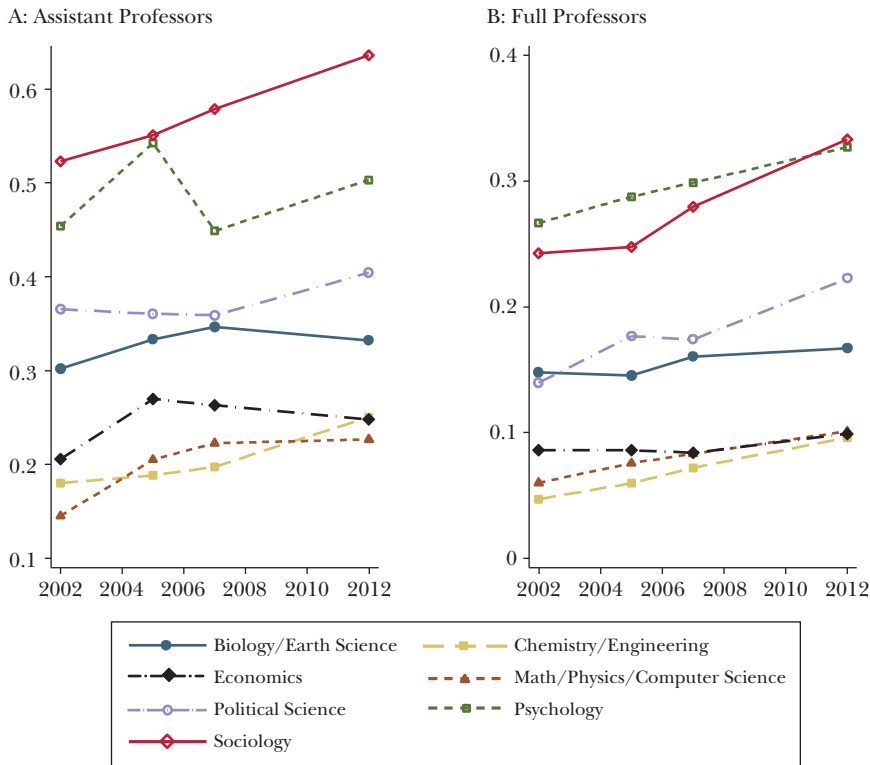
Source: Authors, using data from CSWEP and from the UAQ for PhD-granting departments from 1993 to 2017.

Note: The departments included are Brown University, Carnegie Mellon University, Columbia University, Cornell University, Duke University, Harvard University, Massachusetts Institute of Technology, New York University, Northwestern University, Princeton University, Stanford University, University of California—Berkeley, University of California—Los Angeles, University of California—San Diego, University of Chicago, University of Michigan—Ann Arbor, University of Minnesota, University of Pennsylvania, University of Wisconsin—Madison, and Yale University.

of Oklahoma. Top-50 departments are as ranked by the National Science Foundation according to field-specific research expenditures.⁴ Figure 3 shows trends in the share of female assistant and full professors across seven discipline groups. For ease of presentation, we combine data from chemistry and four types of engineering departments (chemical, civil, electrical, and mechanical). We also combine biology and earth science, and math, computer science, physics, and astrophysics. In general, there is an upward trend in the share of female faculty at all ranks over this ten-year period. Hard sciences have the lowest share of female professors at all ranks, while the social sciences have the highest. Economics remains solidly within the lowest group in terms of female faculty shares, alongside physics, math, and engineering, and far below the biological and other social sciences. At the senior level, economics seems to have lost some ground relative to other sciences during

⁴Comparable data on top-50 departments is not available going back further in time. Using the NSF Survey of Doctoral Recipients (SDR), Ginther and Kahn (2004) and Ceci, Ginther, Kahn, and Williams (2014) show trends in the share of female assistant professors and tenured faculty across disciplines since 1973. However, the SDR samples doctoral recipients from all US academic institutions, and is not necessarily representative of faculty at top departments.

Figure 3

Representation of Women in Top-50 Departments, 2002–2012*(share female)*

Source: Authors, using data from the Nelson Diversity Surveys and CSWEP.

this period. In all cases the share of women is decreasing with rank (note the y-axes for graphs A and B are different).

Do Women Study Different Fields of Economics than Men and Has the Distribution of Women across Fields Changed over Time?

While the survey data from the Committee on the Status of Women in the Economics Profession allow us to track the career progression of female academic economists over time, much less is known about another dimension of women's representation in economics—their distribution across fields of study. Understanding how trends in research areas differ for men and women may be important for understanding differential trends in publishing and tenure. Field choice may

affect entry into tenure-track positions in economics, publication rates, and the probability of publishing in top journals, all of which may also affect the probability of earning tenure. However, the limited evidence estimating differences in professional success across economics fields is mixed. Recent work shows that field choice explains a large share of the gender gap in research output (Ductor, Goyal, and Prummer 2018), while Ginther and Kahn (2004) find that broad fields are an insignificant predictor of tenure among a sample of assistant professors in 1989. Several recent papers in economics document the contemporary distribution of women across fields, but to our knowledge, the existing research cannot provide insight into how fields of study have changed over the past few decades.

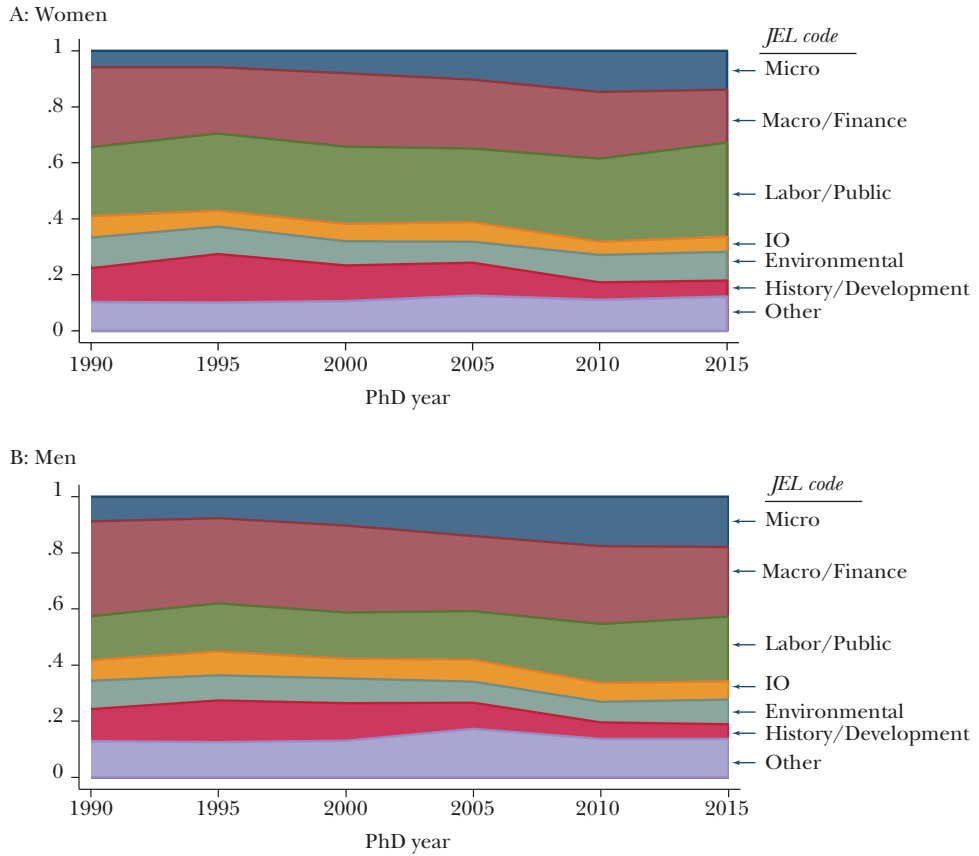
Using data from the National Bureau of Economic Research Summer Institute 2001–2016, Chari and Goldsmith-Pinkham (2017) show that the distribution of female economists at this event is not uniform across fields. Women are particularly scarce in macro and finance, and more abundant in labor and other applied microeconomic fields. Beneito, Pilar, Boscá, Ferri, and García (2018) use data from the annual AEA meetings from 2010–2016 to show the percentage of female authors in five subfields according to the *Journal of Economic Literature* subject codes of the sessions. For the most recent years, the authors also use machine learning to classify the paper abstracts by topic. Similar to Chari and Goldsmith-Pinkham (2017), they find that female representation is substantially lower in macro, finance, and mathematical and quantitative methods than in applied micro and other fields. An important caveat about these findings is that participation at both the AEA annual meetings and the NBER Summer Institute may be nonrepresentative across both gender and field, and again, little is known about how the gender composition across economic fields has changed over a longer period.

To provide a broader perspective on the evolution of women across fields and over time, we have collected information on recipients of PhDs in economics from 1991–2017, including the recipient's name and the JEL code of their dissertation. This information comes from the Doctoral Dissertations in Economics lists published annually in the *Journal of Economic Literature*, and represents almost all major PhD-granting departments in the United States. To classify the gender of each doctoral recipient, we use two databases that allow us to determine the probability that a given name is female: the Social Security Administration name files and the Genderize.io database for an international dimension. We match the first names in our data to these probabilities, and assign gender to those with a probability of being female that is above 0.8 or below 0.2.⁵ In total, we identify the gender

⁵The first database is the Social Security Administration name files, which include all names with at least five occurrences in a given year based on applications for a US Social Security card at birth. Because this data is only representative of US-born individuals, and a large share of PhD recipients in economics are foreign-born, we also match to the Genderize.io database, which contains over 200,000 distinct names from 79 countries. Both datasets contain the number of male and female incidences of the name. We designate a name as female if the probability that the name is female is higher than 0.8, and male if the probability is lower than 0.2. We are able to match 88.5 percent of the individuals in our data to a name in at least one of the two databases, and we assign a gender to 83 percent of the total sample.

Figure 4

Dissertation Topics of Women and Men by Year
(share)



Source: Authors, using data from the annual list of Doctoral Dissertations in Economics, 1991–2017.

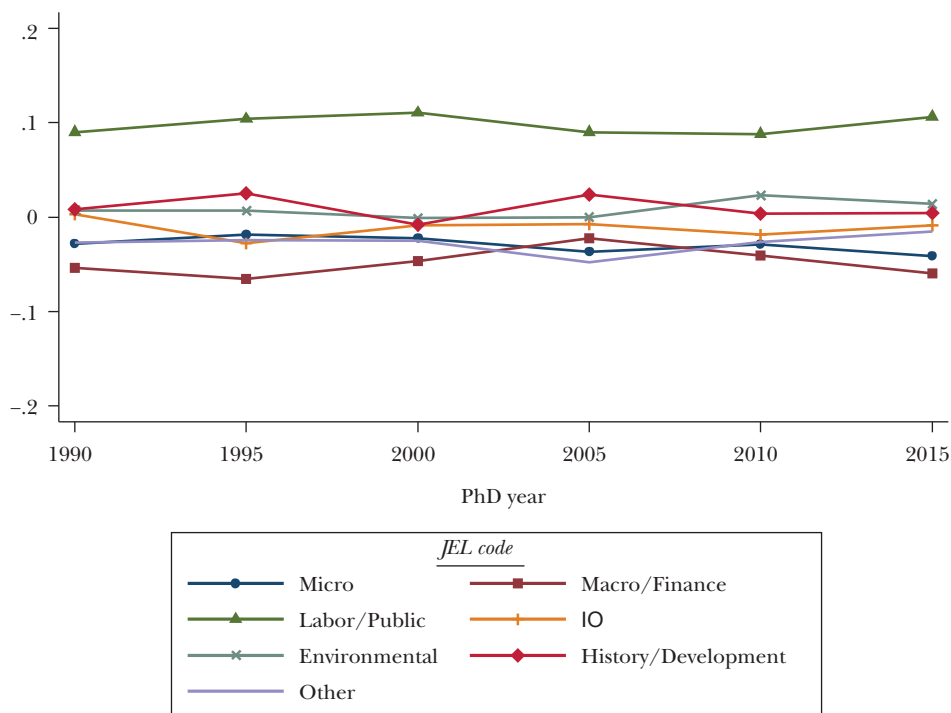
Note: Data collapsed into five-year bins for smoothness. The 1990 bin contains data from 1991–1994 and the 2015 bin contains data from 2015–2017; all other bins contain 5 years of data.

of 23,442 out of 28,209 individuals over 26 years. About 29 percent of new PhDs over this period are female, and the trend in the share of female PhD recipients over time mirrors the CSWEP data above.

Figures 4A and 4B show the distribution of women and men across seven fields over time.⁶ In general, the distributions of men and women across these fields are very similar. The higher representation of women in labor/public is apparent,

⁶We have collapsed the JEL codes into seven categories for ease of presenting results. “Micro” is JEL code D; “Macro/Finance” is codes E, F, and G; “Labor/Public” is H, I, and J; “IO” is L; “Environmental” is Q; “History/Development” is N and O; and “Other” contains the remaining JEL codes A, B, C, K, M, P, R, Y, and Z, which all represent a relatively small share of PhD dissertations.

Figure 5

Difference between Share of Women and Share of Men in Particular Fields of Economics

Source: Authors, using data from the annual list of Doctoral Dissertations in Economics, 1991–2017.

Note: Data was collapsed into five-year bins for smoothness. The 1990 bin contains data from 1991 to 1994 and the 2015 bin contains data from 2015 to 2017; all other bins contain five years of data.

but female economists are well-represented in all fields. In terms of changes over time, women are more likely to write dissertations in micro and labor/public than they were in the early 1990s, and somewhat less likely to study macro/finance and history/development. However, in large part these trends reflect broader trends in the profession, and very similar changes in field choice can be seen among men.

In fact, the distribution of PhD recipients across fields has not evolved differentially for men and women since the early 1990s. To show this more clearly, Figure 5 plots the difference between the share of women in a particular field and the share of men in that field over time. While it is certainly the case that women are more likely than men to study topics in labor and public economics and less likely to do dissertation research in macro and finance across the entire time period, there is virtually no evidence of differential trends. (Because of the gender imbalance in economics, there are still more men than women who graduate with a dissertation classified as labor or public every year.) It is not entirely clear why a higher fraction of women than men choose labor-oriented research topics. One commonly

discussed hypothesis is that women on average have stronger interests in studying individual behavior. A survey of AEA members in 2008 showed that, while there were no gender differences in responses to beliefs about core economic concepts, female economists are more likely to support the need for government intervention versus market solutions (May, McGarvey, and Whaples 2014). This bias in choice of field could be sustained over time if the research environment across different fields is an important factor in what graduate students choose to study; that is, the higher share of female faculty in labor economics might encourage female students to study labor through role model effects.

This lack of change in the relative gender composition across fields over time is important for two reasons. First, it suggests that as the share of female PhD recipients has risen, the more recent female cohorts are no different in terms of their broad research interests. Second, differential trends in field choice over time cannot explain the observed changes in the gender gap in the share of PhD recipients who become assistant professors and who are later tenured.

The graphs shown here use data starting in 1991 because this was the last time that the *Journal of Economic Literature* subject codes were substantially redesigned. Focusing on this period means that the JEL codes are comparable over time. However, it is possible to collect similar data going back further. For a longer-term perspective, we have also compiled data from the early 1970s. JEL codes in this period were not completely comparable with those used today. In particular, “core” areas of economic theory including micro and macro theory were categorized together in a “General Economics” category, though the applied categories are reasonably comparable for our purposes.⁷ There is still no evidence of differential trends by gender in these applied areas. In particular, the difference in the share of women compared to men who study labor and public economics has remained constant at about 0.1 since at least the early 1970s.

How Do Women’s Academic Careers in Economics Compare With Men’s?

Women’s representation in economics departments tends to fall as academic rank increases. As shown above (Figures 1 and 2), the female share of full professors in research-oriented departments ranges from 8 to 13 percent, from 20 to 25 percent for assistant professors, and from about 25 to 30 percent for PhD students. Simple “lock-step” models tracking cohorts of PhD recipients, reported annually by the Committee on the Status of Women in the Economics Profession, show a distinct drop-off from last-year-in-rank assistant professors to last-year-in-rank

⁷The full category name is “General Economics; including Economic Theory, History of Thought, Methodology, Economic History, and Economic Systems.” The comparable categories include environmental, development, IO, labor/public, and international economics. See Cherrier (2017a) for a history of JEL codes, including a list of categories in this time period.

associate professors for PhD cohorts from the mid-1980s through 2003 (CSWEP 2018). This suggests that the economics career pipeline is “leaky” at the stage when most academics receive tenure.

Studies using micro-data tend to confirm that something goes wrong for female economists at the tenure stage. Using longitudinal data on all AEA members from the 1960s through the 1980s, McDowell, Singell, and Ziliak (2001) find that women were less likely than men with similar characteristics to be promoted to both associate professor and full professor. However, they also find that women’s promotion prospects improved in the 1980s, leaving no unexplained gender differences in promotion for individuals observed in 1989. In contrast, Ginther and Kahn (in this journal, 2004) find clear evidence of a leaky pipeline in a sample restricted to AEA members who were assistant professors at PhD-granting departments in 1989—women in this sample were less likely to get tenure than men and took longer to achieve it. Ten years after receiving their PhDs, female economists were 21 percentage points less likely than men to have a tenured academic job. Differences in productivity, including number of publications, publication quality, and citations, explained only 30 percent of this promotion gap. In the same paper, Ginther and Kahn find a similar result using the 1972 to 1991 PhD cohorts from the National Science Foundation Survey of Doctoral Recipients, which has limited data on publication quality but does have information on family characteristics. Controlling for the presence of young children, which had an impact on promotion independent of productivity, leaves a substantial portion of the gender difference in tenure probabilities unexplained. The authors conduct the same analysis for other disciplines using the Survey of Doctoral Recipients data, and found that the gender promotion gap in economics was distinctive. There were negligible gender gaps in the transition to tenure in statistics and the sciences, and only an 8 percent gap in the other social sciences. In engineering, women were more likely than men to have been promoted after ten years.

In a later study of women’s careers in academic social science that examined cohort differences using the 1981–2008 waves of the Survey of Doctoral Recipients, Ginther and Kahn (2014) find that, although there were gender differences in tenure probabilities for the 1980 cohort of PhDs in other social science disciplines, these had disappeared for the 1999 PhDs, while a 20 percent gender gap persisted in economics. They conclude: “Economics is the one field where gender differences in tenure receipt seem to remain even after background and productivity controls are factored in and even for single childless women” (p. 311). Similarly, they find no significant gender differences in promotion to tenure or full professor in the sciences overall after controlling for demographic, family, and productivity covariates (Ginther and Kahn 2009).

In an omnibus study on women in academic science written in collaboration with two psychologists, Ginther and Kahn examine recent career progression in math-intensive fields of study and find evidence of gender inequality only in economics (Ceci, Ginther, Kahn, and Williams 2014). Most of these disciplines made progress towards gender equality in income and promotion between the mid-1990s

and 2010, while economics did not. In geosciences, engineering, math/computer science, and physical sciences, men and women now enter PhD programs at rates proportionate to their representation in college majors and are equally likely as PhD students to be hired into tenure-track positions.

In addition to the persistent gender gap in promotions to tenured positions, Ceci et al. (2014) also find significant gaps in academic salaries and job satisfaction among economists that have not decreased (and in some cases have increased) over time. In the 1995 Survey of Doctoral Recipients data, female assistant professors in economics were paid lower salaries than male assistant professors, but the difference was not significant. By 2010, the average salary gap in these data had increased and become significant. Over the same period, the relative salaries of female full professors fell as well, to 74 percent of male salaries by 2010, though there are no significant salary differences at research-intensive universities with PhD programs (so-called “R1” institutions). Women in the sciences tend to report being less satisfied with their jobs than male scientists, but the gender differences in the 1997 and 2010 Survey of Doctoral Recipients were generally small and fell over time. The gap in job satisfaction among economists, in contrast, was large in 1997 and grew by 2010, with men becoming more likely to report being very or somewhat satisfied with their jobs and women becoming less likely to do so.

Gender gaps in job satisfaction may not be surprising, given the disadvantages women appear to face in promotion and pay, but what might explain these differences in substantive career progression? Many studies have shown that women in science, technology, engineering, and mathematics fields, including economics, have fewer publications than men at equivalent stages of their career, though there appears to be no difference in hours worked (Ceci et al. 2014). Ginther and Kahn (2004) report that, ten years post-PhD, women in the 1989 cohort of assistant professors have 0.3 fewer top-10 publications and 3.8 fewer articles in other journals, though these differences do not explain most of the promotion gap. The 1995 and 2008 Survey of Doctoral Recipients data includes the number of articles accepted in refereed journals in the previous five years and, according to this metric as well, female assistant professors published less than male assistant professors. Between 1995 and 2008, this gap increased and became significant, with women publishing less and men publishing more (Ceci et al. 2014). A recent study based on a broader database of journal articles from EconLit (with gender identified for 80 percent of authors) finds that the raw gender gap in research output for all economists has been relatively constant at around 50 percent since the late 1980s, though 43 percent of this gap can be explained by differences in experience and field (Ductor, Goyal, and Prummer 2018).

A leading hypothesis for why female academics are less productive is that women have more intense domestic responsibilities; indeed, the evidence from most science, engineering, technology, and mathematics fields is that publications by single childless females are not significantly different from publications by single childless men. This is not the case in economics and the physical sciences, however, where there is a significant gender gap among the childless as well. Gender norms

that assign more nurturing roles to women may also influence productivity through the way that time on the job is allocated. Studies of faculty in science, technology, engineering, and mathematics have found a gender discrepancy in time use, with women spending more time on teaching, service, and other nonresearch academic activities such as mentoring students (Xie and Shauman 2003; Misra, Lundquist, Holmes, and Agiomavritis 2011). Female faculty are more likely to volunteer for low-reward tasks (tasks unlikely to contribute to one's chances for promotion), and lab experiments confirm that women volunteer, and are asked to volunteer, more than men (Babcock, Recalde, Vesterlund, and Weingart 2017). However, we are not aware of any economics-specific evidence on professional time allocation.

Evidence for Barriers

If women's relative failure to advance in departments of economics cannot be explained by the gender gap in productivity, the possibility of differential treatment arises. As we will discuss, a number of recent papers explore the role of gender *per se* in the economics profession, examining the possible causes of differential attrition and the persistent gap in tenure probabilities. Taken together, this work builds a case that female economists face substantial barriers throughout their career. These barriers may influence persistence in the profession by reducing expectations of future success, impeding research activity and publication outcomes, or affecting the probability of promotion even conditional on observed productivity.

Barriers that act to limit women from becoming tenured economists may start earlier. For example, Figure 1 shows that only about one-third of undergraduate economics majors are women. Also, Figure 1 shows that attrition rates in economics PhD programs were higher for women than men until the mid-2000s (as shown by the gap between the share of women who were first-year PhD students and the share that were new PhDs). In this symposium, the paper by Buckles discusses the research on policies that have been used in trying to raise the share of women at all stages of the economics career pipeline, while the paper by Boustan and Langan looks at the heterogeneity across departments in the share of women admitted to and completing PhD programs. We focus here primarily on issues affecting the research productivity of female economists.

An accumulating body of evidence suggests that early-career female economists may be adversely affected by limited access to the mentoring and social networks that support research activities, as well as by potential biases in the referee process. For example, a lack of senior female mentors may disadvantage assistant professors, especially if important information about publishing and tenure is transmitted informally within departments or research networks. In an effort to expose female assistant professors to successful female role models, boost research productivity, and help prepare them for the tenure process, the CSWEP Mentoring Program, CeMENT, matches junior female faculty with senior mentors. The program has been routinely oversubscribed, enabling a randomized control trial of the program

to be conducted in the 2000s. This evaluation found that CeMENT significantly increased the publication rates and grant funding of participants, bolstering the argument that a lack of mentoring may be important for women (Blau, Currie, Croson, and Ginther 2010).

Barriers in social network formation that hinder mentoring in a male-dominated field may lead men and women to have different research collaboration and coauthorship networks as well (McDowell, Singell, and Stater 2006). Although women in economics have a higher share of coauthored papers, their coauthorship patterns are distinct from those of men in ways that are predictive of lower output—fewer coauthors, higher clustering, and more collaboration with the same coauthors (Ductor, Goyal, and Prummer 2018). Coauthored publications also appear to be evaluated differently based on the gender of the authors. Male and female economists receive similar credit for sole-authored papers of similar quality in terms of their impact on tenure decisions (Sarsons 2017a). However, women receive significantly less credit for coauthored work, particularly when they coauthor with men. This contrasts with evidence from sociology, where Sarsons finds that men and women benefit equally from coauthored work.

Women and men in economics may also face different experiences throughout the publishing process. Several papers have tested for outright discrimination against women in manuscript review, but the empirical evidence is mixed. Ferber and Teiman (1980) study double-blind reviewing in economics journals and find that the gender gap in acceptance rates is lower when journals use double-blind reviewing. In an experiment of single-blind versus double-blind reviewing, Blank (1991) finds women fare slightly better under a double-blind reviewing system, but the estimated effects are not significant. Abrevaya and Hamermesh (2012) find no evidence of gender discrimination or altruism based on the gender pairing of reviewers and authors in the review process at a top field journal, though the journal uses a double-blind review process. (Of course, reviewers are often able to determine the identity and gender of authors if the paper is posted online.) An important limitation of this gender-pairing research design, however, is that it may fail to identify gender bias in the peer review process if women and men both discriminate against female authors. Card, DellaVigna, Funk, and Iriberry (2018) study referee decisions at four leading economics journals and similarly find no evidence of differential gender bias among reviewers or editors. However, they show that both male and female referees appear to hold female authors to a higher standard (as measured by citation counts), resulting in a substantial difference in the probability that female-authored papers receive a revise and resubmit. Similarly, Grossbard, Yilmazer, and Zhang (2018) show that papers in demographic economics journals with female authors receive more citations. Hengel (2017) adds a different dimension to the evidence that higher editorial standards are imposed on women in economics. She finds that economic research papers written by female authors spend six months longer under review at one top journal, although female-authored papers are more readable (using five different measures of writing clarity) and the gender gap in readability grows over the peer-review process. Hamermesh (2013) finds that, regardless of the reason, female authors have been substantially

underrepresented in top journals since the 1980s. While the evidence is not conclusive, differences in coauthorship networks and potential bias in the publishing process may both contribute to this gap.

External recognition through conference participation may also serve as a barrier to success for women. Women are underrepresented at high-profile conferences in economics compared to the overall share of female assistant professors, which is important if tenure committees use these presentations as a measure of prestige or external recognition of quality work (Chari and Goldsmith-Pinkham 2017).

Finally, the evaluation process for tenure and promotion may systematically disadvantage women. Evidence has been accumulating that implicit bias, which can lead to discrimination on the basis of unconscious attitudes and associations, is a problem in academia and can affect both hiring and promotion decisions on many margins (for a discussion in this journal, see Bayer and Rouse 2016). For example, faculty evaluating curriculum vitae with randomly assigned names are more likely to positively evaluate and hire male applicants for tenure-track jobs (Steinpreis, Anders, and Ritzke 1999). Letters of recommendation written for individuals applying for academic positions use different adjectives to describe men and women, and the characteristics used to describe women are viewed more negatively in hiring decisions (Madera, Hebl, and Martin 2009; Schmader, Whitehead, and Wysocki 2007). More specific to economics, equally productive female economists in Italy are less likely to be promoted to associate or full professor when randomly assigned to an all-male promotion committee, but there is no gender gap when women are assigned to a mixed-gender committee (De Paola and Scoppa 2015).

Even policies that have been supported on the grounds of gender equity may create biases against women's success. Antecol, Bedard, and Stearns (2018) examine the effect of gender-neutral tenure-clock stopping policies, which allow assistant professors who have children to extend their tenure clock. They find that such policies substantially increase the probability that men get tenure in their first job, but reduce the probability that women get tenure. Observed publishing outcomes suggest that men use the additional time on the tenure clock to continue to work and publish while women do not. Moreover, this study also finds that a large and significant gap in the probability of tenure remains even when controlling for the number of publications in top-five and non-top-five journals.

Evidence of gendered expectations of performance exists in many other high-skilled occupations as well. In a study of physician referral practices, Sarsons (2017b) finds that female surgeons are more heavily penalized for negative patient outcomes, while male surgeons are more strongly rewarded after positive outcomes. Another study, of misconduct by financial advisors, finds female financial advisors engage in less-costly types of misconduct on average, but are also significantly more likely relative to men to face harsh punishments following misconduct (Egan, Matvos, and Seru 2017). Finally, men serving on promotion committees across academic disciplines evaluate female candidates less favorably when there are women on the committee as well (Bagues, Sylos-Labini, and Zinovyeva 2017). The contrast between economics and other academic disciplines in the lack of progress

that has been made in reducing gender inequalities, however, suggests that biases within institutions of economics may be particularly pervasive.

Discussion

Following the considerable growth in women's representation among economics students and faculty during the 1970s and 1980s, progress has leveled off in the last two decades. Economics has made less headway than the science, technology, engineering, and mathematics fields in terms of increasing the share of female undergraduate majors and PhD recipients (Bayer and Rouse 2016), which will make it even more difficult to close the faculty gender gap in economics going forward. Furthermore, common explanations for female academic disadvantage, such as heavier domestic responsibilities and an aversion to math intensity, fail to explain why economics is falling behind these other fields in terms of female persistence and promotion probabilities. What can explain the unique challenges that women seem to face in economics?

An adversarial and aggressive culture within academic economics is often advanced as a causal force in women's stalled progress in the profession, though its impact is difficult to quantify. Economics seminars, for example, have a reputation for being particularly hostile environments. The culture of an academic discipline can have gendered implications if women either fail to fully adapt to the culture or if they receive differential treatment as a result of it. Female economists appear to be less likely to engage in practices that are positively correlated with professional success, suggesting an inability or unwillingness to adapt to professional norms. For example, male academics self-cite more than female academics in many fields, but the male-to-female self-cite ratio is twice as high and more persistent in economics (King, Bergstrom, Correll, Jacquet, and West 2017). Applied economics fields attract a higher proportion of women, but this work is still seen by some as less rigorous or less important than traditionally male-dominated topics. Anecdotal evidence suggests that women may choose to go into less-male-dominated fields or leave academia altogether based on early experiences with toxic environments that men are more likely to tolerate.

It is obviously difficult to obtain quantitative estimates of the extent of outright harassment of women in economics. We do know that there are many reports of women in economics experiencing inappropriate behavior in job interviews, seminars, meetings, and at conferences (Shinall 2018). In addition, the language used to describe female economists on at least one anonymous online forum is often sexual and derogatory, in a way that it is not for men (Wu 2017). Recent evidence suggests that gender harassment is a problem in academics more broadly (National Academies of Sciences, Engineering, and Medicine, 2018). Such behavior is often normalized and tolerated in male-dominated settings, making it difficult to change. Thus, the National Academies of Sciences offer several evidence-based recommendations to address harassment in the university setting that may be directly relevant

to economics. In particular, they advise reducing the importance of hierarchical relationships and implementing “power-diffusion” mechanisms such as mentoring networks. They also argue that taking explicit actions to achieve greater gender equity in the hiring and promotion process is an essential step in creating a diverse and respectful environment.

The evidence summarized above suggests two primary mechanisms through which the barriers against women in economics may operate: differences in productivity between men and women, and differences in how they are evaluated. Women may be on average less productive than men due to childbearing and other family responsibilities, a higher propensity to engage in service activities instead of research, or differences in the type of research in which they choose to invest their time. The distinct experiences of men and women in the profession may also contribute to productivity gaps that arise as a result of differences in collaborative networks, access to mentors, and gender harassment. But gender gaps conditional on productivity are also larger in economics than in other academic disciplines, suggesting that a second factor explaining female disadvantage in economics may be disparate assessment of men and women. It appears that women are held to higher standards than men of equal ability, and need to publish more, higher-quality work to achieve equal levels of success in this profession.

Continued progress toward equality in academic economics will require a widespread awareness that these barriers exist, accompanied by a concerted effort to remove opportunities for bias in the hiring and promotion process. However, first steps have been slow in coming. A 2008 survey of AEA members found, in addition to substantial differences in the policy views of male and female economists, a meaningful gender gap in their beliefs on equal opportunity in the profession (May, McGarvey, and Whaples 2014). While 76 percent of female AEA members believed that opportunities for economics faculty in the US favor men, fewer than 20 percent of men shared the same view. In fact, one-third of male economists felt that opportunities in economics actually favor women. To the extent that such beliefs persist, they are a major obstacle to the development of new diversity initiatives.

Diversifying the economics profession is important, because a greater breadth of individual perspectives will affect what is taught in the classroom, what research questions are asked, and how policy discussions are addressed. In addition, to the extent that women’s stalled progress in economics is the result of discrimination or biased assessment, as recent evidence suggests, continued action to remove these barriers can be justified both on the basis of simple fairness and also on the benefits of creating an environment where equal work yields equal rewards.

■ *We are grateful to Juliana Helo and Sangeetha Ramamurthy for excellent research assistance, and to Dick Startz and Meredith Startz for helpful comments.*

References

- Abrevaya, Jason, and Daniel S. Hamermesh. 2012. "Charity and Favoritism in the Field: Are Female Economists Nicer (to Each Other)?" *Review of Economics and Statistics* 94(1): 202–207.
- Antecol, Heather, Kelly Bedard, and Jenna Stearns. 2018. "Equal but Inequitable: Who Benefits from Gender-Neutral Tenure Clock Stopping Policies?" *American Economic Review* 108(9): 2420–41.
- Babcock, Linda, Maria P. Recalde, Lise Vesterlund, and Laurie Weingart. 2017. "Gender Differences in Accepting and Receiving Requests for Tasks with Low Promotability." *American Economic Review* 107(3): 714–47.
- Bagues, Manuel, Mauro Sylos-Labini, and Natalia Zinovyeva. 2017. "Does the Gender Composition of Scientific Committees Matter?" *American Economic Review* 107(4): 1207–38.
- Bartlett, Robin L. 1998. "CSWEP: 25 Years at a Time." *Journal of Economic Perspectives* 12(4): 177–183.
- 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.
- Bell, Carolyn Shaw. 1973. "Report of the Committee on the Status of Women in the Economics Profession." *American Economic Review* 63(2): 508–11.
- Beneito, Pilar, J. E. Boscá, Javier Ferri, and Manu García. 2018. "Women across Subfields in Economics: Relative Performance and Beliefs." FEDEA Working Paper No 2018-06.
- Bergmann, Barbara R. 1998. "Two Cheers for CSWEP?" *Journal of Economic Perspectives* 12(4): 185–189.
- Blank, Rebecca M. 1991. "The Effects of Double-Blind versus Single-Blind Reviewing: Experimental Evidence from *The American Economic Review*." *American Economic Review* 81(5): 1041–67.
- Blau, Francine D., Janet M. Currie, Rachel T. A. Croson, and Donna K. Ginther. 2010. "Can Mentoring Help Female Assistant Professors? Interim Results from a Randomized Trial." *American Economic Review* 100(2): 348–52.
- Card, David, Stefano DellaVigna, Patricia Funk, and Nagore Iriberry. 2018. "Are Referees and Editors in Economics Gender Neutral?" October https://eml.berkeley.edu/~sdellavi/wp/EditorGender_Oct24.pdf.
- Ceci, Stephen J., Donna K. Ginther, Shulamit Kahn, and Wendy M. Williams. 2014. "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 23953.
- Cherrier, Beatrice. 2017a. "Classifying Economics: A History of the JEL Codes." *Journal of Economic Literature* 55(2): 545–79.
- Cherrier, Beatrice. 2017b. "'The American Economic Association Declares that Economics is not a Man's Field': The Missing Story." *The Undercover Historian* (blog), March 31, 2017.
- Cherrier, Beatrice, Cleo Chassonnery-Zaigouche, and John Singleton. 2018. "A Game of Mirrors? Economists' Models of the Labor Market and the 1970s Gender Reckoning." *The Undercover Historian* (blog), March 6, 2018.
- CSWEP. 1973. "Combating Role Prejudice and Sex Discrimination: Findings of the American Economic Association Committee on the Status of Women in the Economics Profession." *American Economic Review* 63(5): 1049–61.
- CSWEP. 2018. "2017 Report on the Status of Women in the Economics Profession." *CSWEP News*, Issue I. <https://www.aeaweb.org/content/file?id=6388>.
- De Paola, Maria, and Vincenzo Scoppa. 2015. "Gender Discrimination and Evaluators' Gender: Evidence from Italian Academia." *Economica* 82(325): 162–88.
- Ductor, Lorenzo, Sanjeev Goyal, and Anja Prummer. 2018. "Gender & Collaboration." Cambridge-INET Working Paper 1807.
- Egan, Mark L., Gregor Matvos, and Amit Seru. 2017. "When Harry Fired Sally: The Double Standard in Punishing Misconduct." NBER Working Paper 23242.
- Ferber, Marianne A., and Michelle Teiman. 1980. "Are Women Economists at a Disadvantage in Publishing Journal Articles?" *Eastern Economic Journal* 6(3/4): 189–193.
- Forget, Evelyn L. 2011. "American Women and the Economics Profession in the Twentieth Century." *Æconomia* 1(1): 19–30.
- Fourcade, Marion, Etienne Ollion, and Yann Algan. 2015. "The Superiority of Economists." *Journal of Economic Perspectives* 29(1): 89–114.
- Friedman, Milton. 1998. "A Comment on CSWEP." *Journal of Economic Perspectives* 12(4): 197–99.
- 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.

Ginther, Donna K., and Shulamit Kahn. 2009. "Does Science Promote Women? Evidence from Academia 1973–2001." Chap. 5. in *Science and Engineering Careers in the United States*, edited by Richard B. Freeman and Daniel F. Goroff. Chicago, IL: University of Chicago Press for NBER Science Engineering Workforce Project.

Ginther, Donna K., and Shulamit Kahn. 2014. "Academic Women's Careers in the Social Sciences." Chap. 11 in *The Economics of Economists: Institutional Setting, Individual Incentives, and Future Prospects*, edited by Alessandro Lanteri and Jack Vromen. Cambridge University Press.

Grossbard, Shoshana, Tansel Yilmazer, and Lingrui Zhang. 2018. "The Gender Gap in Citations: Lessons from Demographic Economics Journals." HCEO Working Paper 2018-078.

Hamermesh, Daniel S. 2013. "Six Decades of Top Economics Publishing: Who and How?" *Journal of Economic Literature* 51(1): 162–72.

Hengel, Erin. 2017. "Publishing While Female. Are Women Held to Higher Standards? Evidence from Peer Review." Cambridge Working Papers in Economics 1753, Faculty of Economics, University of Cambridge.

King, Molly M., Carl T. Bergstrom, Shelley J. Correll, Jennifer Jacquet, and Jevin D. West. 2017. "Men Set Their Own Cites High: Gender and Self-Citation across Fields and Over Time." *Socius* 3.

Madera, Juan M., Michelle R. Hebl, and Randi C. Martin. 2009. "Gender and Letters of Recommendation for Academia: Agentic and Communal Differences." *Journal of Applied Psychology* 94(6): 1591–99.

May, Ann Mari, Mary G. McGarvey, and Robert Whaples. 2014. "Are Disagreements among Male and Female Economists Marginal at Best?: A Survey of AEA Members and Their Views on Economics and Economic Policy." *Contemporary Economic Policy* 32(1): 111–132.

McDowell, John M., Larry D. Singell Jr, and Mark Stater. 2006. "Two to Tango? Gender Differences in the Decisions to Publish and Coauthor." *Economic Inquiry* 44(1): 153–68.

McDowell, John M., Larry D. Singell Jr, and James P. Ziliak. 2001. "Gender and Promotion in the Economics Profession." *ILR Review* 54(2): 224–44.

Misra, Joya, Jennifer Hickes Lundquist, Elissa

Holmes, and Stephanie Agiomavritis. 2011. "The Ivory Ceiling of Service Work." *Academe* 97(1): 22–26.

National Academies of Sciences, Engineering, and Medicine. 2018. *Sexual Harassment of Women: Climate, Culture, and Consequences in Academic Sciences, Engineering, and Medicine*. Consensus Study Report. Washington, DC: The National Academies Press.

Nelson, Donna J. 2004. "Nelson Diversity Surveys." Norman, OK: Diversity in Science Association. <http://drdonnajnelson.oucreate.com/diversity/top50.html> (accessed June 3, 2018).

Sarsons, Heather. 2017a. "Gender Differences in Recognition for Group Work." https://scholar.harvard.edu/files/sarsons/files/full_v6.pdf.

Sarsons, Heather. 2017b. "Interpreting Signals in the Labor Market: Evidence from Medical Referrals." https://scholar.harvard.edu/files/sarsons/files/sarsons_jmp.pdf.

Schmader, Toni, Jessica Whitehead, and Vicki H. Wysocki. 2007. "A Linguistic Comparison of Letters of Recommendation for Male and Female Chemistry and Biochemistry Job Applicants." *Sex Roles* 57(7–8): 509–514.

Shinall, Jennifer Bennett. 2018. "Dealing with Sexual Harassment." *CSWEP News*, Issue I. <https://www.acaweb.org/content/file?id=6769>.

Steinpreis, Rhea E., Katie A. Anders, and Dawn Ritzke. 1999. "The Impact of Gender on the Review of the Curricula Vitae of Job Applicants and Tenure Candidates: A National Empirical Study." *Sex Roles* 41(7–8): 509–528.

Stevenson, Betsey, and Hanna Zlotnick. 2018. "Representations of Men and Women in Introductory Economics Textbooks." *AEA Papers and Proceedings* 108(5): 180–85.

Strober, Myra. 2016. *Sharing the Work: What My Family and Career Taught Me about Breaking Through (and Holding the Door Open for Others)*. MIT Press.

Wu, Alice H. 2017. "Gender Stereotyping in Academia: Evidence from Economics Job Market Rumors Forum." Working Papers 2017-09, Princeton University, Woodrow Wilson School of Public and International Affairs, Center for Health and Wellbeing. <https://drive.google.com/file/d/0BwjFN4HbBrDBbnFqZzdLWThDb0U/view>.

Xie, Yu, and Kimberlee A. Shauman. 2003. *Women in Science: Career Processes and Outcomes*. Cambridge: Harvard University Press.

NBER WORKING PAPER SERIES

PUBLICATION, COMPENSATION, AND THE PUBLIC AFFAIRS DISCOUNT:
DOES GENDER PLAY A ROLE?

Lori L. Taylor
Kalena E. Cortes
Travis C. Hearn

Working Paper 26022
<http://www.nber.org/papers/w26022>

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
June 2019

Special thanks for helpful comments and suggestions from Jeff Denning, Joanna Lahey, and Abigail Payne as well as the conference and seminar participants at the Association for Education Finance and Policy and Texas A&M's Bush Schools Quantitative Brown Bag Series. The authors thank Adria Escobedo and Tate Cannon who provided outstanding research assistance. Any errors are attributable to the authors. Institutional support from Texas A&M's Bush School of Government and Public Service is gratefully acknowledged. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2019 by Lori L. Taylor, Kalena E. Cortes, and Travis C. Hearn. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Publication, Compensation, and the Public Affairs Discount: Does Gender Play a Role?

Lori L. Taylor, Kalena E. Cortes, and Travis C. Hearn

NBER Working Paper No. 26022

June 2019

JEL No. J01,J16,J30,J31

ABSTRACT

This paper presents on three new styled facts: first, schools of public affairs hire many economists; second, those economists are disproportionately female; and third, salaries in schools of public affairs are, on average, lower than salaries in mainline departments of economics. We seek to understand the linkage, if any, among these facts. We assembled a unique database of over 2,150 faculty salary profiles from the top 50 Schools of Public Affairs in the United States as well as the corresponding Economics and Political Science departments. For each faculty member we obtained salary data to analyze the relationship between scholarly discipline, department placement, gender, and annual salary compensation. We found substantial pay differences based on departmental affiliation, significant differences in citation records between male and female faculty in schools of public affairs, and no evidence that the public affairs discount could be explained by compositional differences with respect to gender, experience or scholarly citations.

Lori L. Taylor
Texas A&M University
College Station, TX 77843
lltaylor@tamu.edu

Travis C. Hearn
Texas A&M University
College Station, TX 77843
thearn171@gmail.com

Kalena E. Cortes
The Bush School
Texas A&M University
4220 TAMU
1049 Allen Building
College Station, TX 77843
and IZA
and also NBER
kcortes@tamu.edu

I. INTRODUCTION

Salary equity studies have long been used to illuminate pay disparities among higher education faculty based on gender (e.g., Ginther and Hayes, 2003); race and ethnicity (e.g., Webber and Gonzales Canche, 2015); organizational loyalty (e.g., Barbezat, 2004; Masakure, 2016), and scholarly discipline (e.g., Ehrenberg, McGraw, and Mrdjenovic, 2006; Curtis and Kisielewski, 2015). Our study contributes to the literature by disentangling the contributions of discipline, departmental affiliation and gender in faculty pay differentials for schools of public affairs.

Three key observations motivate our analysis. First, we observe that schools of public affairs hire many economists. Second, we observe that those economists are disproportionately female. Finally, we observe that salaries in schools of public affairs are, on average, lower than salaries in mainline departments of economics. We seek to understand the linkage, if any, among these three new styled facts.

Using various data sources, we assembled a unique database of over 2,150 faculty salary profiles from the top 50 Graduate Schools of Public Affairs in the United States as well as the corresponding Economics and Political Science departments in their home universities. For each faculty member we collected compensation data to analyze the relationship between field of specialization, department placement, and annual salary compensation. In addition, we also control for measures of scholarly productivity, such as, citation counts, h-indices, and i10-indices.

There are a couple of key takeaways from our analysis here. First, we found substantial pay differences based on departmental affiliation. Economists in schools of public affairs were paid at a *significant discount* compared to their counterparts employed in traditional departments of economics. Second, the public affairs discount could *not* be explained by compositional differences. Controlling for gender, experience and various indicators of research productivity only

widened the salary gap between economics departments and schools of public affairs. Lastly, we found evidence that citation records differed by gender, discipline and departmental affiliation; and that salary differences between male and female faculty members were not statistically significant once these differences in citation records were taken into account. We found no evidence that the return to citations was systematically higher (or lower) for male than for female faculty members in the institutions of higher education under analysis.

II. RELEVANT LITERATURE

The economics literature includes a remarkable number of salary analyses in the higher education contexts. Nearly all of these studies are at least loosely based on hedonic wage models in which the individual's salary is a function of worker characteristics, job characteristics and location characteristics.¹ Most of these studies focus on quantifying faculty salary differences by some demographic of interest, typically by gender, race/ethnicity or scholarly discipline. Many focus exclusively on salary differentials within economics.

One of the key challenges in those analyses of faculty salary has been the measurement of productivity. Common measures of research productivity include: numbers of publications (Ginther, 2003; Barbezat, 2004; Umbach, 2007; Binder et al., 2010; Renzulli, Reynolds, Kelly, and Grant, 2013; Tuckman and Tuckman, 1976); numbers of citations (Hamermesh, Johnson and Weisbrod, 1982; Gibson, Anderson, and Tressler, 2017; Gibson and Burton-McKenzie, 2017; Hamermesh, 2018) or grant dollars received (Binder et al., 2010; Renzulli et al., 2013). Common measures of teaching productivity include: numbers of teaching awards (Carlin et al., 2013);

¹ For a recent review of the higher education literature, see Gibson, Anderson and Tressler (2017) and Hamermesh (2018). For a more general review of the salary equity literature, see Blau and Kahn (2017).

numbers of instructional preparations (Binder et al., 2010); numbers of classes taught (Renzulli et al., 2013); or percent time in teaching (Umbach, 2007).

Of keen interest (at least to economists) have been studies documenting and explaining salary premia for economists vis-à-vis faculty in other academic disciplines. Academic economists enjoy some of the highest average salaries among university faculty. Ehrenberg, McGraw, and Mrdjenovic (2006) found that, at the full professor level, only faculty in engineering and computer science programs earn more than those in economics departments. More recently, Curtis and Kisielewski (2015) found that professors' salaries in economics exceeded those of their social scientist peers in sociology, anthropology, political science and government at each of the assistant, associate, and full professor ranks. For the period of 2002 through 2016, salaries of economists also grew at a faster rate than any their other social science counterparts, rising twice as quickly as anthropology and political science and at three times the rate of sociology for new assistant professors in particular (Curtis and Kisielewski, 2015).

There are several reasons why academic economists enjoy such a salary premium over their colleagues in other social sciences. Human capital theory predicts that workers with similar characteristics—such as professors with similar degrees, responsibilities, and years of experience—will receive similar compensation while workers with dissimilar characteristics will receive dissimilar compensation (e.g., Park, 2012). Even though their teaching tasks might be similar to those of other academics, economists could earn a premium based on their comparative advantage in producing specific, highly-valued academic products. Freeman (1999) argued that economics, with more powerful analytic tools, stronger students (i.e., based on Graduate Record Examinations [GRE] scores), and more mathematically demanding curricula than in sociology or political science, was “the strongest social science” and therefore commanded the highest wages.

Alternatively, external labor market pressure could account for differences in salary. Economists enjoy unique access to high-value alternative sources of income, such as consulting fees, private investment, and corporate board membership (Fourcade, Ollion, and Algan, 2015), which suggest a high earnings potential for economists who choose to pursue opportunities outside academia. Relatively high wages for economics faculty are therefore justified because universities must compete with the demand for economists in the non-academic labor market (Boyle, 2008; Ehrenberg, 2004b; Buchanan and Tollison, 1981).

On the other hand, some researchers attribute the higher salaries in economics, at least in part, to the male-dominated nature of the discipline (Bellás, 1994; Bellás, 1997; Smart, 1991; Perna, 2003). Researchers have found evidence that female salaries in higher education are systematically below those of males (e.g., Ginther and Hayes, 2003; Ginther and Kahn, 2004; Umbach, 2007) even after controlling for differences in experience and productivity. To the extent that this bias persists, salaries could be higher in economics simply because there are fewer women in the profession.

The disproportionately male composition of economics departments may also exert an influence on pay through citation practices. Dion, Sumner, and Mitchell (2018) find that male scholars not only cite other male scholars more often than they cite female scholars, but are also more likely to cite themselves than are female scholars. In economics specifically, male-authored papers are cited at higher rates than either female-authored papers or mixed-authorship papers (Ferber and Brün, 2011). Considering that the marginal value of a citation may be as high as \$1,300 per year (Diamond Jr., 1986), gendered citation practices can result in disproportionate financial and career rewards for scholars in male-dominated subfields of social science (Dion, Sumner, and Mitchell, 2018).

Thus, the literature suggests that gender composition could at least partially explain the salary gap between economics departments and schools of public affairs. If more female academic economists choose to pursue careers in schools of public affairs, we might expect lower pay for economists in those fields due to a greater proportion of women on the faculty.

Much of the research examining faculty salary dispersion between disciplines relies heavily on survey data and very few to date rely on administrative university-level faculty data.² Thus, our study contributes to this literature by analyzing a uniquely assembled dataset of over 2,150 faculty salary profiles from the top 50 Graduate Programs in Public Affairs in the United States and their corresponding Economics and Political Science departments at those universities, rather than survey data.³ Unlike surveys conducted by groups such as the American Association of University Professors (Ehrenberg, 2004a) or American Political Science Association (Claypool et al., 2017), the multi-institutional faculty-level data in our study allows us to specifically distinguish faculty by their discipline. For each faculty member we concatenated compensation data to analyze the relationship between specialization, department placement, and salary. When possible we also collected information of the faculty's level of productivity as measured by their citation counts, h-indices, and i10-indices.

² Only a small number of studies about salary dispersion within or between departments use university-level faculty salary data, and these typically use data from only a single university or a small sample (e.g., Boudreau et al., 1997; Volk, Slaughter, and Thomas, 2001; Hilmer, Hilmer, and Lush, 2012).

³ Several studies (Fairweather, 1995; Umbach 2007, 2008) utilize the National Survey of Postsecondary Faculty (NSOPF), administered by the National Center for Education Statistics. Friedman (1991) conducted a telephone survey of department heads at public administration programs, one element of which aimed to determine how many economists were employed by public affairs programs. The Carnegie Survey of Faculty also appears in the Langton and Pfeffer (1994) study about labor market characteristics, while Gomez-Mejia and Balkin (1992) opted to design their own survey.

III. DATA SOURCES AND DESCRIPTIVE STATISTICS

We approach our inquiry into the relative salaries of economists from the perspective of a leading school of public affairs. Thus, we began our analysis by identifying a sampling frame based on the universities housing the top graduate programs in public affairs, according to the U.S. News and World Report (USNWR) rankings for 2018. We chose to limit our search to schools of public affairs in the top 50 of the USNWR rankings to ensure a sufficiently large sample while maintaining a focus on sectoral leaders. Once we identified the leading schools of public affairs, we sought publicly available faculty salary data for each of the corresponding colleges or universities. No private institutions of higher education (IHEs) meeting the search criteria voluntarily published salary data publicly. However, we extracted payroll information for the 33 public IHEs from numerous sources, including university and state government websites, nonprofit organizations committed to transparency, and newspapers or other periodicals in the state of interest. If salary data were unpublished, we submitted open records requests to obtain the sought-after data). See Table 1 for descriptive statistics

[Table 1 about here]

Not all of these original datasets provided the same information. Each publicly-available payroll dataset included name, title, and salary. However, some datasets provided multiple figures for salary, while others provided only one. Some datasets also provided full-time equivalent (FTE) numbers, but not others. The reporting years varied. Where available, we calculated the most recent salary using FTE information for each faculty. However, we recognize that there are institution-specific differences in the measurement of salary. As a result, our analyses will include IHE fixed effects in all estimated regression models.

For each of the 33 IHEs, we retained salary data for three departmental types: public affairs, economics, and political science. We restricted the sample to regular faculty members at the assistant, associate and full professor levels. Thus, we excluded adjunct, visiting, distinguished, and emeritus positions, as well as administrative faculty such as deans or department heads.

The next step in compiling a useful dataset required visiting each university's website and matching information from the faculty directory with the names listed in the dataset for that institution. We coded each faculty member's department using a series of indicator variables based on the faculty department website where their profile appeared. To determine a faculty member's sex, highest degree held, year of degree and academic discipline, we examined information available in their faculty biography or the curriculum vitae (CV) posted on their faculty page. If such information were unavailable, we performed a broader Internet search to locate any missing details about an individual, leading either to a personal website or to their PhD information posted with an affiliated organization.

We constructed indicator variables for whether or not a faculty member was an economist or a political scientist based on the field in which they obtained their PhD and their departmental affiliations. Faculty members with a degree in economics (including variants like agricultural economics or econometrics) were classified as economists, as were all faculty members in departments of economics. Faculty members with degrees in political science, government, foreign affairs or international relations were classified as political scientists, as were all faculty members in departments of political science. Individuals with degrees in political economy or "political science and economics" were assigned a one for both the economist indicator variable and the political scientist indicator variable.⁴ If professors listed only their PhD-granting institution,

⁴ The analysis was not sensitive to this coding choice because the number of individuals was small.

without any additional PhD information and were not affiliated with a traditional economics or political science department, we coded their field of degree as missing. In addition, if the available online resources indicated that the faculty member did not hold a PhD, then he or she was excluded from the analysis.⁵ We coded as missing any educational attainment data that could not be located online, which occurred most often when a professor did not provide a CV publicly. Some faculty advertised their CV as available on request, but we did not send any CV requests for this study, opting to use only publicly posted information.

Because we were interested in analyzing differences in salary by gender, we also collected the information on faculty's sex from the same public sources as discipline and department information. Then, we coded each faculty member as male or female based on our familiarity with the individuals, photographs posted to their faculty profile or personal website, and any gendered third-person pronouns included in their own biographical materials.

Finally, we followed Hamermesh (2018) and measured faculty productivity using citation metrics commonly available through Google Scholar. We captured total citations, h-indices (which Hamermesh (2018) characterizes as combining both the breadth and depth of impact) and i10-indices (which measure the number of papers with at least 10 citations) for each individual with a Google Scholar profile.⁶ Unfortunately, there are a number of faculty members who do not have a Google profile or have not made their profiles public. For these faculty members, we used multiple imputation to fill in the missing citation data for some of our analyses and excluded the records

⁵ Nearly all of the individuals excluded on these grounds were employed by schools of public affairs and held juris doctorates.

⁶ The h-index is calculated by counting the number of publications for which a scholar has been cited by other scholars at least that same number of times. For example, a h-index of 16 means that the scholar has published at least 16 papers that have each been cited at least 16 times. The i10-index is the number of publications with at least 10 citations. For example, an i10-index of 12 means that 12 publications of the scholar have each been cited at least 10 times.

with missing data for other estimates of the same specifications. None of our results were substantively changed by the modeling choices regarding missing data.

Table 2 provides summary statistics on the distribution of faculty by departmental affiliation, sex, and annual salary compensation. As these sample summary statistics make clear, we were able to confirm each of our motivating observations. First, leading schools of public affairs employ a large number of economists. The 33 leading public affairs departments in our sample employed more than 100 economists, or 12 percent of the economists in the sample.

Second, a disproportionate number of the economists employed by schools of public affairs were female. As Table 2 illustrates, 19 percent of the faculty in departments of economics were female; whereas 23 percent of the economists in departments of political science were female and 35 percent of the economists in schools of public affairs were female. In other words, the fraction female among economists in schools of public affairs was nearly double the fraction female in traditional departments of economics. This difference among economists between departments of economics and public affairs was statistically significant by any common metric.

Third, average salaries in schools of public affairs were lower than those in traditional economics departments. On average—and without adjustment for faculty rank or institution reporting differences—salaries were 33.5 percent higher in departments of economics than they were in schools of public affairs. Among economists, average salaries were 11 percent higher in departments of economics than they were in schools of public affairs.

[Table 2 about here]

IV. EMPIRICAL ANALYSIS AND DISCUSSION OF RESULTS

Of course, average salaries are not the best comparisons. Therefore, consider the hedonic salary models reported in Table 3. In general, we estimated the following functional form,

$$\ln(\text{Salary}_{id}) = \alpha + \text{Faculty}_{id} \cdot \beta + X_{id} \cdot \delta + \tau_j + \mu_{id}$$

where $\ln(\text{Salary}_{id})$ is the log of annual salary for faculty member i in department d , Faculty_{id} is a vector of faculty characteristics, such as, departmental affiliation (school of public affairs (omitted category), economics department, and political science department), field of specialization (economist, political scientist, and other fields (omitted category)), female indicator, professor rank indicator (assistant professor (omitted category), associate professor, and full professor), and experience level of the faculty as measured by the log of years since their Ph.D. degree; X_{id} is a vector of variables capturing various measures of faculty productivity, such as, total citation count, h-index, and i10-index; τ_j are university fixed effects; and lastly, μ_{id} is the error term.

The first column in Table 3 shows a parsimonious model specification with university fixed effects (which allowed us to control for systematic, unobservable differences by institution such as differences in fringe benefits, salary reporting protocols or reporting years) and indicators for departmental affiliation and discipline, but no other controls. This model indicated that the average salary for economists in economics departments was 17 percent (the exponent of the coefficient of $0.154 = 1.17$) higher than the average salary for economists at the same institution who were not in economics departments, and 32 percent (the exponent of $0.154 + 0.124 = 1.32$) higher than the average salary for faculty members in schools of public affairs who were neither economists nor political scientists (the omitted category). Thus, our first cut at the data suggested that economists in schools of public affairs were paid at a significant discount vis-à-vis those in traditional departments of economics.

[Table 3 about here]

This pattern led us to hypothesize that salaries were lower in schools of public affairs because the faculty members who chose to work in such contexts were systematically different from those who chose to work in traditional departments of economics. Those differences could have existed upon arrival (such as gender) or could have been influenced by context (such as scholarly productivity).

Alternative Explanation #1: Faculty Sorting by Gender

It is clear from Table 2 that there is departmental sorting by gender. Moreover, it is well-documented that the economics profession has had, and continues to have, problems with gender equity.⁷ So, the apparent public affairs discount could simply reflect the greater share of female faculty in schools of public affairs. To explore this hypothesis, the second column in Table 3 presents a model that included an indicator for female faculty member. As the table makes clear, we found a significant, negative differential for female faculty members, but controlling for gender did not eliminate the public affairs discount.

Next, we considered a model that included not only an indicator for female faculty, but also the interactions between gender and discipline (to allow for the possibility that there are gender-linked salary differentials among economists, but not among political scientists or among other public affairs disciplines) and the interactions between gender and department affiliation (to allow for the possibility that bias is more strongly expressed in traditional economics departments than in schools of public affairs). As column (3) of Table 3 illustrates, this model suggested that female faculty members were paid significantly less than male faculty members regardless of discipline or department. All of the gender interaction terms were jointly insignificant ($F(4, 2110) = 1.34, p = 0.25$). Again, controlling for gender did not eliminate the public affairs discount.

⁷ See, for example the discussion in Lundberg and Stearns (2019) and Ceci et al. (2014). See also Wu (2018).

Alternative Explanation #2: Faculty Sorting by Experience and Rank

Maybe there were systematic differences between male and female faculty members or between departments of economics and schools of public affairs with respect to faculty rank and experience. The fourth and fifth columns in Table 3 present models that included faculty rank and years since degree (in logs), and the interactions between these two measures of productivity and the economist and political scientist indicators (to allow the effects of rank and experience to vary by discipline).⁸ Because some of the records do not have year of degree, we report both a model in which we dropped observations with missing data on experience (shown in column (4)) and a model in which we multiply imputed years since degree using 100 imputations (shown in column (5)). For observations for which years since degree was imputed, each interaction term was treated as just another variable, meaning that the data were transformed and then the interactions were imputed, as recommended in Von Hippel (2009) and Seaman, Bartlett, and White (2012).

As Table 3 illustrates, the results shown in columns (4) and (5) were not sensitive to the imputation. In both models, we found that controlling for years of experience and faculty rank greatly reduced the magnitude of the difference between male and female salaries, but the gender difference remained large and statistically significant. Again, all of the sex interaction terms were jointly insignificant ($F(4, 1930) = 0.51$, $p = 0.73$ in column (4); $F(4, 2099) = 0.74$, $p = 0.57$ in column (5)) suggesting that the systematic difference between male and female salaries was not a function of discipline or departmental affiliation. Furthermore, this modeling change was associated with a modest increase in the economics department premium, so there was no evidence that the public affairs discount arose from sorting by sex or seniority.

⁸ Although some argue that faculty rank is endogenous, excluding rank as a control can produce misleading conclusions, particularly in the case of gender equity, showing bias where none exists and vice versa (Boudreau et al., 1997).

Alternative Explanation #3: Faculty Sorting by Other Indicators of Research Productivity

Experience and professorial rank are highly imperfect measures of scholarly productivity. Suspecting that there were other dimensions of research productivity that could help explain the gender differential and the public affairs discount, we turned our attention to the faculty citation metrics available from Google Scholar—citation counts, h-indices and i10-indices.

We note that many researchers do not make their Google Scholar profiles public. Therefore, we were unable to collect citation metrics for more than one third of the researchers in our analysis sample. Again, we relied on multiple imputation to fill in the blanks.

Furthermore, we could not conclude that the citation metrics were missing completely at random. As shown in panel A of Table 4, multinomial probit models indicate that the probability that faculty members lacked a public Google Scholar profiles was an increasing function of the number of years since their doctoral degrees. We found this pattern whether or not the model included imputed values for years since degree. In addition, full professors appeared somewhat more likely to have a public Google Scholar profile than assistant or associate professors, once we controlled for differences in graduation date. Therefore, it was necessary to control for faculty rank and years since degree in the imputation model for each of the Google Scholar metrics.

The remaining panels of Table 4 provide reasons to believe that variations in the citation metrics could at least partially explain the observed differences in salary by sex and departmental affiliation. Although the variables were seldom individually significant, the hypothesis that the female indicator and all its interactions were jointly zero was rejected at the 5-percent level for each of the citation metrics. The hypothesis that the economics department indicator and the interaction between female and the economics department indicator were jointly zero was rejected at the five percent level for both the h-index and i10-index models, but not the citation count model,

suggesting that departmental affiliation matters for at least some of the citation metrics. Intriguingly, where departmental affiliation is statistically significant (i.e., in the h-index and i10-index models), the evidence suggests that scholarly productivity among male economists is lower in traditional departments of economics than in schools of public affairs, all other things being equal, while scholarly productivity among female economists is unrelated to departmental affiliation. On the other hand, the h-indices and i10-indices were systematically lower for female economists than for male economists in both traditional economics departments and schools of public affairs, although the differential was smaller and only significant at the 10-percent level in traditional departments of economics.

[Table 4 about here]

Next, Table 5 presents variations on column (5) of Table 3, which we added the citation metrics and appropriate interaction terms. The interactions between gender and departmental affiliation, which were uniformly negligible in preliminary analyses, have been dropped to avoid over-parameterization.

In the first column, we present a model that omitted observations without citation data. In all other specifications, the citation-based indicators of productivity were multiply imputed to address the problem of missing data. Because whether or not the Google Scholar data were missing depended on the years since degree, it was necessary to use years since degree to impute the citation metrics. As a result, all of these models exclude observations for which years since degree was missing.

As the table illustrates, we found a significant, positive and multi-collinear relationship between the citation metrics and faculty salaries. As shown in column (4), each additional one-unit increase in the h-index was associated with a one percent increase in faculty salary, all other

things being equal. Increases in citation count had no incremental effect on salary, once the h-index and i10-index were taken into account, and increases in the i10-index were slightly negative, holding the other citation metric constant (shown column (2)). However, when each of the metrics was evaluated separately, there was consistent evidence that increases in citation led to increases in salary (columns (3)-(5)).

[Table 5 about here]

Incorporating citation metrics into the analysis altered the estimated relationship between gender and salary. The marginal effect of gender shrank, and in the imputed specifications became statistically insignificant. In none of the specifications was the marginal effect of being a female economist statistically significant at the 5-percent level (although it was significant at the 5-percent level in the model where citation counts were not imputed, and therefore fewer highly experienced scholars were included).

Interestingly, the public affairs discount widened. As column (2) shows, controlling for citation metrics as well as years since degree and faculty rank, we estimated that economists in schools of public affairs earned at least 28 percent more than otherwise similar faculty members, and economists in departments of economics earned 17 percent more than economists in schools of public affairs. On the other hand, political scientists were better paid in a school of public affairs than in a traditional department of political science, even after controlling for research productivity. As Table 5 illustrates, political scientists in mainline political science departments were paid 10 percent less, on average, than political scientists in schools of public affairs.

We wondered if perhaps the public affairs discount arose from systematic differences in the relative prestige of the various departments. After all, our sample was restricted to public universities that are home to leading schools of public affairs—which does not necessarily imply

that they are also home to leading departments of economics or political science. One could argue that the premium for political scientists in leading schools of public affairs could arise from differences in departmental prestige (as in Ehrenberg, McGraw, and Mrdjenovic, 2006) although it would seem difficult to make such a case for the public affairs discount vis-à-vis a less selective set of departments of economics.

Nevertheless, we explored this possibility as well, by adding two interaction terms to the model. The first is the interaction between the economics department indicator and the US News and World Report rankings for the university's department of economics; the second is the interaction between the political science department indicator and the US News and World Report ranking for the university's department of political science. (We do not add the main effects of the departmental rankings because those are subsumed in the university fixed effects.) As the last column in Table 5 makes clear, controlling for the departmental rankings added nothing to the analysis.

We wondered if the results regarding the vanishing gender differential and persistent public affairs discount could be driven by the salary or reporting practices at an outlier university. Therefore, we performed a jackknife-inspired sensitivity analysis of our preferred specification (column (2)), by excluding one university at a time from the data file, re-imputing, and re-estimating. In all cases, there was no systematic gender differential once citations were taken into account, and the sign, significance and magnitude of the departmental premiums were largely unaffected. (Tables available upon request.)

[Table 6 about here]

Lastly, we also wondered if the return to citations might be higher among economists than among the other social scientists typically found in schools of public affairs. We found mixed

evidence on that front. As Table 6 illustrates, when we allowed for an interaction between gender or discipline and the citation metrics, we found no evidence that the marginal effect was higher for economists than for political scientists, but some evidence (from the non-imputed models) that the marginal effect of an increase in citations was larger for these two social science disciplines than for other disciplines found in schools of public affairs. Contrary to the findings of Sarsons (2017) we found no reason to believe that the return to citations was systematically different for males than for female faculty members in the institutions of higher education under analysis.

V. CONCLUDING REMARKS

Our salary study provides insight into fiscal practices at highly-ranked public universities around the United States. In agreement with the literature reviewed, we found evidence of lower average wages for women, as well as all faculty in departments that include a higher proportion of women. The female differential disappeared when we controlled for citation record, but the disciplinary and departmental differentials did not. We found some evidence that citation records differed by gender, which is consistent with previous work by Dion, Sumner, and Mitchell (2018) and Ferber and Brün (2011) suggesting that citation practices disadvantage female faculty members. We also found evidence that economists had fewer citations, yet substantially higher wages than their colleagues from other disciplines.

Interestingly, our analysis did reveal a persistent pattern: economists in schools of public affairs were paid at a *significant discount*, while their political scientists' colleagues were paid at a *significant premium*, relative to their counterparts in disciplinary departments. In other words, salaries across disciplines within schools of public affairs were more homogeneous than were the

salaries across disciplines between disciplinary departments in the same universities. Furthermore, this pattern could not be explained by faculty sorting on gender.

Two competing hypotheses could explain the schools of public affairs salary discount for economists (and salary premium for political scientists). *First*, schools of public affairs could have been trying to sustain some outdated, internal equity standard in the face of rising relative salaries for economists (Curtis and Kisielewski, 2015). Buchanan and Tollison (1982) refer to this practice as the homogenization of heterogeneous inputs.

Second, both economist and political scientists who are employed at schools of public affairs are fairly quantitative in nature—that is, their scholarly work is inherently applied and some type of statistical modeling is common in their research. It is possible that the apparent public affairs discount is really just a manifestation of an applied micro or public finance discount. This would imply that there is an “applied discount” for economists who work at schools of public affairs, while their political scientists enjoy an “applied premium.”

Regardless of which competing hypothesis might be true, one thing is for certain, economists who chose public affairs departments experienced an estimated 18 percent salary loss relative to their peers in departments of economics. Card, Mas, Moretti, and Saez (2012) argued that relative perceptions of salary reduce job satisfaction for individuals earning below the median, with individuals also reporting a higher likelihood of seeking employment elsewhere. In order to facilitate intra-departmental cordiality, administrators may consider paying the economists on their faculty a wage that more accurately reflects the relative pay of their peers in economics departments.

References

- Barbezat, Debra A. (2004). A loyalty tax? National measures of academic salary compression. *Research in Higher Education*, 45(7), pp. 761-776.
- Bellas, Marcia L. (1994). Comparable worth in academia: The effects on faculty salaries of the sex composition and labor-market conditions of academic disciplines. *American Sociological Review*, 59(6), pp. 807-821.
- Bellas, Marcia L. (1997). Disciplinary differences in faculty salaries: Does gender bias play a role? *The Journal of Higher Education*, 68(3), pp. 299-321.
- Binder, M., K. Krause, J. Chermak, J. Thacher, and J. Gilroy. (2010). Same work, different pay? Evidence from a US public university. *Feminist Economics*, 16(4), pp. 105-135.
- Blau, F. D., and L. M. Kahn. (2017). The gender wage gap: Extent, trends, and explanations. *Journal of Economic Literature*, 55(3), pp. 789-865.
- Boudreau, Nancy, James Sullivan, William Balzer, Ann Marie Ryan, Robert Yonker, Todd Thorsteinson, and Peter Hutchinson. (1997). Should faculty rank be included as a predictor variable in studies of gender equity in university faculty salaries? *Research in Higher Education*, 38(3), pp. 297-312.
- Boyle, G. (2008). Pay peanuts and get monkeys? Evidence from academia. *The BE Journal of Economic Analysis & Policy*, 8(1), ISSN (Online) 1935-1682.
- Buchanan, James M., and Robert D. Tollison. (1981). The homogenization of heterogeneous inputs. *American Economic Review*, 71(1), pp. 28-38.
- Card, David, Alexandre Mas, Enrico Moretti, and Emmanuel Saez. (2012). Inequality at work: The effect of peer salaries on job satisfaction. *American Economic Review*, 102(6), pp. 2981-3003.
- Carlin, P. S., M. P. Kidd, P. M. Rooney, and B. Denton. (2013). Academic wage structure by gender: The roles of peer review, performance, and market forces. *Southern Economic Journal*, 80(1), pp. 127-146.
- Ceci, S. J., D. K. Ginther, S. Kahn, and W. M. Williams. (2014). Women in academic science: A changing landscape. *Psychological Science in the Public Interest*, 15(3), pp. 75-141.
- Claypool, Vicki Hesli, Brian David Janssen, Dongkyu Kim, and Sara McLaughlin Mitchell. (2017). Determinants of salary dispersion among political science faculty: The differential effects of where you work (institutional characteristics) and what you do (negotiate and publish). *PS: Political Science & Politics*, 50(1), pp. 146-156.
- Curtis, John, and Michael Kisielewski. (2015). The Need to Know: Faculty Salaries in Sociology and Other Disciplines. *Washington: American Sociological Association*.
- Diamond Jr., Arthur M. (1986). What is a citation worth? *The Journal of Human Resources*, 21(2), pp. 200-215.
- Dion, Michelle L, Jane Lawrence Sumner, and Sara McLaughlin Mitchell. (2018). Gendered citation patterns across political science and social science methodology fields. *Political Analysis*, 26(3), pp. 312-327.
- Ehrenberg, Ronald G. (2004a). Don't blame faculty for high tuition: The annual report on the economic status of the profession 2003-04. *Academe*, 90(2), pp. 19-103.
- Ehrenberg, Ronald G. (2004b). Prospects in the academic labor market for economists. *Journal of Economic Perspectives*, 18(2), pp. 227-238.

- Ehrenberg, Ronald G., Marquise McGraw, and Jesenka Mrdjenovic. (2006). Why do field differentials in average faculty salaries vary across universities? *Economics of Education Review*, 25(3), pp. 241-248.
- Fairweather, James S. (1995). Myths and realities of academic labor markets. *Economics of Education Review*, 14(2), pp. 179-192.
- Ferber, Marianne A, and Michael Brün. (2011). The gender gap in citations: does it persist? *Feminist Economics*, 17(1), pp. 151-158.
- Fourcade, Marion, Etienne Ollion, and Yann Algan. (2015). The superiority of economists. *Journal of Economic Perspectives*, 29(1), pp. 89-114.
- Freeman, Richard B. (1999). It's better being an economist (but don't tell anyone). *Journal of Economic Perspectives*, 13(3), pp.139-145.
- Friedman, Lee S. (1991). Economists and public policy programs. *Journal of Policy Analysis and Management*, 10(2), pp. 343-359.
- Gibson, J., D. L. Anderson, and J. Tressler. (2017). Citations or journal quality: Which is rewarded more in the academic labor market? *Economic Inquiry*, 55(4), pp. 1945-1965.
- Gibson, J., and E. Burton-McKenzie. (2017). Are returns to research quality lower in agricultural economics than in economics? *Australian Journal of Agricultural and Resource Economics*, 61(3), pp. 498-514.
- Ginther, Donna K. (2003). Is MIT an exception? Gender pay differences in academic science. *Bulletin of Science Technology & Society*, 23, pp. 21-26.
- Ginther, Donna K., and K. J. Hayes. (2003). Gender differences in salary and promotion for faculty in the humanities 1977-1995. *The Journal of Human Resources*, 38(1), pp. 34-73.
- 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), pp. 193-214.
- Gomez-Mejia, Luis R, and David B. Balkin. (1992). Determinants of faculty pay: An agency theory perspective. *Academy of Management Journal*, 35(5), pp. 921-955.
- Hamermesh, D. S. (2018). Citations in Economics: Measurement, Uses, and Impacts. *Journal of Economic Literature*, 56(1), pp. 115-156.
- Hamermesh, D. S., G. E. Johnson, and B. A. Weisbrod. (1982). Scholarship, citations and salaries: Economic rewards in economics. *Southern Economic Journal*, 49(2), pp. 472-481.
- Hilmer, C. E., M. J. Hilmer, and J. L. Lusk. (2012). A comparison of salary structures between economics and agricultural economics departments. *Applied Economic Perspectives and Policy*, 34(3), pp. 489-514.
- Langton, Nancy, and Jeffrey Pfeffer. (1994). Paying the professor: Sources of salary variation in academic labor markets. *American Sociological Review*, 59(2), pp. 236-256.
- Lundberg, S., and J. Stearns. (2019). Women in economics: Stalled progress. *Journal of Economic Perspectives*, 33(1), pp. 3-22.
- Masakure, O. (2016). The effect of employee loyalty on wages. *Journal of Economic Psychology*, 56, pp. 274-298.
- Park, Toby. (2012). Do faculty members get what they deserve?: A review of the literature surrounding the determinants of salary, promotion and tenure. *Journal of the Professoriate*, 6(1), pp. 28-47.
- Perna, Laura W. (2003). Studying faculty salary equity: A review of theoretical and methodological approaches. In *Higher education: Handbook of theory and research*, pp. 323-388. Springer.

- Renzulli, L. A., J. Reynolds, K. Kelly, and L. Grant. (2013). Pathways to gender inequality in faculty pay: the impact of institution, academic division, and rank. *Research in Social Stratification and Mobility*, 34, pp. 58-72.
- Sarsons, H. (2017). Recognition for group work: Gender differences in academia. *American Economic Review*, 107(5), pp. 141-145.
- Seaman, S. R., J. W. Bartlett, and I. R. White. (2012). Multiple imputation of missing covariates with non-linear effects and interactions: an evaluation of statistical methods. *BMC Medical Research Methodology*, 12(1), pp. 46.
- Smart, John C. (1991). Gender equity in academic rank and salary. *The Review of Higher Education*, 14(4), pp. 511-525.
- Tuckman, B. H., and H. P. Tuckman. (1976). The structure of salaries at American universities. *The Journal of Higher Education*, 47(1), pp. 51-64.
- Umbach, Paul D. (2007). Gender equity in the academic labor market: An analysis of academic disciplines. *Research in Higher Education*, 48(2), pp. 169-192.
- Volk, Cindy S, Sheila Slaughter, and Scott L. Thomas. (2001). Models of institutional resource allocation: Mission, market, and gender. *The Journal of Higher Education*, 72(4), pp. 387-413.
- Von Hippel, P. T. (2009). How to impute interactions, squares, and other transformed variables. *Sociological Methodology*, 39(1), pp. 265-291.
- Webber, K. L., and M. González Canché. (2015). Not equal for all: Gender and race differences in salary for doctoral degree recipients. *Research in Higher Education*, 56(7), pp. 645-672.
- Wu, Alice H. (2018). Gendered Language on the Economics Job Market Rumors Forum. *American Economic Review Papers and Proceedings*, 108, pp. 175-179.

Table 1: Sample Descriptive Statistics

	No. Observations	Mean	Minimum	Maximum
Salary (log)	2,152	11.81 (0.42)	10.60	13.14
School of Public Affairs	2,152	0.32	0	1
Economics Department	2,152	0.34	0	1
Political Science Department	2,152	0.34	0	1
Economist	2,152	0.39	0	1
Political Scientist	2,152	0.40	0	1
Other Fields	2,152	0.21	0	1
Female Faculty	2,152	0.30	0	1
Male Faculty	2,152	0.70	0	1
Assistant Professor	2,152	0.26	0	1
Associate Professor	2,152	0.31	0	1
Full Professor	2,152	0.43	0	1
Year of Ph.D. Degree	1,981	1998.79 (12.84)	1957	2017
Google Scholar's Total Citation Count	1,337	2992.02 (6331.08)	1	100367
Google Scholar's h-index	1,337	17.20 (13.52)	1	102
Google Scholar's i10-index	1,337	26.12 (31.72)	0	281

Notes: Standard deviations are in parentheses for continuous variables. The variable "Other Fields" include other fields whose faculty members are in the school of public affairs, such as, sociology, demography, anthropology, etc. Productivity measures (i.e., total citation count, h-index, and i10-index) were obtained from each faculty's Google Scholar profiles. See Appendix 1 for detailed description of variables listed in Table 1.

Source: Authors' calculations.

Table 2: The Distribution of Faculty by Departmental Affiliation, Gender, and Annual Salary Compensation

Departmental Affiliation	Panel A: Faculty by Department and Gender				Panel B: Compensation (\$000's)	
	No. of Faculty Members	Percent Female	No. of Economists	Percent Female	Average Salary	Average Economist Salary
School Public Affairs	701	40.7%	102	35.3%	\$134	\$161
Economics	685	18.7%	685	18.7%	\$179	\$179
Political Science	666	31.1%	13	23.1%	\$126	\$183
Joint Appointments	100	27.0%	46	21.7%	\$168	\$199
Total	2,152	30.1%	846	20.9%	\$147	\$178

Notes: Annual salary compensation are in current dollars, circa 2017.

Source: Authors' calculations.

Table 3: Ordinary Least Squares (OLS) Regression Results - Hedonic Wage Models

	(1)	(2)	(3)	(4)	(5)
Economics Department	0.154*** (0.033)	0.133*** (0.0329)	0.117*** (0.041)	0.154*** (0.034)	0.147*** (0.033)
Political Science Department	-0.119*** (0.035)	-0.126*** (0.0353)	-0.0896** (0.045)	-0.0875*** (0.033)	-0.0956*** (0.031)
Economist	0.124*** (0.033)	0.119*** (0.0326)	0.123*** (0.041)	0.243*** (0.071)	0.247*** (0.070)
Political Scientist	0.0257 (0.035)	0.0208 (0.0353)	-0.00707 (0.046)	-0.0714 (0.071)	-0.0728 (0.070)
Female Faculty		-0.117*** (0.0159)	-0.124*** (0.030)	-0.0534** (0.027)	-0.0636** (0.025)
Female Faculty x Economist			-0.016 (0.064)	0.0201 (0.058)	0.0221 (0.056)
Female Faculty x Political Scientist			0.0737 (0.069)	0.0618 (0.055)	0.0784 (0.052)
Female Faculty x Economics Department			0.0675 (0.068)	0.0029 (0.059)	0.0152 (0.057)
Female Faculty x Political Science Department			-0.103 (0.067)	-0.0662 (0.052)	-0.0666 (0.051)
Years Since PhD Degree (log)				0.00377 (0.027)	0.00484 (0.027)
Years Since PhD Degree (log) x Economist				-0.0748** (0.033)	-0.0764** (0.034)
Years Since PhD Degree (log) x Political Scientist				0.0402 (0.034)	0.0399 (0.034)
Associate Professor				0.267*** (0.042)	0.265*** (0.041)
Associate Professor x Economist				0.00967 (0.054)	0.00927 (0.052)
Associate Professor x Political Scientist				-0.124** (0.053)	-0.123** (0.051)
Full Professor				0.557*** (0.057)	0.528*** (0.056)
Full Professor x Economist				0.0963 (0.069)	0.118* (0.068)
Full Professor x Political Scientist				-0.0647 (0.069)	-0.0457 (0.068)
Constant	11.74*** (0.016)	11.79*** (0.0169)	11.79*** (0.020)	11.45*** (0.053)	11.46*** (0.053)
University Fixed Effects	Yes	Yes	Yes	Yes	Yes
Imputed Data	No	No	No	No	Yes
Number of Imputed Iterations	--	--	--	--	100
No. Observations	2,152	2,152	2,152	1,981	2,152
Adjusted R ²	0.34	0.35	0.36	0.63	--

Notes: Standard errors (shown in parentheses). The omitted comparison category for professor's rank is Assistant Professor. For column (5), the variables "Years Since PhD Degree (log)" and corresponding interaction terms were imputed with 100 iterations. See Appendix 1 for detailed description of variables listed in Table 3. ***, **, * denotes statistical significant at the 1 percent, 5 percent, 10 percent levels, respectively.

Table 4: Imputing Google Scholar Profile Citation Metrics

	Panel A: Logit Models		Panel B: OLS Models		Panel C: OLS Models		Panel D: OLS Models	
	Probability of a Missing Google Scholar Profile		Log Citation Count		h-index		i10-index	
	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
Economics Department	0.181	0.123	-0.063	-0.060	-4.043***	-3.930***	-18.19***	-17.90***
	(0.178)	(0.170)	(0.141)	(0.140)	(1.372)	(1.333)	(3.605)	(3.502)
Political Science Department	-0.124	-0.094	0.031	0.035	-0.997	-0.648	-3.095	-2.15
	(0.166)	(0.158)	(0.155)	(0.150)	(1.505)	(1.431)	(3.954)	(3.758)
Economist	0.240	0.202	-0.691**	-0.735**	-6.250*	-6.431**	-9.079	-9.591
	(0.373)	(0.364)	(0.333)	(0.334)	(3.232)	(3.145)	(8.491)	(8.270)
Political Scientist	0.156	0.033	-0.213	-0.265	-4.319	-4.691	-9.491	-10.59
	(0.372)	(0.364)	(0.337)	(0.338)	(3.275)	(3.178)	(8.604)	(8.338)
Female Faculty	0.037	0.066	-0.170	-0.132	-1.142	-0.856	-4.13	-3.704
	(0.140)	(0.126)	(0.121)	(0.114)	(1.172)	(1.089)	(3.079)	(2.859)
Female Faculty x Economist	0.202	0.139	0.105	0.078	-4.631*	-4.619*	-18.79***	-18.55***
	(0.303)	(0.289)	(0.255)	(0.250)	(2.476)	(2.382)	(6.504)	(6.255)
Female Faculty x Political Scientist	-0.182	-0.082	0.255	0.226	0.141	0.136	0.408	0.784
	(0.278)	(0.262)	(0.252)	(0.246)	(2.448)	(2.346)	(6.430)	(6.160)
Female Faculty x Economics Department	0.075	0.092	0.022	0.020	3.734	3.497	17.38***	16.76***
	(0.306)	(0.295)	(0.258)	(0.254)	(2.500)	(2.421)	(6.567)	(6.359)
Female Faculty x Political Science Department	0.374	0.213	-0.497**	-0.451*	-1.459	-1.674	-0.878	-1.679
	(0.267)	(0.255)	(0.247)	(0.243)	(2.398)	(2.314)	(6.299)	(6.076)
Years Since PhD Degree (log)	0.514***	0.467***	0.757***	0.713***	4.181***	3.868***	8.675**	7.971**
	(0.140)	(0.139)	(0.136)	(0.139)	(1.324)	(1.285)	(3.478)	(3.379)
Years Since PhD Degree (log) x Economist	-0.123	-0.078	0.340**	0.363**	4.661***	4.772***	12.99***	13.21***
	(0.172)	(0.171)	(0.165)	(0.166)	(1.598)	(1.555)	(4.198)	(4.094)
Years Since PhD Degree (log) x Political Scientist	0.054	0.077	0.07	0.092	1.955	2.095	4.872	5.155
	(0.174)	(0.174)	(0.167)	(0.169)	(1.620)	(1.580)	(4.257)	(4.149)
Associate Professor	-0.175	-0.073	1.010***	1.055***	2.604	2.993*	2.675	3.306
	(0.220)	(0.209)	(0.185)	(0.184)	(1.798)	(1.730)	(4.722)	(4.546)
Associate Professor x Economist	-0.298	-0.450*	-0.382*	-0.391*	-4.001*	-4.201*	-12.89**	-13.14**
	(0.280)	(0.271)	(0.230)	(0.228)	(2.229)	(2.160)	(5.856)	(5.677)
Associate Professor x Political Scientist	-0.131	-0.197	-0.363	-0.393*	-3.068	-3.443	-8.051	-8.691
	(0.273)	(0.263)	(0.230)	(0.228)	(2.234)	(2.154)	(5.869)	(5.657)
Full Professor	-0.543*	-0.405	1.639***	1.711***	12.80***	13.42***	26.32***	27.56***
	(0.293)	(0.284)	(0.267)	(0.267)	(2.595)	(2.498)	(6.816)	(6.569)
Full Professor x Economist	-0.0898	-0.255	-0.234	-0.276	-4.186	-4.701	-15.37*	-16.52**
	(0.355)	(0.347)	(0.316)	(0.315)	(3.068)	(2.966)	(8.059)	(7.802)
Full Professor x Political Scientist	0.07	0.0253	-0.255	-0.272	-5.950*	-6.242**	-17.33**	-17.88**
	(0.355)	(0.347)	(0.319)	(0.318)	(3.093)	(2.996)	(8.125)	(7.869)
Constant	-0.895**	-0.822**	4.060***	4.119***	2.45	2.812	-2.428	-1.379
	(0.367)	(0.329)	(0.250)	(0.252)	(2.428)	(2.353)	(6.377)	(6.185)
F-test: Female and All Interactions			F(5,1211)=3.37	F(5,1284)=2.63	F(5,1211)=3.29	F(5,1284)=3.21	F(5,1211)=4.66	F(5,1284)=4.71
P-values			p=0.005	p=0.020	p=0.006	p=0.007	p=0.003	p=0.003
University Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Imputed Data	No	Yes	No	Yes	No	Yes	No	Yes
Number of Imputed Interactions	--	100	--	100	--	100	--	100
No. Observations	1,981	2,152	1,262	1,337	1,262	1,337	1,262	1,337
Adjusted R ²	--	--	0.71	--	0.57	--	0.46	--

Notes: Standard errors (shown in parentheses). The omitted comparison category for professor's rank is Assistant Professor. For each model reported in column (2), the variables "Years Since PhD Degree (log)" and corresponding interaction terms were imputed with 100 iterations. See Appendix 1 for detailed description of variables listed in Table 4. ***, **, * denotes statistical significant at the 1 percent, 5 percent, 10 percent levels, respectively.

Table 5: OLS Regression Results - Hedonic Wage Models with Citation Measures

	(1)	(2)	(3)	(4)	(5)	(6)
Economics Department	0.210*** (0.029)	0.168*** (0.028)	0.159*** (0.027)	0.179*** (0.027)	0.187*** (0.028)	0.190*** (0.0303)
Political Science Department	-0.0679** (0.030)	-0.100*** (0.026)	-0.104*** (0.026)	-0.100*** (0.026)	-0.103*** (0.026)	-0.0874*** (0.0292)
Economist	0.173** (0.080)	0.295*** (0.068)	0.278*** (0.068)	0.286*** (0.068)	0.267*** (0.068)	0.299*** (0.0680)
Political Scientist	-0.110 (0.080)	-0.041 (0.068)	-0.058 (0.068)	-0.037 (0.068)	-0.040 (0.068)	-0.0379 (0.0680)
Female Faculty	-0.0675** (0.030)	-0.0438* (0.026)	-0.0434 (0.027)	-0.0432 (0.026)	-0.0427 (0.027)	-0.0450* (0.0265)
Female Faculty x Economist	0.016 (0.039)	0.028 (0.035)	0.015 (0.035)	0.032 (0.034)	0.032 (0.035)	0.0294 (0.0346)
Female Faculty x Political Scientist	0.0756** (0.037)	0.0212 (0.033)	0.0197 (0.033)	0.0159 (0.033)	0.00837 (0.033)	0.0224 (0.0327)
Years Since PhD Degree (log)	-0.00922 (0.034)	-0.0547** (0.028)	-0.0546* (0.028)	-0.0428 (0.027)	-0.026 (0.027)	-0.0538* (0.0276)
Years Since PhD Degree (log) x Economist	-0.0609 (0.040)	-0.0981*** (0.033)	-0.0891*** (0.033)	-0.101*** (0.033)	-0.0986*** (0.033)	-0.0963*** (0.0327)
Years Since PhD Degree (log) x Political Scientist	0.0452 (0.041)	0.0384 (0.033)	0.0438 (0.033)	0.0362 (0.033)	0.036 (0.033)	0.0375 (0.0330)
Associate Professor	0.180*** (0.046)	0.241*** (0.042)	0.225*** (0.042)	0.253*** (0.041)	0.262*** (0.042)	0.240*** (0.0416)
Associate Professor x Economist	0.0508 (0.056)	0.0237 (0.052)	0.0212 (0.053)	0.0275 (0.052)	0.0302 (0.053)	0.0206 (0.0523)
Associate Professor x Political Scientist	-0.0824 (0.056)	-0.110** (0.051)	-0.112** (0.052)	-0.109** (0.052)	-0.110** (0.052)	-0.109** (0.0514)
Full Professor	0.385*** (0.066)	0.489*** (0.056)	0.497*** (0.057)	0.501*** (0.056)	0.520*** (0.056)	0.490*** (0.0559)
Full Professor x Economist	0.104 (0.077)	0.0936 (0.067)	0.093 (0.068)	0.102 (0.067)	0.112* (0.068)	0.0873 (0.0670)
Full Professor x Political Scientist	-0.0263 (0.078)	-0.053 (0.067)	-0.0647 (0.068)	-0.0462 (0.068)	-0.0436 (0.068)	-0.0532 (0.0675)
Google Scholar's Total Citation Count (log)	0.0178* (0.009)	0.0154 (0.011)	0.0592*** (0.008)			0.0155 (0.0107)
Google Scholar's h-index	0.0102*** -0.0021	0.00996*** -0.00237		0.00752*** -0.000805		0.00982*** (0.00238)
Google Scholar's i10-index	-0.00176** (0.0007)	-0.00145* (0.0008)			0.00237*** (0.0003)	-0.00141* (0.000802)
US News Ranking in Economics x Econ Dept.						-0.000397* (0.000226)
US News Ranking in Political Science x Poli Sci Dept.						-0.000145 (0.000239)
Constant	11.34*** (0.071)	11.39*** (0.065)	11.23*** (0.061)	11.46*** (0.052)	11.47*** (0.052)	11.38*** (0.0651)
F-test: Female and All Interactions	F(3,1210)=2.95	F(3,1919.7)=1.46	F(3,1926.6)=1.77	F(3,1923.3)=1.51	F(3,1926.6)=1.80	F(3,1915.7)=1.46
P-values	p=0.03	p=0.22	p=0.15	p=0.21	p=0.14	p=0.22
University Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Imputed Data	No	Yes	Yes	Yes	Yes	Yes
Number of Imputed Iterations	--	100	100	100	100	100
No. Observations	1,262	1,981	1,981	1,981	1,981	1,981

Notes: Standard errors (shown in parentheses). The omitted comparison category for faculty's rank is Assistant Professor. For columns (2)-(6), the variables Years Since PhD Degree (log), Google Scholar's Total Citation Count, h-index, and i10-index were imputed with 100 iterations. See Appendix 1 for detailed description of variables listed in Table 5. ***, **, * denotes statistical significant at the 1 percent, 5 percent, 10 percent levels, respectively.

Table 6: OLS Regression Results - Differential Returns to Citations (Google Scholar's Total Citation Count) by Gender and Field

	Panel A: By Gender		Panel B: By Field		Panel C: By Gender and Field	
	(1)	(2)	(1)	(2)	(1)	(2)
Economics Department	0.205*** (0.030)	0.159*** (0.027)	0.211*** (0.029)	0.165*** (0.027)	0.209*** (0.029)	0.163*** (0.027)
Political Science Department	-0.0750** (0.031)	-0.108*** (0.026)	-0.0740** (0.030)	-0.106*** (0.026)	-0.0820*** (0.030)	-0.113*** (0.026)
Economist	0.173** (0.082)	0.298*** (0.069)	-0.137 (0.100)	0.0474 (0.094)	-0.167 (0.114)	0.055 (0.109)
Political Scientist	-0.118 (0.082)	-0.0463 (0.069)	-0.343*** (0.101)	-0.203** (0.095)	-0.462*** (0.119)	-0.270** (0.114)
Female Faculty	0.0492 (0.068)	0.0784 (0.070)	-0.0688** (0.030)	-0.0479* (0.027)	-0.145 (0.120)	-0.0431 (0.129)
Female Faculty x Economist	0.00707 (0.039)	0.0186 (0.035)	0.00822 (0.039)	0.0217 (0.035)	0.0711 (0.160)	-0.00653 (0.172)
Female Faculty x Political Scientist	0.0695* (0.038)	0.0166 (0.033)	0.0765** (0.038)	0.0248 (0.033)	0.422*** (0.155)	0.292* (0.167)
Google Scholar's Total Citation Count (log)	0.0639*** (0.008)	0.0645*** (0.009)	0.00186 (0.014)	0.0179 (0.016)	-0.0032 (0.016)	0.0185 (0.019)
Female Faculty x Total Citation Count (log)	-0.0175* (0.009)	-0.0187* (0.010)			0.0117 (0.018)	-0.000772 (0.019)
Economist x Total Citation Count (log)			0.0812*** (0.017)	0.0674*** (0.019)	0.0863*** (0.019)	0.0666*** (0.021)
Political Scientist x Total Citation Count (log)			0.0605*** (0.017)	0.0422** (0.019)	0.0803*** (0.019)	0.0555** (0.022)
Female x Economist x Total Citation Count (log)					-0.0095 (0.024)	0.00432 (0.025)
Female x Political Scientist x Total Citation Count (log)					-0.0538** (0.023)	-0.0418* (0.025)
Constant	11.16*** (0.073)	11.19*** (0.065)	11.40*** (0.080)	11.37*** (0.078)	11.44*** (0.096)	11.37*** (0.096)
F-test: Female and All Interactions	F(4,1211)=2.11	F(4,1914.3)=1.22	F(3,1210)=2.63	F(3,1920.8)=1.44	F(6,1207)=2.39	F(6,1902.8)=1.58
P-values	p=0.08	p=0.30	p=0.05	p=0.23	p=0.03	p=0.15
University Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Faculty Characteristic Controls ^a	Yes	Yes	Yes	Yes	Yes	Yes
Imputed Data	No	Yes	No	Yes	No	Yes
Number of Imputed Interactions	--	100	--	100	--	100
No. Observations	1,262	1,981	1,262	1,981	1,262	1,916
Adjusted R ²	0.698	--	0.703	--	0.705	--

Notes: Robust standard errors (shown in parentheses). The omitted comparison category for faculty's rank is Assistant Professor. For column (2) in panels A, B, and C, the variables Years Since PhD Degree (log) and Google Scholar's Total Citation Count were imputed with 100 iterations. ^aFaculty characteristics controls include: years since PhD degree (log) and interactions with field of degree; professor rank and interactions with field of degree. See Appendix 1 for detailed description of variables listed in Table 6. ***, **, * denotes statistical significant at the 1 percent, 5 percent, 10 percent levels, respectively.

Appendix 1: Description of Variables

Variable Type	Variable Description
Annual Salary Compensation (log)	<u><i>Dependent Variable</i></u> : Annual salary in current dollars, as publicly reported by the respective institutions of higher education (IHEs). In cases where multiple salary figures were available, this indicator was full-time equivalent salary including reported benefits. The data represent the most recent reporting year as of July 2018 (which in most cases was 2017). Inconsistencies in reporting necessitate the use of IHE fixed effects for all salary models.
Departmental Affiliation Indicators	School Public Affairs (<i>omitted category</i>) Economics Political Science
Degree Field Indicators	Other Fields (<i>omitted category</i>) Economics Political Science
Gender Indicators	Male Faculty (<i>omitted category</i>) Female Faculty
Rank Indicators	Assistant Professor (<i>omitted category</i>) Associate Professor Full Professor
Experience	Years Since PhD Degree (log)
Productivity Measures	<u><i>Total Citation Count</i></u> : Total number of citations. <u><i>h-index</i></u> : Characterizes faculty's productivity in the field by combining both the breadth and depth of impact of their scholarly work (Hamermesh, 2018). The h-index is calculated by counting the number of publications for which a scholar has been cited by other scholars at least that same number of times. For example, a h-index of 16 means that the scholar has published at least 16 papers that have each been cited at least 16 times. <u><i>i10-index</i></u> : Measures the number of papers with at least 10 citations. For example, an i10-index of 12 means that 12 publications of the scholar have each been cited at least 10 times.

Notes: The productivity measures are from each faculty's Google Scholar profile as of November 2018. "Other Fields" include other fields whose faculty members are in the school of public affairs, such as, sociology, demography, anthropology, etc.

Source: Publicly available data collected from a variety of sources.

Appendix 2: List of Institution Names by Alphabetical Order

	Name of Institution	No. of Faculty Members	Percent of Faculty Members
1	Arizona State University	68	3.16
2	Baruch College	80	3.72
3	Florida State University	52	2.42
4	Georgia Institute of Technology	39	1.81
5	George Mason University	91	4.23
6	Georgia State University	80	3.72
7	Indiana University Bloomington	96	4.46
8	John Jay College	46	2.14
9	North Carolina State University	42	1.95
10	Ohio State University	73	3.39
11	Rutgers University - Newark	49	2.28
12	Texas A&M University - College Station	93	4.32
13	University at Albany - State University of New York (SUNY)	52	2.42
14	University of Arizona	47	2.18
15	University of California - Berkeley	89	4.14
16	University of California - Los Angeles	129	5.99
17	University of Central Florida	63	2.93
18	University of Colorado - Denver	41	1.91
19	University of Georgia	52	2.42
20	University of Illinois - Chicago	39	1.81
21	University of Kansas	37	1.72
22	University of Kentucky	29	1.35
23	University of Maryland - College Park	71	3.3
24	University of Michigan	113	5.25
25	University of Minnesota	72	3.35
26	University of Missouri - Columbia	35	1.63
27	University of Nebraska - Omaha	30	1.39
28	University of North Carolina - Chapel Hill	62	2.88
29	University of Texas - Austin	100	4.65
30	University of Virginia	94	4.37
31	University of Washington	73	3.39
32	University of Wisconsin - Madison	71	3.30
33	Virginia Commonwealth University	44	2.04
No. Observations		2,152	100.00

Source: Authors' calculations.

Equal but Inequitable: Who Benefits from Gender-Neutral Tenure Clock Stopping Policies?[†]

By HEATHER ANTECOL, KELLY BEDARD, AND JENNA STEARNS*

Many skilled professional occupations are characterized by an early period of intensive skill accumulation and career establishment. Examples include law firm associates, surgical residents, and untenured faculty at research-intensive universities. High female exit rates are sometimes blamed on the inability of new mothers to survive the sustained negative productivity shock associated with childbearing and early childrearing in these environments. Gender-neutral family policies have been adopted in some professions in an attempt to “level the playing field.” The gender-neutral tenure clock stopping policies adopted by the majority of research-intensive universities in the United States in recent decades are an excellent example. But to date, there is no empirical evidence showing that these policies help women. Using a unique dataset on the universe of assistant professor hires at top-50 economics departments from 1980–2005, we show that the adoption of gender-neutral tenure clock stopping policies substantially reduced female tenure rates while substantially increasing male tenure rates. However, these policies do not reduce the probability that either men or women eventually earn tenure in the profession. (JEL I23, J16, J24, J32, J44)

Women are underrepresented in many high-skilled professions at labor market entry and leave these professions at higher rates than men as they age. Figure 1 documents female representation among physicians, lawyers, business professionals, and postsecondary teachers with advanced degrees. Women are underrepresented in these occupations at entry, more underrepresented as they age, and the decline with age is particularly pronounced among top earners. The widening gender gap in the share of top earners over the period when many professionals have children, can

*Antecol: Robert Day School of Economics and Finance, Claremont McKenna College, Claremont, CA 91711 (email: hantecol@cmc.edu); Bedard: Department of Economics, University of California, Santa Barbara, Santa Barbara, CA 93106 (email: kelly.bedard@ucsb.edu); Stearns: Department of Economics, University of California, Davis, Davis, CA 95616 (email: jestearns@ucdavis.edu). This paper was accepted to the *AER* under the guidance of Hilary Hoynes, Coeditor. We are grateful to Shelly Lundberg, Heather Royer, Dick Startz, and Catherine Weinberger, as well as seminar participants at the 2016 Society of Labor Economist Meetings, the 2016 All California Labor Conference, the 2017 American Economics Association Meetings, the University of Houston, Purdue University, and Uppsala University for helpful comments and suggestions. We also thank Sabrina Benyammi, Chang Lee, Jorge Dominguez Rodriguez, and Nicole Stedman for helpful research assistance. The authors declare that they have no relevant or material financial interests that relate to the research described in this paper.

[†]Go to <https://doi.org/10.1257/aer.20160613> to visit the article page for additional materials and author disclosure statement(s).

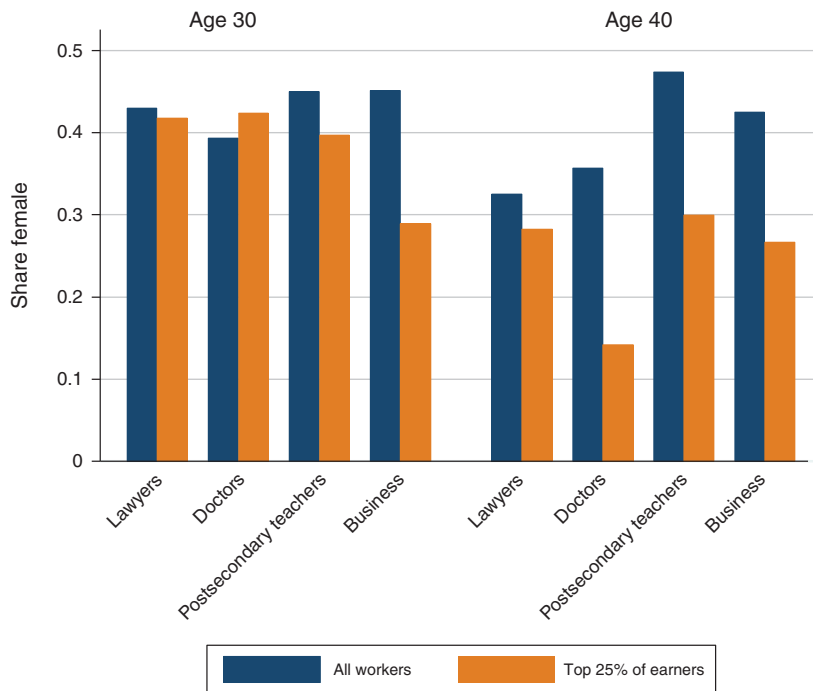


FIGURE 1. SHARE FEMALE AMONG FULL-TIME WORKERS WITH ADVANCED DEGREES IN SELECTED HIGH SKILL OCCUPATIONS

Notes: Figure shows individuals born in 1970 at age 30 and age 40. Data at age 30 come from the 2000 Census and data at age 40 come from the 2010 American Community Survey. Sample is conditional on having a master's, professional, or doctoral degree. Sample is restricted to workers who report working 30 hours per week or more. Top earners are defined as those in the top quartile of the occupation-specific wage income distribution among workers with advanced degrees. Among the cohort born in 1970, women hold 54 percent of all master's, professional, and doctoral degrees obtained by age 30.

be interpreted as a “family gap,” and suggests that even professional women have trouble maintaining a successful career while raising a family.

An important feature of many high skill occupations is that human capital accumulation and promotion tracks are very steep during the first decade of one's career. Lawyers, academics, and executives often work in “up-or-out” environments, where individuals who miss a set window of opportunity for advancement are never able to do so (Rosen 1990; Demougin and Siow 1994; O'Flaherty and Siow 1995). Having children may therefore reduce the probability that women are promoted because early productivity falls. Short family leave policies do not mitigate this issue because they do not adequately account for the prolonged productivity loss associated with starting a family.¹ This problem is especially acute at research-intensive universities, where tenure decisions are made at a fixed point in time. For this reason,

¹There is also a high wage penalty associated with taking time off in high skill occupations (Goldin and Katz 2011). A particular concern with female-specific policies is the belief that managers and colleagues will view women who use them as less committed to the job or less productive. Some have argued that gender-neutral policies may benefit women through reducing social stigma against use (Rehel and Baxter 2015; Lazear and Rosen 1990; Dahl, Løken, and Mogstad 2014).

many universities have adopted gender-neutral tenure clock stopping policies for family-related reasons (American Association of University Professors 2001).

We study the tenure clock stopping policies adopted by many universities over recent decades. These policies are unusual in that assistant professors are not required to take time off. Instead, they are allowed to stop their tenure clock for one year after childbirth or adoption. In theory, no research is expected during this time. This type of policy better accounts for the extended period of reduced productivity associated with having a child. However, gender-neutral clock stopping policies will not level the playing field in terms of tenure outcomes if men are able to use the extra time more productively or strategically than women. In this case, these policies could actually increase the gender gaps in the profession. Yet to date, there is little empirical evidence about the effects of tenure clock stopping policies on tenure outcomes.

We have compiled a unique dataset on the professional and publication outcomes of all assistant professors hired at the top-50 economics departments from 1980–2005 to study the impact of these policies on tenure rates. We find that men are 17 percentage points more likely to get tenure in their first job once there is an established gender-neutral clock stopping policy in place, while women are 19 percentage points less likely. These policies substantially increase the gender gap in tenure rates. The primary mechanism driving these effects is an increase in the number of top-5 journal publications by men with no such increase by women. However, this does not mean that more women leave academia. Among those who start their career at a top-50 department, we find no evidence that gender-neutral clock stopping policies reduce the share of women who ultimately get tenure in the profession. We also find no consistent evidence that women are either hurt or helped by clock stopping policies that only apply to women.

I. Tenure Clock Stopping Policies

Family-friendly policies are designed to allow new parents to better balance family and career, but little is known about the causal effects of access to these policies for high-skill American professionals.² Yet there are strong correlations between job flexibility and the gender wage gap, which suggests that women value amenities such as the ability to take leave or have flexible hours (Goldin 2014). At the same time, Goldin and Katz (2012) show that as pharmacy became a more family-friendly occupation, the share of female workers rose and the gender wage gap disappeared. American universities have also become more family-friendly. Many have recently adopted family-friendly policies including short term family leave, allowing part-time work, and allowing assistant professors to stop their tenure clocks. However, there is little empirical evidence showing how these policies affect productivity and employment outcomes.

²Family leave can reduce the “family gap” in wages by improving job continuity (Waldfogel 1998; Lundberg and Rose 2000) and increasing female employment rates (Ruhm 1998; Misra, Budig, and Böckmann 2011; Stearns 2018). Paternity leave may increase childcare time among fathers even several years later (Schober 2014). Other studies do not find any behavioral effects within the household (Ekberg, Eriksson, and Friebe 2013). These policies may also have unintended consequences. Blau and Kahn (2013) find that these policies lead to higher rates of part-time work and employment in lower-level positions. The effects of US paid family leave benefits are concentrated among less advantaged groups (Rossin-Slater, Ruhm, and Waldfogel 2013).

Assistant professors are evaluated for tenure near the end of a fixed probationary period of about seven years, giving them time to produce a portfolio of work that signals their true productivity.³ A fixed probationary period disadvantages people who experience large, temporary negative productivity shocks, such as having children. In part for this reason, women are less likely than men to have children before tenure, and the gender gap in tenure rates is larger for those with children (Mason and Goulden 2002).

To better account for these productivity shocks when making tenure decisions, and to reduce the work-family tradeoff, tenure clock stopping policies have become common at universities over the past 30 years. These policies allow eligible assistant professors to stop their tenure clock for one year.⁴ Importantly, such policies are independent of leave-taking; meaning that assistant professors do not face a tradeoff between forgoing income while on leave and gaining the extra time on their tenure clock.⁵ Tenure clock stopping policies typically cover having children in addition to serious illness or personal issues and caring for sick or elderly relatives. While clock stopping for non-birth related reasons requires dean approval, time exclusions for new parents are automatically approved at many universities.⁶ If the clock is stopped, departments and outside letter-writers are supposed to fully discount the additional time spent as an assistant professor. However, it is not clear how these individuals are actually evaluated.⁷

These clock stopping policies can be classified into two main types: policies that only apply to mothers, and policies that are gender neutral. Gender-neutral policies extend equal benefits to new mothers and fathers, whereas female-only policies are only available to women.⁸ Both types of policies typically stop the tenure clock for one year for each new child, up to a maximum of two. Table 1 reports gender-neutral and female-only policy adoption dates for the universities in our sample. Early policies were more likely to be female-only as women shoulder more of the burden of bearing and caring for children, and therefore face a higher cost of having children early in their careers. However, gender-neutral tenure clock stopping policies are now more common, and many universities that originally adopted a female-only policy have since converted to a gender-neutral version.

³The American Association of University Professors (2014) guidelines state that the probationary period should not exceed seven years. This means that assistant professors are usually evaluated for tenure during year six.

⁴Some universities independently provide course reductions to new parents, but because it is often at the discretion of the department chair (e.g., Mason et al. 2005) it is not considered here. There is little evidence that the introduction of course reduction policies are correlated at the university level. To the best of our knowledge, only 4 out of the 51 universities in our sample implement or change a course reduction policy alongside a gender-neutral tenure clock stopping policy.

⁵Clock stopping policies are also called tenure clock extension or rollback policies. Some universities do not have a clock stopping policy, but do allow for clock stopping only in the case of a leave of absence. These are not considered in this analysis.

⁶However, parents can choose to go up for tenure at the original time (or earlier), or to opt out of the policy.

⁷Thornton (2009) reports that 37 percent of economics tenure case reviewers do not know the applicable rule for clock stopping adjustment. Manchester, Leslie, and Kramer (2013) find that reviewers view clock stopping as a negative signal of academic commitment. Finkel, Olswang, and She (1994) find a large share of female faculty believe that using family friendly policies will hurt their career.

⁸What we call female-only policies are sometimes instead called primary caregiver policies, and cover parents of either gender who provide more than 51 percent of caregiving activities. Gender-neutral policies sometimes require parents to have "significant" childcare responsibilities, but can be used simultaneously by both parents. While in theory primary caregiver policies could be used by men, reports indicate that male take-up is rare. Female take-up is also low, and one of the common reasons women cite for not stopping their clock is that they thought it would have a negative impact on their tenure case (Mason et al. 2005).

TABLE 1—TENURE CLOCK STOPPING POLICIES AT TOP-50 UNIVERSITIES

	Female-only	Gender-neutral		Female-only	Gender-neutral
Arizona State	2003		University of California-Berkeley	1988	1997
Boston College		1993	University of California-Davis	1988	1997
Boston University			University of California-Irvine	1988	1997
Carnegie Mellon University	1992		University of California-Los Angeles	1988	1997
Columbia University	1995		University of California-San Diego	1988	1997
Cornell University	2005		University of California-Santa Barbara	1988	1997
Duke University		2003	University of Chicago	1996	
Georgetown University		1999	University of Florida		2004
Harvard University		2001	University of Illinois-Urbana-Champaign		1999
Indiana University-Bloomington			University of Iowa		1993
Johns Hopkins University		2000	University of Maryland-College Park		1996
Massachusetts Institute of Technology		2001	University of Michigan-Ann Arbor		1990
Michigan State University			University of Minnesota-Twin Cities		1992
New York University	1999		University of North Carolina-Chapel Hill		2004
North Carolina State University-Raleigh		2002	University of Pennsylvania		1997
Northwestern University		1981	University of Rochester		
Ohio State University		1996	University of Southern California		1993
Pennsylvania State University	1990		University of Texas-Austin	1997	
Princeton University	1970	1991	University of Virginia	1987	2000
Purdue University		1991	University of Washington		
Rice University		1993	University of Wisconsin-Madison		1994
Rutgers University		1992	Vanderbilt University		2003
Stanford University	1971	2002	Washington University in St. Louis		
Texas A&M University-College Station		2003	Yale University		
University of Arizona		1990			

Notes: Table shows policies adopted by 2005. Policies adopted after 2005 or that only stop the tenure clock if parental leave is taken are not shown. The sample includes 49 of the top 50 economics departments with graduate programs as ranked by the 2010 *US News and World Report*, as described in Section III.

Source: Collected directly from universities' bylaws, statutes, policy documents, faculty handbooks, human resource departments, or direct communication with staff.

II. Possible Effects of Tenure Clock Stopping Policies

Tenure clock stopping policies have important implications for assistant professors, given the high fertility rates of young economists. About 74 percent of assistant professors in our sample had a child within ten years of earning their PhD, and 47 percent of men and 44 percent of women had a child within five years—before any standard tenure appraisal date. These policies may affect tenure in several ways. First, even if fertility and research behaviors do not change, parents benefit from an additional year of productivity before their tenure review. Given long publication lags, the extra year may also simply allow more time for papers to get published. If tenure decisions are made using an absolute standard and the additional time on

the tenure clock is fully discounted (e.g., appraisers continue to divide by six when thinking about average productivity instead of seven), then these policies will not negatively affect anyone. But in practice, it is unclear how much tenure evaluators discount “stopped” time (Thornton 2009). If reviewers do not fully discount stopped time, the effect of gender-neutral policies can be either negative or positive depending on the magnitudes of the discounting and productivity effects. Regardless of how stopped time is discounted, these policies will benefit fathers more than mothers if the productivity loss associated with having a child is higher for women.⁹ In this case, women can even end up worse off if tenure evaluations are made using relative, rather than absolute, standards and the extra time helps fathers enough to raise the tenure bar.¹⁰ The use of relative standards similarly implies that gender-neutral policies can change the distribution of outcomes for non-parents compared to parents via changes in the tenure bar.

Second, even assuming fertility does not change, extra time on the tenure clock may directly affect observed productivity through publication submission decisions. Tenure in top-50 economics departments is heavily dependent on publications in highly ranked economics journals, but these journals have very low acceptance rates and lengthy review processes. As the time to the tenure decision increases, there is less risk associated with submitting papers to top journals because there is more time to try again (at potentially lower ranked journals with higher acceptance rates) if unsuccessful. The potential to stop the clock may therefore make assistant professors more likely to submit to top journals early in their career. This leads to more highly ranked publications if this strategy pays off at least some of the time. It is worth emphasizing that this strategy is disproportionately attractive to individuals with lower expected productivity losses surrounding birth, because they anticipate having relatively more time to work on papers before submission or to complete difficult revisions. Gender-neutral clock stopping policies may then again benefit men more than women if new mothers use the extra time less productively or less strategically than new fathers. If either the number or quality of papers published increases, then tenure clock stopping policies may in turn also increase tenure standards within departments.

Third, it is important to recognize that tenure clock stopping policies may affect tenure and publication outcomes even for individuals who ultimately do not have a child prior to tenure. For individuals who plan to have children eventually, the knowledge that they could try to do so before tenure in order to extend their clock if needed may be enough in and of itself to change publication strategies early on in their career.

Fourth, tenure clock stopping policies may also change fertility decisions. Both men and women may be more likely to have a child before tenure if they think an extra year on the clock will at least partially compensate for the time costs associated with having a child. The effect might be larger for women if their opportunity

⁹ It is likely that this is the case. Rhoads and Rhoads (2012) show male assistant professors spend much less time in childcare than female assistant professors. Even if parents spend equal time in childcare, productivity losses are still likely larger on average for women than for men because their costs of childbearing are much higher.

¹⁰ Instead of having a fixed tenure bar, some departments may have a fixed number of tenure slots that assistant professors compete for. In this case, a relative increase in male productivity would also decrease the probability that mothers earn tenure. We refer to both of these possibilities as increasing tenure standards.

cost of pre-tenure childbearing is higher. Or, the effect might be larger for men if it is easier for men to adjust the timing of fertility because their costs are low and their expected benefits are large. Having a child before tenure may lower pre-tenure productivity, especially for women.

Finally, to the extent that tenure decisions are based on an absolute standard, female-only and gender-neutral policies should have similar effects for women. If tenure decisions are instead based on a relative standard, female-only policies may be relatively more advantageous for women. This is because the tenure bar is unlikely to rise as a result of female-only policies, as women constitute a small share of assistant professors and their costs of childbearing are high. But again, the effects depend on the extent to which stopped time is discounted by tenure appraisers, which may vary by policy type.

III. Data

We compiled two unique datasets. The first is a complete list of tenure clock stopping policy adoption dates for 49 of the top-50 economics departments, shown in Table 1.¹¹ The second dataset contains academic employment histories and publication records for all faculty members hired as an assistant professor between 1980 and 2005 at any of the top-50 economics departments in our sample. We have complete data on key outcomes for all individuals; there is no sample selection or attrition.

Assistant professors were identified from archived course catalogs obtained from each university. Information about every person's academic job history and publication record was taken from their most recent curriculum vitae (CV), or found via Internet search.¹² We also collected information on the person's gender, PhD, and the year and institution at which they first received tenure. Our main sample includes 1,392 individuals who started their first job as an assistant professor at a top-50 economics department within two years of receiving their PhD, and who published at least two journal articles within eight years of their PhD. This publication restriction reduces noise in the data, as those with zero or one publication would not be serious candidates for tenure in top departments—they typically leave academia quickly, and essentially never get tenure at their first job. We relax these restrictions in Section V.

¹¹ These are the top 50 economics departments with graduate programs as ranked by the 2010 *US News and World Report*. California Institute of Technology is excluded from our sample because they do not have a stand-alone economics department; rather they offer a PhD in social sciences. We were unable to obtain complete faculty information from Brown and the University of Pittsburgh. There was a three-way tie for the fiftieth ranked department, so our final sample includes 49 universities. The data were obtained directly from each institution, so the exact source of the policy details and adoption dates varies. Records were often obtained from archived copies of university bylaws or statutes, human resource departments, or faculty affairs documents.

¹² We located CVs that report academic job histories and publications for more than 93 percent of the assistant professors in our sample. For the remainder, we found the relevant job history information via professional or personal websites, university records, and other publications. We found publication information using the Web of Science. We were able to find information on our main outcomes of interest for every person in the sample. We also used the Web of Science to fill in publication data in recent years when CVs were out of date. We did not track specific job changes once the individual left academia. Nonacademic jobs include working for government organizations, private sector companies, non-governmental organizations, and non-tenure track positions. However, we do track reentry into tenured or tenure-track academic jobs.

To match each assistant professor to the appropriate tenure clock stopping policy, we construct indicators for whether a female-only or gender-neutral policy was in place in the first year of the individual's first assistant professor appointment. We address lags in program take-up in Section IV. The university at which a person is tied to a clock stopping policy is hereafter referred to as the *policy university*.

The main outcome of interest is whether or not the assistant professor was granted tenure at the policy university. Additional outcomes include tenure in the profession, research productivity, time to tenure, and mobility. Research productivity is measured by the number of non-top-5 peer-reviewed journal articles published within three, five, seven, and nine years of receiving a PhD, as well as the number of articles published in top-5 economics journals within these windows. Top-5 publications include regular and short peer-reviewed articles in the *American Economic Review*, *Econometrica*, *Journal of Political Economy*, *Quarterly Journal of Economics*, and the *Review of Economic Studies*.¹³

To examine fertility effects, we surveyed everyone in our sample for whom we could find a valid email address and asked whether or not they have children and years of birth. While the response rate was relatively high (66 percent), there are large differences in response rates on observables and by cohort. These issues are discussed in detail in Section V. We construct indicators for having any children during the first five years after PhD completion and ever having children. We also construct continuous measures for the number of children born (or adopted) during the first five years and the total number of children.

A few data limitations warrant comment. First, we do not observe actual tenure decisions, only outcomes. Some people who would have been granted tenure leave at or before the tenure decision. We observe only actual promotions to tenured positions, which usually occur when individuals are promoted from assistant to associate professor.¹⁴ We thus identify effects from tenure promotions rather than tenure decisions. Second, we do not observe tenure clock extensions. However, as discussed in Section II, these policies affect all assistant professors who expect to have children, regardless of whether or not they stop their clock. Finally, there are relatively few women hired at each university; the primary estimation sample includes 1,150 men and 242 women hired between 1980 and 2005.

IV. Estimation and Identification

Our primary objective is to estimate the effect of tenure clock stopping policies on the probability that assistant professors at top-50 economics departments are granted tenure. We specifically want to allow for the possibility that the effect of

¹³ We exclude publications in the *American Economic Review Papers and Proceedings* issue, as well as comments, replies, and book reviews in all issues of these five journals. *AER Papers and Proceedings* articles are counted in the regular publications counts. Replies and comments are not counted as publications. We also collected coauthoring information for top-5 publications.

¹⁴ Some departments promote to associate without tenure. We distinguish between associates with and without tenure.

clock stopping policies may be different for men and women. Our baseline specification is therefore as follows:

$$\begin{aligned}
 (1) \quad Y_{ugit} = & \beta_1 GN_{ut} + \beta_2 GN_{ut} \times F_{ugit} + \beta_3 GN_{ut} \times E_{ut} + \beta_4 GN_{ut} \times E_{ut} \times F_{ugit} \\
 & + \beta_5 FO_{ut} + \beta_6 FO_{ut} \times F_{ugit} + \beta_7 FO_{ut} \times E_{ut} + \beta_8 FO_{ut} \times E_{ut} \times F_{ugit} \\
 & + \zeta X_{ugit} + \eta Z_{ut} + \rho_{gt} + \psi_{ug} + \varepsilon_{ugit},
 \end{aligned}$$

where Y is the outcome of interest (e.g., tenure at the policy university), and u , g , i , and t indicate policy university, gender, individual, and the year the policy job started, respectively. The variable F is an indicator for female, and GN is an indicator equal to 1 if individual i starts their career at an institution with a gender-neutral tenure clock stopping policy in place and zero otherwise. The indicator FO is defined equivalently if the individual starts their career at a university with a female-only policy in place. The variable E is an indicator for jobs starting in years zero through three after policy adoption.¹⁵ The calendar year in which a policy is adopted is defined as year zero. This allows the effects in the early years of the policies to differ from the effects in later years. This is important for at least two reasons. First, program uptake generally takes time. Second, if some of the effects come through changing tenure standards, the effects should not be immediate because standards take time to evolve, especially if program uptake evolves as well. The vector X includes indicators for being female (F), PhD program rank, and an indicator for having done a postdoc.¹⁶ The vector Z includes time-varying university level controls collected from the Integrated Post-Secondary Education Data System (IPEDS) including the number of undergraduate students, number of graduate students, faculty size, average salary of full professors, average salary of assistant professors, annual revenue, the fraction of the faculty who are female, and the fraction of the faculty who are full professors.¹⁷ Finally, ρ is a vector of gender-specific year fixed effects for the year the policy job started (cohort effects), ψ is a vector of gender-specific university fixed effects, and ε is an error term. The standard errors are clustered at the policy university level.

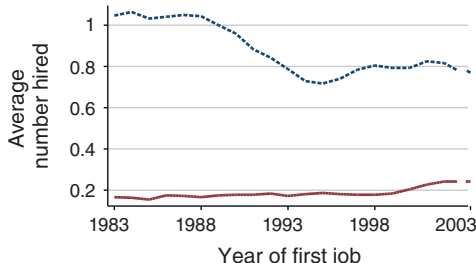
The coefficient β_1 represents the effect of gender-neutral tenure clock stopping policies for men in universities with policies in place for at least four years, and β_2 is the additional effect for women. The coefficients β_3 and β_4 give the differential effects for men and women, respectively, who start their job within three years of the institution's policy adoption compared to their counterparts who start after the policy has been in place for at least four years. Finally, β_5 through β_8 represent analogous effects of female-only policies. A plausible causal interpretation of these coefficients rests heavily on two related issues. First, we must satisfy the common trends assumption by adequately controlling for underlying gender-specific trends in the outcome. Figure 2 shows how the gender composition of assistant professors

¹⁵The results are not sensitive to small changes in the years assigned to E .

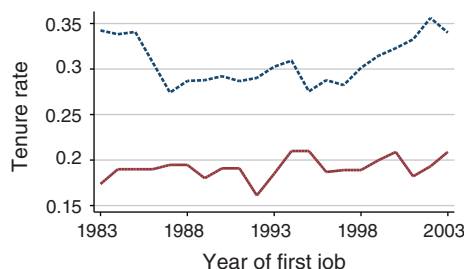
¹⁶We create five tiers of PhD programs based on job placements in the top-50 departments within our sample.

¹⁷These controls reduce concerns that policy adoption is correlated with other university-wide changes that affect hiring or retention. The results are robust to their omission and Section VI shows that they are uncorrelated with policy adoption.

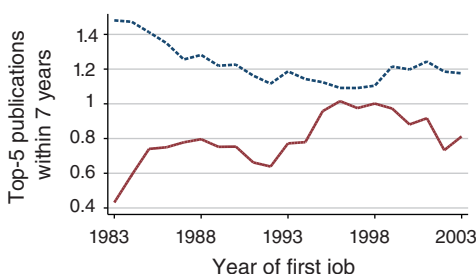
Panel A. Average number of assistant professors hired



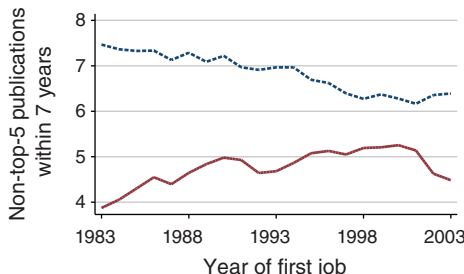
Panel B. Tenure rate



Panel C. Average number of top-5 publications within 7 years



Panel D. Average number of non-top-5 publications within 7 years



--- Men — Women

FIGURE 2. TRENDS IN TOP-50 ECONOMICS DEPARTMENTS

Notes: Sample includes individuals whose first job was an assistant professor at a top-50 university within two years of receiving a PhD and who publish at least two papers within eight years of receiving a PhD. Figures show seven-year moving averages for smoothness. Publications are counted through the seventh year after receiving a PhD.

has changed during our sample, as well as the differential trends in tenure and publication rates. These differences highlight why our main specification flexibly controls for differences between men and women over time. Gender specific university fixed effects similarly allow for the possibility that male and female experiences and outcomes differ within departments. That being said, we discuss a variety of alternative trend options in Section V, and show that our results are generally robust to such strategies. Second, conditional on the available control variables, the timing of policy adoption must be as good as random. As clock stopping policies are adopted at the university level, it is reasonable to assume that they are not driven by hiring and tenure trends in a single department. Nevertheless, we explore this issue in Section VI.

V. Results

Table 2 presents our main results for equation (1). For ease of interpretation, we present total effects for men and women hired in years zero through three after policy adoption ($\beta_1 + \beta_3$ for men and $\beta_1 + \beta_2 + \beta_3 + \beta_4$ for women) in panel A and total effects for those hired in years four or later (β_1 for men and $\beta_1 + \beta_2$ for women) in panel B. Panel B shows that men whose first job was at a top-50

TABLE 2—THE EFFECTS OF CLOCK STOPPING POLICIES ON THE PROBABILITY OF TENURE AT THE POLICY UNIVERSITY

	Total effects (1)	Male-female (2)	GNCS-FOCS (3)
<i>Panel A. Policy effects years 0–3</i>			
Men FOCS	–0.008 (0.067)	–0.181 (0.140)	
Women FOCS	0.172 (0.140)		
Men GNCS	0.051 (0.079)	0.068 (0.145)	0.060 (0.098)
Women GNCS	–0.017 (0.107)		–0.189 (0.164)
<i>Panel B. Policy effects years 4+</i>			
Men FOCS	0.002 (0.075)	–0.047 (0.128)	
Women FOCS	0.049 (0.101)		
Men GNCS	0.176 (0.083)	0.370 (0.146)	0.173 (0.085)
Women GNCS	–0.194 (0.106)		–0.244 (0.104)
Sample size	1,392		

Notes: The sample includes 49 universities. Standard errors are clustered at the policy university-level. The outcome is an indicator variable equal to one if the individual becomes a tenured associate professor at the policy university and zero otherwise. GNCS indicates a gender-neutral tenure clock stopping policy and FOCS indicates a female-only policy. The first column shows the total effect for each group, the second column shows the difference in the male and female coefficients for each policy type, and the third column shows the difference between the effect of the GNCS and FOCS policies for men and women. Additional controls: All models also include gender-specific indicators for the year the policy job started, gender-specific university indicators, a female indicator, PhD rank, an indicator for having done a postdoc, and time-varying university characteristics (number of undergraduates, number of graduate students, faculty size, average salary of full professors, average salary of assistant professors, annual revenue, fraction of faculty who are female, and fraction of faculty who are full professors).

university with a gender-neutral tenure clock stopping policy in place for more than three years have a 17.6 percentage point tenure rate advantage over men at the same university prior to the implementation of any policy (β_1). In contrast, the adoption of a gender-neutral policy negatively affects women in our sample. Column 2 shows that the gender gap in tenure attainment at the typical policy university increased by 37 percentage points after implementation of the policy ($-\beta_2$). Not only are women less likely to get tenure compared to men, but they are also 19.4 percentage points less likely to get tenure relative to women at the same university with no clock stopping policy ($\beta_1 + \beta_2$ in column 1). Both the male-female difference-in-difference and the absolute male and female differences are statistically significant at conventional levels.

Unlike gender-neutral policies, female-only clock stopping policies appear to have small positive effects for women, but the estimates are very imprecise. Further, column 3 shows that the gender-neutral effect relative to the female-only effect is positive for men and negative for women ($(\beta_1 - \beta_5)$ and $(\beta_1 + \beta_3) - (\beta_5 + \beta_6)$, respectively). But given the small number of schools that only adopt a female-only

policy and the fairly short time span between the adoption of female-only and gender-neutral policies at universities that switch, the point estimates should be interpreted with caution. For this reason, most of the discussion focuses on gender-neutral policies, but the female-only effects are reported in all tables.

Panel A of Table 2 shows that gender-neutral clock stopping policies do not have significant effects on tenure rates for those hired in years zero through three after implementation. While the point estimates are again positive for men and negative for women, they are much smaller and statistically insignificant. This suggests that take-up may be low in the early years of the policy. Or, if the gender-neutral policies cause tenure standards to rise, it might take time for this process to evolve.

While the gender-neutral policy effects are large, it is important to remember that the vast majority of assistant professors are affected by these policies, not just those induced to have a child pre-tenure. As shown later in this section, 74 percent of our sample have children within ten years of completing their PhD. This means that many individuals may be directly impacted even if their fertility does not change. It is also important to emphasize that these policies affect individuals who do not have a child before tenure; just knowing that they could do so to get extra time on the clock might be enough to change their productivity or publication strategies. Also note that tenure rates are falling by approximately 10 percentage points for men and are rising by about 10 percentage points for women at institutions that have not yet adopted a clock stopping policy over the sample period. As we find that tenure rates for women fall by about 17 percentage points after a gender-neutral policy is adopted, this implies that much of the effect comes from a lack of increase in female tenure rates in these departments, rather than a large absolute decline in tenure rates. The increase in tenure rates for men relative to no policy is similarly partially driven by the absence of this declining trend once a gender-neutral policy is implemented.

In order to interpret the above estimates as causal impacts of clock stopping policies, there cannot be differential trends in tenure rates between universities that do and do not adopt them. To test for pre-trends, Table 3 shows the results from a modified event study specification, in which we additionally include three sets of indicators for periods prior to the adoption of any clock stopping policy.¹⁸ For universities that switch from a female-only to gender-neutral policy, the pre-periods are defined as years until the first policy adoption, as either type of policy may theoretically affect future tenure outcomes. We find no evidence of significant pre-trends in the years leading up to first policy implementation, reducing concerns about the validity of the research design. Given that both Tables 2 and 3 show no effects of either type of policy in the pre-periods or the early years of the policy, we continue to estimate equation (1) but only report the effects for individuals hired more than three years after policy implementation in subsequent tables.

Table 4 shows that the main results are robust to alternate specifications and samples. To facilitate comparisons, column 1 reports the estimates from Table 2.

¹⁸ Given the female sample size, we cannot estimate separate effects for each cohort. Instead, we create bins of 1–6, 7–12, and 13–18 years before policy implementation. Reducing the number of years in each window produces noisier estimates for women given the small sample size in each cell, but results in no indication of robust systematic pre-trends in tenure rates.

TABLE 3—CLOCK STOPPING POLICIES AND TENURE AT THE POLICY UNIVERSITY: ALLOWING FOR PRE-PERIOD EFFECTS

	Relative to the period prior to the first policy		
	Total effects (1)	Male-female (2)	GNCS-FOCS (3)
<i>Panel A. Pre first policy year effects</i>			
Men 13–18 years before first policy	−0.081 (0.066)	−0.126 (0.131)	
Women 13–18 years before first policy	0.045 (0.125)		
Men 7–12 years before first policy	0.017 (0.042)	−0.007 (0.120)	
Women 7–12 years before first policy	0.023 (0.109)		
Men 1–6 years before first policy	—	—	
Women 1–6 years before first policy	—		
<i>Panel B. Policy effects years 0–3</i>			
Men FOCS	0.015 (0.068)	−0.150 (0.137)	
Women FOCS	0.164 (0.142)		
Men GNCS	0.086 (0.076)	0.117 (0.136)	0.071 (0.096)
Women GNCS	−0.031 (0.098)		−0.195 (0.170)
<i>Panel C. Policy effects years 4+</i>			
Men FOCS	0.034 (0.069)	0.001 (0.111)	
Women FOCS	0.033 (0.090)		
Men GNCS	0.231 (0.073)	0.462 (0.141)	0.197 (0.081)
Women GNCS	−0.231 (0.109)		−0.264 (0.112)
Sample size	1,392		

Notes: The sample includes 49 universities. Standard errors are clustered at the policy university level. The outcome is an indicator variable equal to one if the individual becomes a tenured associate professor at the policy university and zero otherwise. GNCS indicates a gender-neutral tenure clock stopping policy and FOCS indicates a female-only policy. The first column shows the total effect for each group, the second column shows the difference in the male and female coefficients for each policy type, and the third column shows the difference between the effect of the GNCS and FOCS policies for men and women. All models also include the additional controls listed in the Table 2 notes.

Column 2 removes the time-varying university characteristics, postdoc indicator, and PhD rank. Column 3 expands the sample to include individuals with less than two publications or who started their first job more than two years after earning their PhD. Column 4 adds department tier-by-gender cohort effects.¹⁹ Column 5

¹⁹The top tier includes Harvard, MIT, Stanford, University of Chicago, Princeton, University of California Berkeley, Yale, Northwestern, University of California-San Diego, and University of Rochester. The second tier includes all other universities.

TABLE 4—CLOCK STOPPING POLICIES AND TENURE AT THE POLICY UNIVERSITY: ALTERNATIVE SAMPLES AND SPECIFICATIONS

	Baseline (1)	Excludes individual and university controls (2)	Removes sample restrictions (3)	Adds tier-gender specific trends (top 10) (4)	Removes gender interactions (5)	Additional individual controls (6)
<i>Panel A. Policy effects years 4+</i>						
Men FOCS	0.002 (0.075)	−0.005 (0.072)	−0.024 (0.074)	−0.039 (0.078)	−0.016 (0.064)	−0.040 (0.064)
Women FOCS	0.049 (0.101)	0.045 (0.098)	0.075 (0.095)	0.120 (0.117)	0.120 (0.104)	0.127 (0.075)
Men GNCS	0.176 (0.083)	0.158 (0.075)	0.140 (0.084)	0.158 (0.086)	0.161 (0.067)	0.128 (0.082)
Women GNCS	−0.194 (0.106)	−0.188 (0.104)	−0.156 (0.095)	−0.141 (0.138)	−0.045 (0.084)	−0.161 (0.103)
<i>Panel B. Male-female difference</i>						
FOCS	−0.047 (0.128)	−0.051 (0.118)	−0.099 (0.111)	−0.159 (0.145)	−0.136 (0.092)	−0.168 (0.092)
GNCS	0.370 (0.146)	0.346 (0.134)	0.296 (0.138)	0.300 (0.174)	0.206 (0.069)	0.289 (0.136)
Sample size	1,392	1,392	1,591	1,392	1,392	1,392

Notes: The sample includes 49 universities. Standard errors are clustered at the policy university level. The outcome is an indicator variable equal to 1 if the individual becomes a tenured associate professor at the policy university and 0 otherwise. GNCS indicates a gender-neutral tenure clock stopping policy and FOCS indicates a female-only policy. All models also include the additional controls listed in the Table 2 notes, except when otherwise specified. Policy effects for years 0–3 are included in all models. Column 6 includes covariates that could explain the effects on tenure: number of non-top-5 and top-5 publications within seven years of receiving a PhD, average number of coauthors on top-5 publications, children born within five years of PhD completion, and an indicator for missing fertility information.

removes the university and cohort gender interactions. Column 6 adds covariates that could explain the effects on tenure (number of non-top-5 and top-5 publications within seven years of receiving a PhD, average number of coauthors on top-5 publications, and children born within five years of PhD completion).²⁰ While the estimates are generally consistent across specifications, it is important to highlight the smaller female estimate in column 5. As men constitute more than 80 percent of the sample, in the absence of gender-specific time and university fixed effects, the trends and intercepts are essentially male. The lower female coefficient highlights the importance of these gender-specific fixed effects. It is also important to note that the somewhat smaller and noisier point estimates in column 6 are consistent with upcoming results showing that some of these variables are important pathways through which gender-neutral policies ultimately affect tenure rates.

While the adoption of gender-neutral tenure clock stopping policies help men and hurt women, these policies do not cause women to leave the profession altogether. Column 1 of Table 5 clearly shows that these policies have little impact on the probability of earning tenure in the profession (at any college or university) for

²⁰We also include a missing indicator for individuals who did not respond to our fertility survey.

TABLE 5—CLOCK STOPPING POLICIES, TENURE SOMEWHERE, TIME TO TENURE, AND JOB CHURNING

	Main sample				Tenure somewhere		No tenure at policy university	
	Tenure somewhere (1)	Pre-tenure years at policy university (2)	Leave early (3)	Leave late (4)	Years to tenure (5)	Number of jobs to associate (6)	Move down (7)	Move up (8)
<i>Panel A. Policy effects years 4+</i>								
Men FOCS	0.026 (0.059)	−0.279 (0.342)	0.090 (0.080)	0.002 (0.028)	−0.308 (0.411)	−0.067 (0.120)	0.006 (0.075)	−0.016 (0.041)
Women FOCS	0.063 (0.125)	−0.116 (0.591)	0.064 (0.161)	−0.011 (0.103)	1.104 (0.886)	0.270 (0.229)	−0.139 (0.142)	0.062 (0.108)
Men GNCS	−0.005 (0.044)	−0.263 (0.287)	0.103 (0.070)	0.098 (0.040)	−0.907 (0.417)	−0.205 (0.177)	0.007 (0.073)	0.048 (0.043)
Women GNCS	0.074 (0.202)	−0.222 (0.927)	0.095 (0.201)	0.097 (0.138)	1.805 (1.120)	0.518 (0.248)	−0.251 (0.114)	0.136 (0.092)
<i>Panel B. Male-female difference</i>								
FOCS	−0.037 (0.143)	−0.163 (0.712)	0.026 (0.157)	0.012 (0.106)	−1.412 (1.051)	−0.337 (0.279)	0.145 (0.163)	−0.078 (0.134)
GNCS	−0.078 (0.206)	−0.040 (1.033)	0.008 (0.222)	0.001 (0.140)	−2.713 (1.303)	−0.722 (0.338)	0.258 (0.142)	−0.089 (0.102)
Sample size	1,373	1,392	1,392	1,351	1,121	1,140	983	983

Notes: The sample includes 49 universities. Standard errors are clustered at the policy university level. GNCS indicates a gender-neutral tenure clock stopping policy and FOCS indicates a female-only policy. All models also include the additional controls listed in the Table 2 notes. Policy effects for years 0–3 are included in all models.

either men or women.²¹ The remaining columns further explore the impact of clock stopping policies on measures of time to tenure and mobility.

Although it is tempting to think that the availability of clock stopping will lengthen the average time spent as an assistant professor, the landscape is much more complicated. The introduction of clock stopping might change the probability that men and women leave prior to the tenure decision. It might also lengthen the time spent at the policy university, but then shorten the length of a new clock if they move to a new department after being denied tenure (or in anticipation of a tenure denial). Column 2 of Table 5 shows that there is no significant effect of these policies on the average number of years spent as an assistant professor at the policy university, but this masks substantial heterogeneity. Columns 3 and 4 show that men and women are both about 10 percentage points more likely to leave before the tenure decision and 10 percentage points more likely to stay longer than the length of the standard tenure clock, though only the male effect of staying longer is statistically significant.²² Next, column 5 shows that gender-neutral clock stopping policies decrease average time to tenure in the profession by 0.9 years for men. This is not surprising since they are more likely to earn tenure at the policy university. In contrast,

²¹ The sample has 19 fewer people because they are still untenured assistant professors in their second or third job as of 2017. Although they did not get tenure at the policy university, we do not yet know if they will eventually get tenure somewhere else.

²² We allow for differences in the length of the clock across departments by using the modal number of years as an assistant professor in the absence of a clock stopping policy to define the length of the tenure clock. Leaving late is defined as being an assistant professor at the policy university more than this number of years. Leaving early is defined as leaving before year five.

TABLE 6—CLOCK STOPPING POLICIES AND PUBLICATION OUTCOMES

	Top publications				Non-top publications			
	By year 3 (1)	By year 5 (2)	By year 7 (3)	By year 9 (4)	By year 3 (5)	By year 5 (6)	By year 7 (7)	By year 9 (8)
<i>Panel A. Policy effects years 4+</i>								
Men FOCS	0.089 (0.071)	0.278 (0.159)	0.318 (0.218)	0.308 (0.266)	0.452 (0.284)	0.908 (0.470)	0.694 (0.697)	0.699 (0.884)
Women FOCS	−0.186 (0.182)	−0.297 (0.317)	−0.645 (0.441)	−0.481 (0.554)	0.018 (0.571)	−0.271 (0.823)	−0.875 (0.885)	−1.414 (1.212)
Men GNCS	0.003 (0.092)	0.355 (0.184)	0.561 (0.241)	0.663 (0.273)	0.075 (0.345)	0.464 (0.495)	0.057 (0.697)	−0.048 (0.923)
Women GNCS	−0.265 (0.148)	−0.069 (0.256)	−0.115 (0.403)	−0.158 (0.480)	−0.628 (0.467)	−0.955 (0.682)	−0.779 (1.054)	−1.192 (1.329)
<i>Panel B. Male-female difference</i>								
FOCS	0.275 (0.166)	0.576 (0.358)	0.963 (0.506)	0.789 (0.604)	0.434 (0.640)	1.179 (1.000)	1.570 (1.074)	2.113 (1.437)
GNCS	0.268 (0.166)	0.424 (0.260)	0.677 (0.427)	0.821 (0.537)	0.703 (0.531)	1.419 (0.829)	0.836 (1.162)	1.144 (1.420)
Sample size	1,392	1,392	1,392	1,392	1,392	1,392	1,392	1,392

Notes: The sample includes 49 universities. Standard errors are clustered at the policy university level. GNCS indicates a gender-neutral tenure clock stopping policy and FOCS indicates a female-only policy. All models also include the additional controls listed in the Table 2 notes. Policy effects for years 0–3 are included in all models.

women take about 1.8 years longer (imprecisely measured) to get tenure. Column 6 shows that gender-neutral policies increase the gender gap in the number of jobs held before tenure. Columns 7 and 8 show that conditional on not getting tenure at the policy university, men are no more likely to move to higher or lower ranked departments, while women are 25 percentage points less likely to move down.²³ In other words, some of the women who are denied tenure because of gender-neutral policies move to similarly ranked departments. This is consistent with rising tenure standards pushing out some marginal women.

As tenure at top-50 departments is highly dependent on publications, publishing may be a key pathway through which gender-neutral clock stopping policies operate. Table 6 shows the effects of clock stopping policies on the cumulative number of top-5 and non-top-5 publications by the third, fifth, seventh, and ninth years since PhD completion. We report results at several points in time because there is no obvious single year at which to evaluate publications. By year five, men exposed to a gender-neutral policy have 0.36 more top-5 publications than men at the same university without a policy. This grows to 0.56 by year seven.²⁴ In contrast, there is no increase in the number of top-5 publications for women. Columns 5–8 show analogous estimates for non-top-5 publications. While the point estimates suggest

²³ An important caveat is that department rankings are inherently noisy and may change over time. We define moves as up, down, or lateral based on tiers consisting of similarly ranked departments. Of movers, about 86 percent of our sample move down, 6 percent move up, and 8 percent stay in the same tier.

²⁴ The effect levels out at about 0.66 and remains constant in subsequent years.

that women also publish about one fewer non-top-5 papers after a gender-neutral policy is implemented, the estimates are imprecise.²⁵

It is also worth noting that the male publication results are consistent with the results on time to tenure and tenure rates. As discussed in Section II, gender-neutral clock stopping policies might induce some men to choose a riskier publication strategy—aim high. Those with early publication success may not ultimately need to stop their clock. For others, success might take longer and induce clock stopping. In still other cases, some men might leave earlier when it becomes clear that the riskier strategy will not pay off.

Fertility is another potentially important pathway through which clock stopping policies may affect tenure outcomes. As previously discussed, it is important to realize that a large fraction of the sample is potentially impacted by gender-neutral clock stopping policies when thinking about the magnitudes of the effects. In our sample, 86 (79) percent of male (female) economists eventually have children and 74 (72) percent do so within ten years of PhD completion. Not only are a large fraction of people eligible to stop their clock without a change in fertility, but there is also considerable room for assistant professors to adjust the timing of their fertility and have a child pre-tenure.

Unfortunately, our fertility sample suffers from severe non-random reporting. The response rate is significantly higher for later cohorts, those who stay in academia, earn tenure at the policy school, and start their career in a top-10 department. For example, those who never earn tenure are 32 percentage points less likely to respond than those who earn tenure at the policy university.²⁶ Given the selection on observables, and the fact that it is unclear whether non-respondents are differentially likely to have children, we cannot estimate causal effects for fertility. Additionally, substantial variation in response rates across universities and over cohorts makes it even more difficult to estimate and interpret the fixed effects regressions estimated above.²⁷ However, Table 7 provides descriptive evidence that universities with clock stopping policies in place have higher pre-tenure fertility rates than those without. The share of men who have at least one child in the first five years after PhD completion is 11 percentage points higher when a gender-neutral policy is available, and they have 0.18 more children on average during this same period. Both of these differences in means are statistically significant. Women exposed to a gender-neutral policy are 12 percentage points more likely to have a child pre-tenure and have 0.23 more children during the first five years of their career, but only the latter difference is statistically significant. We detect no significant differences in pre-tenure fertility for women exposed to a female-only policy compared to no policy, but the magnitudes are quite similar to those of the gender-neutral policies. While only descriptive, these results are consistent with the idea that some men and women change the timing of fertility when a clock stopping option exists. There are no significant differences

²⁵The point estimates of the effects of female-only policies on men are positive and occasionally marginally significant for top and non-top publications. These estimates for men are considerably smaller and never significant if the universities that switch from female-only to gender-neutral policies are excluded. On the other hand, the effects of gender-neutral policies are unchanged when these universities are excluded. The main results on tenure are also very similar. These results are available upon request.

²⁶These results are available upon request.

²⁷For example, we do not observe both a male and female response on either side of the gender-neutral clock stopping policy in 20 of the 34 universities with gender-neutral policy changes.

TABLE 7—CLOCK STOPPING POLICIES AND FERTILITY

	Men		Women			
			Have not adopted FOCS		Have not adopted GNCS	
	No GNCS mean (1)	GNCS mean (2)	No GNCS mean (3)	GNCS mean (4)	No FOCS mean (5)	FOCS mean (6)
Any birth	0.454 (0.498)	0.563 (0.498)	0.396 (0.492)	0.516 (0.508)	0.396 (0.492)	0.500 (0.513)
Number of births	0.632 (0.776)	0.813 (0.822)	0.451 (0.601)	0.677 (0.748)	0.451 (0.601)	0.650 (0.745)
Ever have children	0.859 (0.348)	0.848 (0.360)	0.736 (0.443)	0.871 (0.341)	0.736 (0.443)	0.750 (0.444)
Total number of children	1.876 (1.085)	1.732 (0.986)	1.429 (1.056)	1.581 (0.807)	1.429 (1.056)	1.400 (0.995)
Sample size	581	112	91	31	91	20

Notes: Sample includes survey respondents listed in the column heading categories. Standard deviations are in parentheses. As we have shown that it takes time for policies to take hold, observations for the relevant 0–3 year periods after adoption are excluded from sample. Columns 1, 3, and 5 are means for individuals not covered by GNCS or FOCS as listed, and columns 2, 4, and 6 are the means for those covered by GNCS or FOCS.

in the number of children ever born, suggesting that these policies shift timing rather than increase fertility. To the extent that the cost of having a child is higher for women, these effects are consistent with the tenure and publication results.

Taken as a whole, the results in this section show that gender-neutral clock stopping policies increase male tenure rates and decrease female tenure rates. On the other hand, we find no evidence that female-only policies have significant negative effects on female tenure rates. And while imprecisely measured, the effects of female-only and gender-neutral policies on other outcomes such as publications and fertility are similar for women. Together, these results suggest that the increase in male publishing in top journals resulting from gender-neutral clock stopping policies is likely driving the increase in the gender gap in tenure rates.

VI. Threats to Identification

Interpreting the above results as causal requires the assumption that policy adoption was exogenous and that clock stopping policies did not change hiring decisions. Although these assumptions are not directly testable, Tables 8 and 9 provide evidence that they are reasonable. Column 1 of Table 8 shows there is no change in the gender mix of assistant professors after the adoption of gender-neutral or female-only policies. The remaining columns similarly show that these policies are not correlated with PhD rank, having done a postdoc, or publishing while in graduate school.

Table 9 comes at endogeneity from the other direction by asking whether it is possible to predict policy adoption using either levels or changes in institutional characteristics. Each cell in Table 9 is from a separate regression. Each cell in column 1 (2) reports the point estimate from a cross-sectional OLS regression

TABLE 8—CHANGES IN HIRING AFTER CLOCK STOPPING POLICY ADOPTION

	Female (1)	PhD rank (2)	Post-doc (3)	Number of top publications by year 1 (4)	Number of non-top publications by year 1 (5)
FOCS	0.006 (0.033)				
GNCS	0.004 (0.042)				
Men FOCS		0.182 (0.197)	−0.010 (0.025)	0.006 (0.024)	0.128 (0.126)
Women FOCS		0.275 (0.343)	−0.034 (0.044)	−0.021 (0.037)	−0.106 (0.277)
Men GNCS		0.221 (0.180)	0.005 (0.026)	0.038 (0.037)	0.000 (0.169)
Women GNCS		0.603 (0.363)	0.041 (0.061)	−0.001 (0.018)	−0.142 (0.210)
Sample size	1,392	1,392	1,392	1,392	1,392

Notes: The sample includes 49 universities. Standard errors are clustered at the policy university-level. Outcomes are listed in each column heading. GNCS indicates a gender-neutral tenure clock stopping policy and FOCS indicates a female-only policy. The column 1 specification also includes indicators for the year the policy job started and the policy university. Columns 2–5 include gender-specific indicators for the year the policy job started, gender-specific university indicators, and a female indicator.

with an indicator for adopting a female-only (gender-neutral) tenure clock stopping policy by 2005 on a single institutional characteristic (the average level from 1980–1989).²⁸ For example, the first cell in column 1 reports that a 1,000 student increase in the undergraduate population is associated with a statistically insignificant 0.6 percentage point decrease in the probability that an institution ever adopts a gender-neutral tenure clock stopping policy. The remaining cells in columns 1 and 2 similarly show that the remaining time-varying university controls do not predict policy adoption. Columns 3 and 4 examine whether changes (trends) in university characteristics predict future policy adoption. More specifically, we show that changes in university characteristics from the early 1980s to the late 1980s do not predict gender-neutral or female-only policy adoption.²⁹ Taken as a whole, Table 9 provides no evidence that levels or trends in institutional characteristics predict tenure clock stopping policy adoption.

VII. Discussion

By combining two original datasets—one on assistant professors hired at the top-50 economics departments from 1980–2005 and the other on tenure clock

²⁸This pre-dates gender-neutral policy adoption for all but one university (Northwestern), which is excluded from the analysis. Princeton and Stanford are excluded from the female-only analysis, as they adopt policies before the beginning of our sample period. The University of California campuses are collapsed into a single average because this was a system choice.

²⁹To mitigate non-reporting issues (data for some institutional measures are missing in some years), the early 1980s is defined as the average from 1980–1984 and the late 1980s is defined as the average from 1985–1989.

TABLE 9—PREDICTING CLOCK STOPPING POLICY ADOPTION

	Does level from 1980–1989 predict FOCS adoption by 2005 (1)	Does level from 1980–1989 predict GNCS adoption by 2005 (2)	Do changes from 1980–1984 to 1985–1989 predict FOCS policy adoption by 2005 (3)	Do changes from 1980–1984 to 1985–1989 predict GNCS policy adoption by 2005 (4)
Number of undergraduate students (in 1,000s)	–0.006 (0.005)	0.002 (0.006)	0.005 (0.060)	0.033 (0.068)
Number of graduate students (in 1,000s)	0.004 (0.022)	–0.013 (0.024)	0.025 (0.082)	0.155 (0.091)
Faculty size (in 100s)	–0.008 (0.013)	0.003 (0.015)	0.024 (0.074)	0.059 (0.085)
Average salary for professors (in 1,000s)	0.020 (0.012)	0.004 (0.013)	0.020 (0.020)	0.020 (0.022)
Average salary for assistant professors (in 1,000s)	0.032 (0.027)	0.011 (0.031)	0.024 (0.035)	0.050 (0.039)
Annual revenue (in 10,000,000s)	0.000 (0.035)	0.043 (0.038)	–0.073 (0.070)	0.106 (0.075)
Fraction of faculty female	–1.501 (1.615)	–1.403 (1.779)	–0.973 (3.994)	3.140 (4.503)
Fraction of faculty full professors	0.643 (0.679)	–0.582 (0.753)	2.347 (1.513)	–2.293 (1.752)
Private university	0.081 (0.132)	–0.156 (0.147)	—	—
Sample size	41	43	82	86

Notes: The sample includes universities that do not have a gender-neutral clock stopping policy at the beginning of the sample period, and the University of California campuses are collapsed into a single average because the policy was a system decision. The gender-neutral policy samples (columns 2 and 4) therefore includes 43 institutions. Columns 1 and 3 additionally exclude universities that implemented a female-only policy before 1980 and include 41 institutions. Each cell is a separate regression. Columns 1 and 2 are cross-sections with an indicator for ever adopting the policy listed in the column heading regressed on the average value for the specified explanatory variable as well as a constant. Columns 3 and 4 are two period panels that includes the listed variable, a period indicator to capture trends, and institution fixed effects.

stopping policies at these universities—we study the impact of gender-neutral tenure clock stopping policies on tenure rates for men and women. While the objective of these policies might be to increase family friendliness and level the playing field for women, our results show that, at least for economics, they accomplished the opposite. Once established, gender-neutral clock stopping policies decrease female tenure rates at the policy university by 19 percentage points while increasing male tenure rates by 17 percentage points. We further show that the primary mechanism driving the tenure results appears to be that men publish more in top-5 journals after the policies are implemented, but women do not. Because women do not similarly increase their productivity, fewer are granted tenure in their first job. However, we find no evidence that gender-neutral clock stopping policies reduce the fraction of women who eventually earn tenure in the profession. We also find no consistent evidence that women are either hurt or helped by female-only policies.

These results imply that gender-neutral tenure clock stopping policies do not adequately account for the true gender-specific productivity losses associated with having children. As a result, gender-neutral policies actually increase the gender gap

in economics at research-intensive universities. Even though women are not more likely to leave the profession, these policies appear to increase job churning and the length of time it takes women to get tenure. However, it is important to highlight that these policies could have different effects in disciplines with different propensities to do postdocs, emphasis on publishing books and obtaining grants, and length of publication lags. As such, more work is needed to assess the impact of tenure clock stopping policies across disciplines.

Economics professors are not the only high-skill professionals that face rigid and important promotion decisions early in their careers. Other academics, lawyers, financial professionals, and some types of doctors are also likely to be promoted based on early measures of success. There is evidence of family gaps in each of these professions, especially among top earners, which suggests a need for more family-friendly policies. In theory, gender-neutral policies that attempt to level the playing field by adjusting measures of productivity to account for early childrearing sound promising. However, at least in economics, such policies have unintended consequences that actually hurt women. It therefore seems likely that these types of policies may have unintended consequences in other high skill occupations as well.

REFERENCES

- American Association of University Professors.** 2001. *Statement of Principles on Family Responsibilities and Academic Work*. Washington, DC: American Association of University Professors.
- American Association of University Professors.** 2014. *Recommended Institutional Regulations on Academic Freedom and Tenure*. Washington, DC: American Association of University Professors.
- Antecol, Heather, Kelly Bedard, and Jenna Stearns.** 2018. "Equal but Inequitable: Who Benefits from Gender-Neutral Tenure Clock Stopping Policies? Dataset." *American Economic Review*. <https://doi.org/10.1257.aer.20160613>.
- Blau, Francine D., and Lawrence M. Kahn.** 2013. "Female Labor Supply: Why Is the United States Falling Behind?" *American Economic Review* 103 (3): 251–56.
- Dahl, Gordon B., Katrine V. Løken, and Magne Mogstad.** 2014. "Peer Effects in Program Participation." *American Economic Review* 104 (7): 2049–74.
- Demougin, Dominique, and Aloysius Siow.** 1994. "Careers in Ongoing Hierarchies." *American Economic Review* 84 (5): 1261–77.
- Ekberg, John, Rickard Eriksson, and Guido Friebel.** 2013. "Parental Leave—A Policy Evaluation of the Swedish "Daddy-Month" Reform." *Journal of Public Economics* 97: 131–43.
- Finkel, Susan Kolker, Steven Oslwang, and Nian She.** 1994. "Childbirth, Tenure, and Promotion for Women Faculty." *Review of Higher Education* 17 (3): 259–70.
- Goldin, Claudia.** 2014. "A Grand Gender Convergence: Its Last Chapter." *American Economic Review* 104 (4): 1091–119.
- Goldin, Claudia, and Lawrence F. Katz.** 2011. "The Cost of Workplace Flexibility for High-Powered Professionals." *Annals of the American Academy of Political and Social Science* 638: 45–67.
- Goldin, Claudia, and Lawrence F. Katz.** 2012. "The Most Egalitarian of All Professions: Pharmacy and the Evolution of a Family-Friendly Occupation." NBER Working Paper 18410.
- Lazear, Edward P., and Sherwin Rosen.** 1990. "Male-Female Wage Differentials in Job Ladders." *Journal of Labor Economics* 8 (1): S106–23.
- Lundberg, Shelly, and Elaina Rose.** 2000. "Parenthood and the Earnings of Married Men and Women." *Labour Economics* 7 (6): 689–710.
- Manchester, Colleen Flaherty, Lisa M. Leslie, and Amit Kramer.** 2013. "Is the Clock Still Ticking? An Evaluation of the Consequences of Stopping the Tenure Clock." *Industrial and Labor Relations Review* 66 (1): 3–31.
- Mason, Mary Ann, and Marc Goulden.** 2002. "Do Babies Matter? The Effect of Family Formation on the Lifelong Career of Academic Men and Women." *Academe* 88 (6): 21–27.
- Mason, Mary Ann, Angelica Stacy, Marc Goulden, Carol Hoffman, and Karie Frasch.** 2005. *University of California Faculty Family Friendly Edge: An Initiative for Tenure-Track Faculty at the University of California*. Oakland, CA: University of California.

- Misra, Joya, Michelle Budig, and Irene Böckmann. 2010. "Work-Family Policies and the Effects of Children on Women's Employment Hours and Wages." *Community, Work and Family* 14 (2): 139–57.
- O'Flaherty, Brendan, and Aloysius Siow. 1995. "Up-or-Out Rules in the Market for Lawyers." *Journal of Labor Economics* 13 (4): 709–35.
- Rehel, Erin, and Emily Baxter. 2015. "Men, Fathers, and Work-Family Balance." Center for American Progress. Washington, DC: Center for American Progress.
- Rhoads, Steven E., and Christopher H. Rhoads. 2012. "Gender Roles and Infant/Toddler Care: Male and Female Professors on the Tenure Track." *Evolutionary Behavioral Sciences* 6 (1): 13–31.
- Rosen, Sherwin. 1990. "Contracts and the Market for Executives." NBER Working Paper 3542.
- Rossin-Slater, Maya, Christopher J. Ruhm, and Jane Waldfogel. 2013. "The Effects of California's Paid Family Leave Program on Mothers' Leave-Taking and Subsequent Labor Market Outcomes." *Journal of Policy Analysis and Management* 32 (2): 224–45.
- Ruggles, Steven, Katie Genadek, Ronald Goeken, Josiah Grover, and Matthew Sobek. 2015. "Integrated Public Use Microdata Series: Version 7.0." Minneapolis: University of Minnesota.
- Ruhm, Christopher J. 1998. "The Economic Consequences of Parental Leave Mandates: Lessons from Europe." *Quarterly Journal of Economics* 113 (1): 285–317.
- Schober, Pia S. 2014. "Parental Leave and Domestic Work of Mothers and Fathers: A Longitudinal Study of Two Reforms in West Germany." *Journal of Social Policy* 43 (2): 351–72.
- Stearns, Jenna. 2018. "The Long-Run Effects of Wage Replacement and Job Protection: Evidence from Two Maternity Leave Reforms in Great Britain." Unpublished.
- Thornton, Saranna R. 2009. "An Examination of the Implementation and Utilization of Stop the Tenure Clock Policies in Canadian and U.S. Economics Departments." Paper presented at the American Social Sciences Association Conference, San Francisco, CA.
- Waldfogel, Jane. 1998. "The Family Gap for Young Women in the United States and Britain: Can Maternity Leave Make a Difference?" *Journal of Labor Economics* 16 (3): 505–45.

Gender Bias in Teaching Evaluations

Authors: Friederike Mengel , Jan Sauermann , Ulf Zölitz

Journal of the European Economic Association, Volume 17, Issue 2, April 2019, Pages 535–566, <https://doi.org/10.1093/jeea/jvx057>
10 February 2018

<https://academic.oup.com/jeea/article-abstract/17/2/535/4850534?redirectedFrom=fulltext&login=false>

Abstract

This paper provides new evidence on gender bias in teaching evaluations. We exploit a quasi-experimental dataset of 19,952 student evaluations of university faculty in a context where students are randomly allocated to female or male instructors. Despite the fact that neither students' grades nor self-study hours are affected by the instructor's gender, we find that women receive systematically lower teaching evaluations than their male colleagues. This bias is driven by male students' evaluations, is larger for mathematical courses, and particularly pronounced for junior women. The gender bias in teaching evaluations we document may have direct as well as indirect effects on the career progression of women by affecting junior women's confidence and through the reallocation of instructor resources away from research and toward teaching. (JEL: J16, J71, I23, J45)

Are all economics graduate cohorts created equal? Gender, job openings, and research productivity

Authors: John P. Conley, Ali Sina Önder & Benno Torgler

Published: 28 May 2016

Volume 108, pages 937–958, (2016)

<https://link.springer.com/article/10.1007/s11192-016-1987-1>

Abstract

Using life cycle publication data of 9368 economics PhD graduates from 127 U.S. institutions between 1987 and 1996, we compare research productivities of male and female graduates, and how these correlate with macroeconomic conditions prior to starting graduate studies and with availability of academic jobs at the time of graduation. We find that availability of academic jobs is positively correlated with research productivity for both male and female graduates. Unfavorable employment conditions prior to starting graduate education are negatively correlated with female graduates' research productivity and positively correlated with male graduates' research productivity.

The Earnings and Consulting Income of US Health Economists: Results from the 2012 Survey of the American Society of Health Economists

Authors: John H. Cawley, Michael A. Morrissey and Kosali I. Simon

https://www.journals.uchicago.edu/doi/epdf/10.1162/AJHE_a_00014

Abstract

This paper presents data from the first-ever survey of members of the American Society of Health Economists (ASHEcon) that was conducted in 2012. We present summary statistics of health economist earnings by rank and type of employer, and estimate log earnings models as a function of education, experience, type of employer, and research productivity. The results indicate that (1) academic salaries for health economists have risen in real terms since the previous survey in 2005; (2) we find no statistically significant evidence of disparities in academic salaries between men and women, or between whites and nonwhites; (3) there is a salary premium associated with earning a PhD at one of the top economics departments; and (4) we cannot reject the null hypothesis of no difference in salary by type of employer. We also report on the extent of consulting activities, and provide the first published data on the hourly consulting rates charged by health economists.

Gender promotion differences in economics departments in Japan: A duration analysis

Authors: Ana Maria Takahashi and Shingo Takahashi

<https://doi.org/10.1016/j.asieco.2015.09.002>

Abstract

Using a unique data set from our survey of academic economists in [Japan](#), we present the first detailed study of gender promotion gaps in Japanese academia. The length of time from initial appointment to promotion to associate professor is greater for women than men, largely due to women spending more time as lecturers, the lowest academic rank. The gender gaps in promotions from associate professor to full professor are more complex. Childless women are promoted *faster* than childless men. However, since the effects of marriage and children are negative for women, this ‘reverse gender gap’ disappears for childless married academics, and women's time to promotion becomes substantially longer than men's if they have children.

Introduction

Studies on gender promotion differences in academic labor markets have found substantial gender promotion differences in universities in both the US and UK. McDowell, Singell, and Ziliak (2001) used a sample of economists from the American Economic Association (AEA) directory and found that female academics have an approximately 12% lower probability than males of being promoted to full professor after controlling for various characteristics including publication quality. Ginther and Kahn (2004) also used a sample from the AEA directory and found that the probability of receiving tenure is 13.5% lower for women. Ginther and Hayes (2003), using a sample from the US Survey of Doctorate Recipients (SDR) in the humanities, showed that being female decreases the probability of achieving tenure by 6.8%. Ward (2001), using a sample of academics from the UK, showed that the probability that a male academic becomes a full professor is 10% higher than a comparable female, after controlling for numerous personal, job, and human capital characteristics.²

In contrast to the abundance of literature in the US and the UK, there have been few empirical studies about the gender promotion gap within Japanese universities despite a growing public interest in gender equality. In 1999, the Japanese government enacted the *Basic Law for Gender Equal Society*. Consequently, in 2000, the Association of National Universities set out an *Action Plan* stipulating that each national university should increase the proportion of female academics to 20% by 2010. In 2008, the Ministry of Education, Sports, Science and Technology (MEXT) announced that selected universities would

receive 6 million yen for each female academic they hire.³ Despite such interest in achieving gender equality in academia, little is known about gender differences within Japanese academia. A study conducted by the EPMEWSE (2008, 51-52)⁴ showed that women on average have lower ranking posts than men after controlling for age. However, the study did not control for any other differences between male and female academics. Therefore, a detailed study of gender promotion differences is called for.

We conducted the first detailed study of gender promotion differences in Japanese academia by using data that we collected via a mail survey administered in 2008. Our data set contains complete information on the year of each promotion, the exact timing of job mobility, and the types of universities where each respondent previously worked. In addition, data have personal information such as when respondents married and the ages of their children. Furthermore, our data set contains detailed information on the publication record of each academic. Thus, we are able to conduct a duration analysis of promotion while controlling for important time-varying covariates such as the number of young children and marital status at each point in time during the interval prior to promotion.

The academic labor market in Japan is of interest to researchers for the following reasons. According to our conversations with several academics, there seems to be a belief that promotions are granted almost automatically based on age, experience, and education. If promotions are decided deterministically in line with this belief, then there should be little room for gender differences. However, no empirical study to date has tested whether this common wisdom is in fact true. Therefore, whether or not there are gender promotion differences in Japanese universities is still an open empirical question.

Second, there are important institutional differences between Japanese and US universities. Unlike the US, in Japan the ‘up or out’ tenure-track system is almost non-existent, and the majority of employment contracts are on an unlimited term basis (that is, lifetime employment). Consequently, we expect to see a very different picture of gender promotion differences between Japanese and US universities.

The academic labor market is particularly well-suited for the study of gender promotion differences due to the presence of well-defined job ranks that are common across universities. Many past empirical studies that used cross-industry samples typically did not have well-defined ranks that are homogeneous across observations. Consequently, these studies used proxies for job ranks. For example, Winter-Ebmer and Zweimuller (1997) used skill requirements as proxies for job ranks. The use of such proxies always leaves the question of how comparable the job ranks are across observations. Due to well-defined job ranks in academia, we are able to avoid such problems.

The remainder of the paper is organized as follows: Section 2 presents an overview of academic careers in Japan. Section 3 explains our data source, and Section 4 describes

the construction of each variable. Section 5 presents summary statistics and provides an initial look at the data. Section 6 presents our main results. Section 7 concludes with discussions of the results.

Section snippets

Background

There are two types of universities in Japan: public (national, prefectural, and city) and private. National universities are funded by the central government, and include prestigious universities such as the University of Tokyo and Kyoto University. Prefectural and city universities are established by local governments and funded by them and the central government. Private universities are established by private entities and are largely self-supporting.

In 2004, the *National University*

Data

The data utilized in this project were obtained from a survey we administered via a postal questionnaire. The original questionnaire (in Japanese), and its English translation are in the web appendix. The majority of prior studies used publicly available data (Ginther and Hayes, 2003, Ginther and Kahn, 2004, Kahn, 1993, McDowell et al., 2001), but some studies undertook independent surveys (Ward, 2001). In Japan there are no government or private statistics on academics that allow detailed

How we constructed the variables

We utilize hazard function estimation to examine the gender differences in the following three promotion spells: (1) time to first promotion – the length of time from initial academic appointment to promotion to associate professor; (2) time to second promotion – from associate professor to full professor; and (3) to have a longer term view of promotion, time from the initial appointment to becoming a full professor. Let us explain how the survey questions were used to compute the duration of

Summary statistics and an initial look at the data

We eliminated from the analysis the observations for which we cannot determine the years respondents entered academia (6 observations). Respondents who entered academia directly as full professors were dropped from the analysis (22 observations). Respondents who entered academia in 2008 (the year of survey) were dropped because a duration that is shorter than one is treated as zero in our one-year increment data (4 observations). Four other observations were dropped due to incompleteness or

The gender differences in the initial rank at hiring

As described in Table 2, there is considerable variation in initial academic rank. Before presenting our duration analyses, it is useful to see whether there are any gender differences in initial rank assignment. Table 4 column 1 shows an ordered probit regression where the dependent variable takes the value 3 if the initial rank is associate professor, 2 if

it is assistant professor, and 1 if it is lecturer. The regression controls for age at hiring, and includes a dummy indicating if the

Discussions and conclusion

By using a unique survey data set of academic economists from Japanese universities, we conducted the first detailed study of gender differences in time to promotion within Japanese academia. Our results suggest that women have a longer first promotion spell (from their initial rank to associate professor), and that this gender promotion gap stems largely from women staying as lecturers longer, the lowest academic rank. Thus, to achieve gender promotion equality, a close monitoring of gender

Do gender quotas pass the test? Evidence from academic evaluations in Italy*

Manuel Bagues [†] Mauro Sylos-Labini [‡] Natalia Zinovyeva [§]

This version: August 29, 2014

First version: June 21, 2014

Abstract

This paper studies how the presence of women in academic committees affects the chances of success of male and female candidates. We use evidence from Italy, where candidates to Full and Associate Professor positions are required to qualify in a nation-wide evaluation known as *Abilitazione Scientifica Nazionale*. This evaluation was conducted between 2012 and 2014 in 184 academic disciplines and it attracted around 70,000 applications. In each field, committee members were selected from the pool of professors that had volunteered for the task using a random lottery. We estimate the causal effect of committees' gender composition on candidates' chances of success exploiting the existence of this system of random assignment. In a five-member committee, each additional female evaluator decreases by 2 percentage points the success rate of female candidates relative to male candidates. Information from 274,000 individual evaluation reports shows that, in mixed-gender committees, male and female evaluators are equally biased against female candidates, suggesting that the presence of women in the committee affects the voting behavior of male evaluators.

Keywords: gender quotas, discrimination, academic promotions.

JEL Classification: J71, J16, J45.

*We are grateful for all the useful comments and suggestions provided by David Cuberes, Cecilia Garcia-Peñalosa, Moreno Marzolla and participants at presentations at the *Festival dell'Economia* in Trento, Swedish Institute for Social Research (SOFI), the Workshop 'Gender in the Boardroom' (Sheffield), the 7th COSME-FEDEA Gender Workshop and the Annual Meeting of the French Economic Association in Lyon.

[†]Aalto University and IZA, Helsinki, Finland; email: manuel.bagues@aalto.fi

[‡]University of Pisa, Italy; email: mauro.syloslabini@sp.unipi.it

[§]Aalto University, Helsinki, Finland; email: natalia.zinovyeva@aalto.fi

1 Introduction

To achieve gender balance, academia has still a long way to go. In Europe women account for 46% of PhD graduates, 37% of associate professors and only a mere 20% of full professors (European Commission 2013). The US displays similar patterns and in Japan the gender imbalance is even larger (National Research Council 2009, Abe 2012).

There are several possible explanations for the persistent underrepresentation of women in top positions in academia. According to the pipeline theory, it would be mainly a matter of time that women would move their way through a metaphorical pipeline to reach top-level jobs. However, in most disciplines, the share of women among faculty members remains low even after decades of improved recruitment of women at the undergraduate and the doctoral level (Ginther and Kahn 2004, 2009). Gender differences in promotion rates might reflect differences in productivity, perhaps due to the existence of gendered roles at the household level or the lack of female mentors and role models (Blau et al. 2010). Furthermore, some authors have pointed out that women are less likely to apply for promotions (Bosquet, Combes and Garcia-Peñalosa 2013, De Paola, Ponzio and Scoppa 2014), maybe due to existence of gender differences in the preference for competitive environments (Niederle and Vesterlund 2007; Buser, Niederle and Oosterbeek 2014) or in bargaining abilities in the labour market (Babcock, Gelfand, Small and Stayn 2007; Blackaby, Booth and Frank 2005). Moreover, women seem to devote more time to tasks that are socially desirable but might not taken into account in promotion decisions (Vesterlund, Babcock and Weingart 2014).

Beyond these supply-side explanations, the paucity of women in top positions in academia has been also sometimes attributed to the existence of gender discrimination by the (mostly male) evaluators that decide on hiring and promotion decisions.¹ It has been argued that male evaluators might be subject to gender stereotypes.² They are also more likely to be acquainted with male candidates and these connections might perhaps affect the outcome of the evaluations (Zinovyeva and Bagues 2014). The lack of female evaluators might also be detrimental for female candidates if men and

¹Gender discrimination in academia remains a controversial issue. According to a meta-analysis by Ceci and Williams (2011), the more recent empirical evidence fails to support assertions of discrimination in manuscript reviewing, interviewing, and hiring. However, other studies show that female researchers might still receive lower evaluations than male researchers with identical characteristics (e.g. Steinpreis, Anders and Ritzke 1999, Moss-Racusin, Dovidio, Brescoll, Graham and Handelsman 2012).

²For instance, a report by a Spanish governmental organization, the Foundation for Science and Technology, claims that in “academia, promotion is based on a system [...] that benefits men more than women, since the barriers arise when mostly male committees evaluate female candidates and reject their promotion” (Fundación Española para la Ciencia y la Tecnología, *Mujer y Ciencia: La situación de las Mujeres Investigadoras en el Sistema Español de Ciencia y Tecnología*, (FECYT, 2005), page 48).

women differ in terms of which research areas are more valuable. For instance, several authors have documented the existence of gender segregation across different subfields in Economics (Dolado et al. 2012, Hale and Regev 2014). If evaluators overrate the importance of their own types of research, a preference for same-sex candidates might arise.

These arguments have reached policy makers. Gender quotas in scientific committees have been already introduced in countries such as Spain, Finland, Sweden, and Norway, and they are being considered in other countries. A recent report of the European Commission argues that a gender-mixed composition of scientific committees is needed to ensure that constant progress is made towards gender-equality in research and scientific careers.³ However, the empirical evidence on the impact of committees' gender composition is scarce, typically based on small samples, and it is rather inconclusive. Sometimes, applicants seem to benefit from the presence of same-sex evaluators (Li 2012, De Paola and Scoppa 2014), but most often gender does not seem to play any role (Moss-Racusin, Dovidio, Brescoll, Graham and Handelsman 2012; Steinpreis, Anders and Ritzke 1999; Abrevaya and Hamermesh 2012; Jayasinghe, Marsh and Bond 2003) or applicants obtain relatively better evaluations from opposite-sex evaluators (Broder 1993; Ellemers, Heuvel, de Gilder, Maass and Bonvini 2004). From a policy perspective, the lack of more extensive and conclusive evidence is disappointing. Gender quotas are costly for senior women, as they increase disproportionately the time that they have to devote to evaluation committees. A better understanding of its impact on recruitment and promotion decisions is key in order to determine whether they are cost-effective.

Zinovyeva and Bagues (2011) provide the first large-scale experimental evidence on the role of evaluators' gender in academic evaluations. They analyze the evaluations received by 30,000 applicants to Associate and Full Professor positions in all academic disciplines in Spain between 2002 and 2006, when the system known as *Sistema de Habilitación* was in place. The institutional set up provides an interesting randomised natural experiment. Within this system, applicants have to qualify in a centralized evaluation which is performed periodically by evaluation committees at the national level. Each time, committee members are selected by public officials from the corresponding pool of eligible evaluators using a random lottery. As a result, candidates with similar characteristics may, by the luck of the draw, be assigned to evaluation committees with different gender compositions. On average, the gender of evaluators does not have a statistically significant impact on the chances of success of male and female candidates. The authors also show that the impact of evaluators' gender does not depend on the degree of feminization of the field, it is similar across different

³European Commission, *She Figures 2012: Gender in Research and Innovation* (Luxembourg: Publication Office of the European Union, 2013), page 7.

disciplinary groups (Life Sciences, Natural Sciences, Engineering, Social Sciences and Humanities), but it seems to vary across different types of positions. Evaluators exhibit a same sex preference in competitions to Full Professor, but an opposite sex preference in competitions to Associate Professor.

In this paper, we analyze how the gender composition of academic committees affects evaluations using evidence from another large-scale natural randomized experiment: qualification evaluations in Italy. This system of promotions, known in Italian as *Abilitazione Scientifica Nazionale*, is largely similar to the system that was in place in Spain between 2002 and 2006, but also presents some distinctive features. As in the Spanish case, candidates to Associate and Full Professor positions are required to qualify first in an evaluation performed at the national level by a committee whose members are selected randomly from a pool of eligible evaluators in the field. However, while in Spain evaluations involved a research seminar given by the candidate, in Italy the evaluation relied completely on CVs and publications and it does not require any personal interaction between evaluators and candidates. Most importantly, the procedure was extremely transparent. All the relevant information — including candidates' and evaluators' CVs, bibliometric indicators and individual evaluation reports — was publicized online.⁴

Our database includes information on 69,020 applications in 184 different fields. Compared to male applicants, female applicants tend to exhibit a worse CV in terms of the quantity and quality of publications and a lower unconditional success rate. However, we do not observe any significant differences in their conditional success rate: on average, women are as likely to qualify as male candidates with similar observable research production. Next, we examine whether the chances of success of male and female candidates depend on the gender composition of committees. We compare the success rate of male and female candidates in exams where the gender composition of the committee is expected to be similar but, due to the random draw, committees end up having a different gender mix. As expected, these groups of candidates are statistically similar in every observable predetermined dimension. At the same time, we find that the success rate of female applicants is significantly lower whenever, due to the random draw, the committee includes more female evaluators. In a five-member committee, each additional female evaluator decreases by around two percentage points (five per cent) the success rate of female applicants relative to the success rate of male applicants. This effect is mainly driven by evaluations in exams for Associate Professor positions.

The availability of very detailed data also allows to examine more closely the po-

⁴A national evaluation agency collected and publicized information on bibliometric indicators for all candidates. Committees are not obliged to use this information, but there is a clear 'nudge' from the authorities recommending that committees take them into account.

tential mechanisms at work. In our data approximately 14% of candidates voluntarily withdraw from the evaluation process when the lottery is drawn and the identity of evaluators is known. We find that this decision does not depend (significantly) on the gender composition of committees. Hence, the gender composition of committees matters because it affects evaluations, and not because it influences candidates' final decision to apply.

We also analyze the individual voting behavior of committee members. Committee members typically collect information about the quality of candidates, they share it, debate about it and take a vote. Thus, the presence of women in the committee might affect the final outcome directly through their own vote, but also indirectly through their impact on the voting decisions of other committee members. Information from around 274,000 individual voting reports shows that, in mixed-gender committees, male and female evaluators are equally biased against women, suggesting that the presence of women in the committee affects the voting behavior of male evaluators. We also examine the length of evaluation reports. This variable might partly reflect the amount of information that was available about the quality of a candidate. Male and female evaluators write reports of similar length independently of the gender of the applicant.

Overall, the evidence in this paper suggests that, on average, academic evaluators do not exhibit a preference for same-sex candidates. If anything, female candidates tend to be more successful in committees with fewer female evaluators. The potential introduction of gender quotas in these committees would be detrimental both for junior female researchers, who would have lower chances of being promoted, and for senior ones, who would have less time available to work on their research.

The structure of the paper is as follows. First, section 2 discusses the related empirical literature. Section 3 describes the institutional details of the system of *Abilitazione Scientifica Nazionale*. Next, in section 4 we present the data and in section 5 we analyze the role of candidates' and evaluators' gender in evaluations and we explore possible explanations for the observed evidence. Finally, in section 6 we summarise our results and we examine some possible policy implications.

2 Literature review

The role of the evaluators' gender in academic evaluations has been studied by a number of authors.⁵ Typically the main empirical challenge has been to identify variations in the gender composition of committees that are not somehow related to the relative quality of male and female candidates. At least two papers provide experimental evidence. Moss-Racusin, Dovidio, Brescoll, Graham and Handelsman (2012) performed

⁵A related literature also analyses the role of evaluators' gender in non academic occupations (e.g. Bagues and Esteve-Volart 2010) or in sport activities (e.g. Sandberg 2014).

a randomized double-blind study where 127 biology, chemistry, and physics professors from research-intensive universities rated the application materials of an undergraduate science student who had (ostensibly) applied for a science laboratory manager position. Both male and female faculty judged a female student to be less competent and less worthy of being hired than an identical male student. Steinpreis, Anders and Ritzke (1999) asked 238 male and female academic psychologists to review the CV of an applicant who was assigned randomly a male or a female name. They do not observe any impact of evaluators' gender on the assessments.

Another empirical strategy is to compare the assessments received by the same application from different independent evaluators. The key assumption in this case is that the gender of candidates does not affect the assignment of evaluators.⁶ Broder (1993) analyzes 1,479 grant proposals to the Economics program of the National Science Foundation. She finds that female reviewers rate female-authored papers lower than do their male colleagues. Abrevaya and Hamermesh (2012) study the referee evaluations received by 2,940 submissions to a leading field journal in Economics. They find no interaction between the referees' and authors' gender. Ellemers, Heuvel, de Gilder, Maass and Bonvini (2004) study the assessments by faculty members of the work commitment of 212 PhD students in the Netherlands and in Italy. Female faculty members are most inclined to hold stereotypical views. Jayasinghe, Marsh and Bond (2003) analyze the ratings obtained by 687 grant proposals at the Australian Research Council. They do not find any evidence suggesting that assessors favor same-sex or opposite sex researchers.

While the above literature considers the behavior of independent individual evaluators, very often hiring and promotion decisions are taken by committees composed of several members. A few articles have studied the role of committees' gender composition. For instance, Li (2012) examines the extent to which evaluators' gender influences peer review at the US National Institutes of Health, using information from around 20,000 successful applications. Taking into account the observable information about the past and future research performance of successful candidates, Li concludes that reviewers exhibit a same-sex preference when they review competing applications, but not when they evaluate new applications. De Paola and Scoppa (2014) study how variations in the gender composition of evaluation committees affect the chances of success of 1,000 applicants to Full and Associate Professor positions in Economics and Chemistry in Italy, documenting a same sex preference.

Zinovyeva and Bagues (2011) provide the first evidence from a large-scale randomized natural experiment involving all academic disciplines in a given country. They analyze the evaluations received by 30,000 applicants for Full and Associate Professor

⁶For instance, it would be problematic if specially demanding evaluators are assigned to evaluate same-sex (or opposite-sex) candidates.

positions in every academic field in Spain between 2002 and 2006. On average, the chances of success of female and male candidates are unaffected by the gender composition of committees, but there might be some heterogeneous effects depending on the position at stake. Similarly, in this paper we contribute to the literature providing evidence from another large-scale randomized natural experiment. In this case, we use information from qualifying exams for Associate and Full Professor positions that were held between 2012 and 2014 in every discipline in Italy. These evaluations involved 70,000 candidates. As in Zinovyeva and Bagues (2011), the institutional setup provides a very transparent source of exogenous variation in committee composition: the random lottery that selects committee members out of the pool of eligible evaluators. Moreover, as we explain in detail below, the availability of very detailed information about committee members' individual voting behavior allows us to shed some light on the interactions that arise within committees.

3 Background

Most Italian universities are public and the recruitment of full and associate professors is regulated by national laws.⁷ Before 2010, recruitment procedures were managed locally by each university. According to some authors, this system fostered nepotism (Durante et al. 2011). In 2010, a two-stage procedure similar to those already in place in other European countries was introduced (e.g. France and Spain).⁸

In the first stage, candidates to Associate Professor (AP) and Full Professor (FP) positions are required to qualify in a national-level exam. Evaluations are conducted separately in 184 scientific fields (in Italian named *settore concorsuale*) designed by the Ministry of Education. A positive evaluation is valid for four years while a negative one implies the exclusion to participate in further national evaluations during the following two years. Qualified candidates can participate in the second stage, which is managed locally by each university. This may consist in either an open competition or the assessment of an internal candidate.⁹

⁷According to *OECD Education at a glance* (2013 edition), in 2011 about 92% of students in tertiary education were enrolled in 66 public universities and the remaining 8% in 29 independent private institutions.

⁸Law number 240/2010, also known as “Gelmini reform” after the name of the minister of Education.

⁹From 2017 on this second option will be viable only for assistant professors who want to become associate.

3.1 The national examination

The first national evaluation, which we study in this paper, was performed between 2012 and 2014, about two years after the reform approval.¹⁰ The time structure of the examinations was as follows. The call for eligible evaluators was published in June 2012. The deadline to apply was August 28. Once the list of eligible evaluators was settled, the Ministry publicized their identities and their CVs. In the meantime, the call for candidates' applications was issued in July. Candidates could apply to multiple fields and positions. The deadline for applications was November 20. Then, in January 2013, when the initial list of applicants was settled but before it was publicized on-line, committee members were selected by random draw.

Soon afterward, and before having access to the list of candidates, each evaluation committee is required to draft and publish on-line a document describing the general criteria that would be used to grant a positive evaluation. For candidates, the deadline to withdraw applications expires two weeks after these criteria are publicized. Finally, evaluation committees are informed about the final list of candidates and the examination takes place. Below we explain in more detail how evaluators are selected and how the evaluation is performed.

3.1.1 Selection of committees

Evaluation committees include five members selected by a random lottery. Four members are drawn from a pool of eligible evaluators affiliated to an Italian university (hereafter 'Italian') and the remaining member is selected from a pool of eligible evaluators who are affiliated to a foreign university (hereafter 'foreign').¹¹ The only constraint to the randomization process is that no university can have more than one evaluator within a single committee. All evaluators are in charge for two years. In case an evaluator resigns, a replacement is selected randomly from the corresponding list of eligible evaluators.

The set of 'Italian' eligible evaluators includes all full professors in the field who volunteered for the task and satisfied some minimum quality requirements. In Math, the Natural and Life Sciences, and Engineering eligible evaluators are required to have a research production above the median of full professors in the field in at least two of the following three dimensions: (i) the number of articles published in scientific journals, (ii) the number of citations, (iii) and the H-index.¹² In the Social Sciences

¹⁰A detailed description of the process is available at <http://abilitazione.miur.it/public/index.php?lang=eng>

¹¹Whenever the pool of foreign professors included less than four professors, all five committee members were drawn from the pool of eligible evaluators based in Italy.

¹²More precisely, this includes Mathematics and IT, Physics, Chemistry, Earth Sciences, Biology, Medicine, Agricultural and Veterinary Sciences, Civil Engineering and Architecture (with the exception of Design, Architectural and Urban design, Drawing, Architectural Restoration, and Urban and

and the Humanities, the research performance of eligible evaluators has to be above the median in at least one of the following three dimensions: (i) the number of articles published in high quality scientific journals (in what follows, ‘A’ journals),¹³ (ii) the overall number of articles published in any scientific journals and book chapters, and (iii) the number of published books.

The set of ‘foreign’ eligible evaluators includes Full Professors affiliated to a university from an OECD country who volunteered for the task. ‘Foreign’ eligible evaluators have to satisfy the same research requirements as ‘Italian’ ones. While ‘Italian’ evaluators work *pro bono*, OECD evaluators receive €16,000 for their participation in two evaluations.

3.1.2 The evaluation

A positive assessment of each candidate requires a qualified majority of four votes. Committees make their assessments based on candidates’ CVs. The evaluation does not involve neither exams nor interviews. In principle, committees have full autonomy on the criteria to be used in the evaluation. Nonetheless, an independent evaluation agency appointed by the ministry collected and publicized information on the research productivity of candidates in the previous ten years, as measured by the three bibliometric indicators and asked the committees to evaluate candidates’ publications also with reference to the above bibliometric indicators. These productivity measures are normalized taking into account time since first publication and job interruptions, typically related to maternity or paternity.

At the end of the process committees release for each candidate a (i) collective evaluation explaining how their final decision was taken and (ii) five individual evaluations explaining each evaluators’ position (see Figure 1).

4 Data

We use information from the first edition of the *Abilitazione Scientifica Nazionale*. For each of the 184 academic fields, the dataset includes information on (i) eligible evaluators; (ii) evaluators that eventually served in committees, (iii) applicants, and (iv) the final outcome of the evaluation. Below we describe in detail the database.¹⁴

Regional Planning), Industrial and Information Engineering, and Psychology.

¹³An evaluation agency determined with the help of several scientific committees the set of journals to be considered as high quality in each field.

¹⁴We collected the CVs of candidates and evaluators and their final evaluations from the webpage of the Ministry of Education. To avoid homonymity problems, we have excluded 14 candidates that had the same name and surname within the same exam.

4.1 Evaluators

Overall 7,239 FPs volunteered to serve in committees. In the average field, the pool of eligible evaluators includes 32 ‘Italian’ FPs and eight ‘foreign’ FPs. Approximately 20% of ‘Italian’ evaluators are women (Table 1, upper panel). The ‘foreign’ pool is less feminized: only 12% of ‘foreign’ evaluators are women. Four out of the five committee members were selected through a random draw out of the pool of ‘Italian’ eligible evaluators and the remaining committee member was drawn out of the pool of ‘foreign’ eligible evaluators.¹⁵ Taking into account the composition of both pools, the expected share of women in the committee is around 18%.¹⁶ The actual gender composition of committees is similar to the expected one: on average 18% of committee members were women. Approximately one out of every thirteen evaluators resigned and was replaced by another eligible evaluator. These replacements increased slightly the share of women in committees, but the difference is not statistically significant (Table 1, lower panel).

4.2 Initial set of applicants

The initial number of applications was equal to 69,020. On average, there were 375 applications per field, 117 in examinations for Full Professor (FP) positions and 258 in examinations for Associate Professor (AP) positions. Several candidates participated in more than one evaluation, either in different fields or in different categories of the same field. The average candidate participated in 1.5 examinations.

As shown in Table 2, most applicants are affiliated to an Italian institution (96%) and, not surprisingly, applicants tend to be older in competitions for a position of FP than in competitions for a position of AP (49 vs. 43 years old). In qualification exams for positions of FP around 31% of candidates are women and, among candidates for positions of AP, the degree of feminization is slightly larger: women account for 41% of candidates. Applicants to FP are more likely to hold a permanent position in an Italian university (74% vs. 47%) and, out of these candidates, three fourths hold a position in the same field to which they are applying. We also collected information on the relative order of application for all candidates. We normalized this variable uniformly between 0 and 1. In principle, the timing of the application can reflect both candidates’ self-confidence and their quality.

CVs include detailed information on candidates’ research production. The average

¹⁵In twenty fields the pool of foreign evaluators did not reach the required threshold of four members. In these cases all five committee members were drawn from the ‘Italian’ pool. Additionally, in two fields the pool of foreign evaluators included originally four members, but after the resignation of the foreign member her replacement was selected from the ‘Italian’ pool.

¹⁶We have calculated the expected gender composition of committees using a simulation with 1000 draws, taking into account that the lottery that decided committee composition was subject to the constraint that committees cannot include more than one member from the same university.

CV has 16 pages and it reports around 64 research outputs, including journal articles (36), books (2), book chapters (7), and patents (0.24). A typical paper is coauthored by 5 authors and only 34% of papers are single authored. The candidate reports to be the first author in 22% of the occasions and the last author in 12%.¹⁷ We also gathered information on two variables that proxy the quality of publications. In the Social Sciences and in the Humanities, we collected information on the total number of articles published in A journals.¹⁸ Candidates in these fields published on average 4 A-journal articles. In Sciences, we also consider the *Article Influence Score* (AIS) of journals.¹⁹ Summing up the AIS of all the publications of a candidate, the total AIS of the average candidate is around 61. Not surprisingly, candidates to FP positions have a relatively stronger publication record. They have a longer publication list (89 vs 53 publications) and their total AIS is also significantly higher (86 vs 49).

Table 2 also provides information about candidates' characteristics, by gender (columns 5-7). Female applicants tend to display slightly longer CVs but their publication record is significantly weaker (Table 2, columns 5-7). They are also younger, less likely to be based abroad, and more likely to have a permanent contract in the same field where they applied. We do not observe any significant difference in the timing of the application between men and women.

4.3 Final set of applicants

Approximately 14% of applicants (9,870 out of 69,020 applications) withdrew their application once the identity and the criteria of evaluators were made public. The incidence of withdrawals was significantly larger in FP examinations than in AP examinations (16% vs 13%). Withdrawals were also more common among female applicants (17% vs 13%).

The evaluation agency of the Ministry of Education collected and publicized detailed information about the research production during the 10 previous years of the final set of candidates. In the Social Sciences and the Humanities, it includes the number of books, the total number of academic articles, and the number of articles published in A-journals. In Scientific fields, the evaluation agency collected information on candidates' number of publications, citations received and on their H-index.

Using these bibliometric indicators, the evaluation agency compared the research productivity of candidates with the research productivity of professors in the category to which they applied. Around 38% of candidates were above the median in each of

¹⁷Some candidates list their name first among the list of authors of a given paper regardless of the actual ordering of authors in the publication.

¹⁸We define A-journals following the criteria of the Italian Evaluation Agency (ANVUR).

¹⁹This indicator is available for all publications in the Thomson Reuters Web of Knowledge. It is related to *Impact Factor*, but it takes into account the quality of the citing journals, the propensity to cite across journals and it excludes self-citations.

the three corresponding bibliometric dimensions. Applicants in FP examinations were relatively stronger than in AP examinations (42% vs. 36% are above the median) as well as male applicants (40% vs 34%).

Around 43% of the applicants that did not withdraw their application managed to qualify. The success rate is slightly lower if we consider all candidates that initially applied (37%). In addition to the final decision of the committee, we also collected information on the length of evaluation reports and on the voting behavior of each committee member.²⁰ The average collective evaluation report included 280 words, while individual reports were slightly shorter, around 176 words. 43% of these reports were favorable to the candidate and most of the time (80%) decisions were taken by unanimity.

Success is strongly correlated with observable research productivity (Figure 2). Among candidates whose quality was below the median in every dimension, only 4% managed to succeed. On the contrary, 63% of candidates that excelled in all three dimensions qualified.

5 Empirical analysis

The empirical analysis is organized as follows. First, we provide some descriptive information about the presence of women along the academic career ladder. Second, we compare the evaluations received by male and female applicants that participated in the *Abilitazione Scientifica Nazionale* and examine whether the gender gap varies with the gender composition of committees. Finally, in order to get a better understanding of the observed patterns, we study the impact of committees' gender composition on candidates' decision to withdraw their applications and we also analyze the information provided by individual voting reports.

5.1 Gender gap in promotions

The proportion of women decreases along the academic career ladder. As shown in Table 3, approximately 45% of assistant professors affiliated to Italian universities are women.²¹ The share of women among Assistant Professors applying for positions of Associate Professor in the national qualification exams is slightly lower, around 43%,

²⁰We conducted a text analysis of the individual evaluation reports. We identified approximately 9,000 different sentences that indicate the evaluator's decision to fail or to pass a given candidate. This sentences were used in approximately 274,000 of the 295,000 available individual evaluation reports.

²¹We collected from the Ministry of Education information on all assistant and associate professors in the field who were affiliated to an Italian university in December 2012, shortly before *Abilitazione Scientifica Nazionale* took place. Unfortunately, we do not have systematic information on the identity of other potential applicants, such as Italian researchers based abroad, researchers based in an Italian research centre (CNR) or potential applicants based in a Italian university who had some alternative contract (i.e. postdoctoral scholarship).

and women constitute only 40% of assistant professors that qualified. Similarly, while 35% of associate professors based in Italian universities are women, in evaluations for Full Professor positions the share of women is equal to 32%, and it is slightly lower among those who qualify, around 30%.

Next, we analyze why female candidates are underrepresented among successful candidates, paying particular attention to the role of committees' gender composition.

5.2 Gender differences in performance

Overall, male candidates tend to be more successful. The (unconditional) gender gap is equal to 2.2 percentage points (Table 4, column 1). In order to disentangle how much this reflects differences in the quality of candidates or in the assessments of evaluators, we compare the chances of success of male and female candidates taking into account their observable research productivity and the field and category where the evaluation is being performed. In particular, we estimate the following equation:

$$Y_{i,e} = \beta_0 + \beta_1 Female_i + \mathbf{X}_i\beta_2 + \mathbf{Z}_i\beta_3 + \mu_e + \epsilon_{i,e} \quad (1)$$

where $Y_{i,e}$ is a dummy variable that takes value one if candidate i qualifies in evaluation e (for instance, evaluation for the position of Associate Professor in Applied Economics) and value zero if the candidate fails to qualify. $Female_i$ is a dummy variable indicating the gender of candidate i . We control for the observable productivity of candidates using the information described in section 4.2 and, to account for possible non linearities, we also consider a set of indicator dummies that take value one if candidate's quality in a certain dimension is above the median in the corresponding category and field ($\mathbf{X}_i$). In addition to direct measures of research productivity, we also take into account a number of individual characteristics that might be correlated with quality, such as candidate's age, country of residence or the relative order of application ($\mathbf{Z}_i$). In the case of candidates based in an Italian university, we also consider their affiliation, the type of contract and whether they hold this contract in the same academic field to which they are applying. A set of dummies at the exam level (μ_e) controls for any overall differences across exams that might affect the success rate of male and female candidates in a similar way. As a result, the coefficient β_1 captures any systematic differences in the success rate of female and male applicants that cannot be explained by candidates' observable characteristics. We cluster standard errors at the field level to account for the fact that the same committee evaluates all candidates within a field.

Most of the unconditional gender gap is explained by differences in candidates' observable research productivity. The residual gender gap is about 0.4 percentage points and it is not significantly different from zero (columns 2 and 3). We also analyze how the gender gap varies depending on the category or the type of field. The gender

gap is slightly larger in competitions to Associate Professor and in the Social Sciences and the Humanities, but it is not significantly different from zero at standard levels (columns 4 and 7).

5.3 Does the gender composition of committees matter?

On average, female candidates are as likely to qualify as male candidates with similar observable research productivity. Next, we investigate whether the presence of female evaluators in the committee makes any difference. We estimate again equation (1) including now the interaction between the gender of the candidate and the share of women among evaluators. Interestingly, while in committees where all evaluators are male there is no gender gap, in mixed gender committees there is a significant gap in favor of male candidates (column 1, Table 5). This effect is slightly larger and more precisely estimated in evaluations of application for AP positions (columns 2 and 3). The magnitude is relatively similar in different disciplinary groups (columns 4 and 5).

The above estimates cannot be easily interpreted as causal effects. They have to be considered with caution due to the potential existence of differences in the (unobservable) quality of male and female candidates across different fields. To address this endogeneity problem, we examine how the gender composition of committees affects male and female candidates' chances of success, exploiting the exogenous variation in committee composition caused by the random draw. More precisely, we compare the success rate of male and female candidates who initially were expected to face an evaluation committee with the same gender composition but, due to the random draw, are assigned to committees with different gender compositions. To avoid any potential selection bias, we consider the pool of initial applicants, independently of whether they withdrew their application when they received information about committee composition. We estimate the following equation:

$$Y_{i,e} = \beta_0 + \beta_1 Female_i + \beta_2 Female_i * Female_e + \beta_3 Female_i * Female_e^{expected} + \mathbf{X}_i \beta_4 + \mathbf{Z}_i \beta_5 + \mu_e + \epsilon_{i,e} \quad (2)$$

where $Y_{i,e}$ takes value one if the candidate qualifies and value zero if the candidate withdrew her application or failed the evaluation. The variable $Female_e$ represents the share of female evaluators in the committee that was initially randomly drawn, before any evaluator resigned, and $Female_e^{expected}$ is the expected share of women in the committee. In order to increase the accuracy of the estimation, we also include information about individual observable productivity and individual characteristics that might be correlated with quality. Standard errors are clustered at the field level.

The coefficient β_2 captures the causal effect of committees' initial gender composition on the success rate of female candidates, relative to male candidates. The key

identifying assumption is that, conditional on the expected composition of the committee, its actual composition is uncorrelated with any unobserved factor that might affect candidates' evaluation. In other words, the consistency of β_2 relies on the assignment being indeed random. The way in which the randomization was conducted – selecting a random sequence of numbers that was then applied to several fields – suggests that there was little room for manipulation. Nonetheless, we explicitly test the randomness of the assignment. As expected, there exists no significant correlation between the random shock to committee composition and any predetermined variable (Table 6).

Table 7 reports the estimates from equation 2. The initial share of women in the committee has a significant negative impact on the relative chances of success of female candidates (column 1). In column 2, we control for candidates' observable research productivity and individual characteristics. The point estimate is slightly larger but, as expected, statistically similar. In column 3, we take into account the fact that some evaluators declined to participate in committees. More precisely, we instrument the final gender composition of the committee using the initial composition of the committee determined by the random draw. Results are very similar, if anything the impact of committees' gender composition is slightly larger. In terms of economic magnitude, an additional woman in a five-member committee decreases the relative chances of success of female candidates by approximately two percentage points ($\Delta Female_e = 1/5$; $\beta_2 = 0.123$). This pattern seems to be driven mainly by the evaluations of applications for AP positions and the magnitude of the effect is very similar in Sciences and in Social Sciences and Humanities (columns 4-7).

5.4 Do female applicants shy away from male evaluators?

The presence of women in the committee has a negative impact on the chances of success of female candidates. This might reflect evaluators' assessments or, alternatively, it might be that the gender composition of committees affects differently male and female candidates' decision to participate in the evaluation process. Some candidates might withdraw their application if they believe that evaluators might have some (direct or indirect) preference for same-sex or opposite-sex candidates. The composition of committees can also affect the information that is available to male and female candidates about their own chances of success. Evaluators might potentially provide more accurate information to same-sex candidates.

Approximately 13% of male candidates withdraw their application before the evaluation is conducted. The proportion of women that decides not to proceed with their application is significantly higher, approximately 17%. However, these gender differences in attrition rates are not related to the gender composition of committees (Table 8, columns 1-3). Women are slightly more likely to withdraw their application when

there are women in the committee, but the coefficient is economically small and it is not statistically significant. A similar result arises when we consider separately exams to FP and exams to AP positions, or exams in the Social Sciences and the Humanities vs. exams in the Sciences (columns 3-8).

5.5 Individual evaluations

Committees that include female evaluators tend to be less favorable towards female candidates. In principle, it is unclear whether this effect is directly driven by the voting behavior of female evaluators or, perhaps, by the reaction of male evaluators to the presence of female evaluators in the committee. In order to shed light on this issue we analyze the individual voting behavior of each committee member. We run the following equation at the level of individual evaluation:

$$Y_{i,c} = \beta_0 + \beta_1 Female_i + \beta_2 Female_i * Female_e + \beta_3 Female_i * Female_e^{expected} + \mathbf{X}_i \beta_4 + \mathbf{Z}_i \beta_5 + \mu_e + \epsilon_{i,c} \quad (3)$$

where the dependent variable $Y_{i,c}$ takes value one if committee member c voted in favor of candidate i and the set of independent variables is similar to the one consider in equation 2. In all-male committees, female candidates tend to obtain a similar number of positive votes as equally qualified male candidates, but in gender-mixed ones women have a significantly worse performance, although in this case the gap is only significant at the 10% (Table 9, columns 1 and 2).

Two possible mechanisms might explain this pattern. First, female evaluators might be biased against female candidates. That would explain why the presence of women in the committee hurts the chances of success of female applicants. Second, maybe male evaluators change their voting behavior in the presence of women in the committee, becoming tougher with female candidates. To disentangle between the two mechanisms, we examine whether male and female evaluators vote differently depending on the gender of the candidate. In order to account for potential differences in the quality of candidates that might be observed by evaluators but not by the econometrician, we run the following equation:

$$Y_{i,c} = \beta_0 + \beta_1 Female_i * Female_c + \mu_i + \epsilon_{ic} \quad (4)$$

where candidate fixed effects control for any characteristics of candidates that are observable by all committee members and the error term ϵ_{ic} captures any remaining differences in the evaluation criteria of committee members, for a given candidate.

If anything, female evaluators are slightly more favorable towards female candidates than male evaluators, but the difference is not statistically significant (column 3). The

gender gap observed in mixed-gender committees cannot be explained exclusively by the behavior of female evaluators. Both of male and female evaluators sitting in mixed-gender committees are relatively less favorable towards female candidates.²²

Finally, we examine the length of collective and individual evaluation reports. As we point out in a companion paper, evaluation reports tend to be longer when evaluators and candidates work in the same institution, they are co-authors or they do research in the same subfield (Bagues, Sylos-Labini and Zinovyeva 2014). The length of reports is also decreasing with the number of candidates that a committee has to evaluate, suggesting that committees with many candidates spend less time on each evaluation. In sum, the length of evaluation reports might reflect the amount of information that was available to evaluators about the quality of the candidate at the time. We run equation (3) using as dependent variable the length of committees' collective report. Evaluators write slightly longer reports on female candidates, but this gap is only significant at the 10% and it does not vary with the gender composition of the committee (Table 10, column 1). The length of individual reports does not depend either on the gender of the candidate (columns 2 and 3). In sum, there is no evidence suggesting that male or female evaluators have different information about the quality of male and female candidates.

6 Conclusions

In this paper, we analyze how the gender composition of evaluation committees affects the chances of success of candidates. We use the exceptional evidence provided by qualification evaluations to Full and Associate Professor positions in Italy. These competitions involved around 70,000 applicants and 1,000 evaluators and, due to their institutional design, they provide an extraordinary large-scale natural randomized experiment. The composition of evaluation committees is decided using a random lottery. As a result, groups of male and female candidates with similar characteristics face committees with different gender compositions.

On average, women are as likely to qualify than equally productive men. However, this gender gap varies significantly depending on the gender composition of evaluation committees. The presence of female evaluators in the committee has a significant negative impact on the success rate of female and male candidates. Each additional woman in the committee decreases the chances of success of female applicants by two percentage points relative to male applicants. Our analysis also shows that this effect is due to evaluators' behaviour, and not to candidates' potential decision to

²²This information should be considered with certain caution, given that committee members share information and discuss each case before reaching a final decision. Another potential source of concern is that, sometimes, committee members who are in a minority might decide strategically to report a positive vote in their evaluation report, in order not to antagonize candidates.

withdraw from the competition. Information from individual votes within committees suggests that this bias partly reflects the differential behavior of male evaluators when women are present in the committee. Nonetheless, we are unable to provide a clear-cut explanation of why committees including female evaluators are relatively tougher with female candidates.

Our results have important policy implications for the introduction of gender quotas in evaluation committees in academia. These quotas have been introduced in several countries and they are intended to improve meritocracy and increase the presence of women in the upper echelons of the academic career. One possible justification would be the that all-male committees are biased in favor of male candidates (or the type of research done by men). Our results are not consistent with this hypothesis. In fact, mixed gender committees are less favorable towards women than all-male committees.

There are certain features of quotas that are not captured by our analysis. Evaluators that are explicitly chosen to represent a minority might behave differently, perhaps being more inclined to look positively at candidates belonging to their own group. Moreover, maybe once quotas are introduced, the strategic incentives of evaluators can be affected. For instance, given the disproportionate time that senior women would have to spend sitting in committees, there might be an incentive to increase the number of women that qualify. Keeping in mind these limitations, our results suggest that the introduction of gender quotas in evaluation committees might have unintended consequences. Quotas might be detrimental both for senior female researchers, who would have to spend more time sitting in committees, and for junior ones, who might have lower chances of success. A back of the envelope calculation suggests that a rule requiring the presence of at least 40% of women (or men) would imply that 700 less women qualify. Moreover, female full professors would have to form part of committees at least 3 times more often than men, and they would probably end up having less time to devote to their own research than their male colleagues.²³

²³We have chosen the 40% threshold following the example of countries such as Spain or Norway. In our data, only 18% of eligible evaluators are women. In order to satisfy a 40% quota, this implies that women would have to participate in committees at least 3 times as often as men $[(0.4/0.18)/(0.6/0.82)]$. On the other hand, 26,275 female candidates participated in these qualification exams. According to the estimate in Table 7, column 3, with a 40% quota 711 additional women would have failed to qualify $[-0.123*(0.40-0.18)*26,275]$.

References

- Abrevaya, Jason and Daniel S. Hamermesh (2012), "Charity and Favoritism in the Field: Are Female Economists Nicer (to Each Other)?," *Review of Economic and Statistics*, 94(1), 202-207.
- Abe, Yukiko (2012), "The Academic Labor Market in Japan and the Presence of Women," *CSWEP Newsletter*, Fall, 9-10.
- Babcock, Linda, Michele Gelfand, Deborah Small, and Heidi Stayn (2007), "Who gets to the bargaining table: The influence of gender and framing on the propensity to initiate negotiations," *Journal of Personality and Social Psychology*, 93, 600-613.
- Bagues, Manuel and Berta Esteve-Volart (2010), "Can Gender Parity Break the Glass Ceiling? Evidence from a Repeated Randomized Experiment," *Review of Economic Studies*, 77(4), 1301-28.
- Bagues, Manuel, Mauro Sylos-Labini and Natalia Zinovyeva (2014) "Can Transparency Overcome Evaluation Biases in Academia? Evidence from Italy", Aalto University, mimeo.
- Blackaby, David, Alison L. Booth and Jeff Frank (2005), "Outside Offers And The Gender Pay Gap: Empirical Evidence From the UK Academic Labour Market," *Economic Journal*, 115(501), F81-F107
- Blau, Francine D., Janet M. Curie, Rachel T.A. Croson and Donna K. Ginther (2010), "Can Mentoring Help Female Assistant Professors? Interim Results from a Randomized Trial," *American Economic Review*, 100(2), 348-352.
- Bosquet, Clément, Pierre-Philippe Combes and Cecilia Garcia-Peñalosa (2013), "Gender and Competition: Evidence from Academic Promotions in France", *Sciences Po Economics Discussion Papers* 17.
- Broder, Ivy E. (1993), "Review of NSF Economics Proposals: Gender and Institutional Patterns," *American Economic Review*, 83(4), 964-970.
- Buser, Thomas, Muriel Niederle and Hessel Oosterbeek (2014), "Gender, Competitiveness and Career Choices," *Quarterly Journal of Economics*, forthcoming.
- Ceci, Stephen J. and Williams, Wendy M. (2011), "Understanding Current Causes of Women's Underrepresentation in Science," *Proceedings of the National Academy of Sciences*, 108.
- CSWEP (2008), "Report of the Committee on the Status of Women in the Economics Profession 2008," *American Economic Review: Papers & Proceedings*, 99(2), 692-699.
- De Paola, Maria, Michela Ponzio and Vincenzo Scoppa (2014), "Gender Differences in Attitudes Towards Competition: Evidence from the Italian Scientific Qualification," University of Calabria, mimeo.
- De Paola, Maria and Vincenzo Scoppa (2014), "Gender Discrimination and Evaluators' Gender: Evidence from the Italian Academy," *Economica*, forthcoming.
- Dolado, Juan Jose, Florentino Felgueroso and Miguel Almunia (2012), "Are Men and Women-Economists Evenly Distributed across Research Fields? Some New

- Empirical Evidence,” *Journal of the Spanish Economic Association*, 3(3), 367-393.
- Durante, Ruben, Giovanna Labartino and Roberto Perotti (2011), “Academic Dynasties: Decentralization and Familism in the Italian Academia,” *NBER working paper* n. 17572.
- Ellemers, Naomi, van den Heuvel, Henriette, de Gilder, Dick, Maass, Anne and Alessandra Bonvini (2004), “The underrepresentation of women in science: Differential commitment or the queen bee syndrome?” *British Journal of Social Psychology*, 43, 1-24.
- Ginther, Donna K. and Shulamit Kahn (2004), “Women in Economics: Moving Up or Falling Off the Academic Ladder,” *Journal of Economic Perspectives*, 18(3), 193-214.
- Ginther, Donna K. and Shulamit Kahn (2009), “Does Science Promote Women? Evidence from Academia 1973-2001,” NBER chapters in: *Science and Engineering Careers in the United States: An Analysis of Markets and Employment*, 163-194.
- Hale, Galina and Tali Regev (2014), “Gender Ratios at Top PhD Programs in Economics,” *Economics of Education Review*, forthcoming.
- Jayasinghe, Upali W., Marsh, Herbert W. and Nigel Bond (2003), “A multilevel cross-classified modelling approach to peer review of grant proposals: the effects of assessor and researcher attributes on assessor ratings”, *Journal of the Royal Statistical Society*, 166(3), 279-300.
- Li, Danielle (2012), “Gender bias in NIH Peer Review: Does it Exist and Does it Matter?”, MIT, mimeo
- National Research Council (2009), *Gender Differences at Critical Transitions in the Careers of Science, Engineering, and Mathematics Faculty*, Washington D.C.: The National Academies Press.
- Niederle, Muriel and Lise Vesterlund (2007), “Do Women Shy Away From Competition? Do Men Compete Too Much?”, *Quarterly Journal of Economics*, 122(3), 1067-1101.
- Moss-Racusin, Corinne A., John F. Dovidio, Victoria L Brescoll, Mark J. Graham and Jo Handelsman (2012), “Science faculty’s subtle gender biases favor male students,” *Proceedings of the National Academy of Sciences*, 109(41), 16474-16479.
- Sandberg, Anna (2014), “Competing biases: Effects of gender and nationality in sports judging,” Stockholm School of Economics, mimeo.
- Steinpreis, Rhea E., Anders, Katie A. and Dawn Ritzke (1999) “The Impact of Gender on the Review of the Curricula Vitae of Job Applicants and Tenure Candidates: A National Empirical Study,” *Sex Roles*, 41(7), 509-527.
- Vesterlund, Lise, Babcock, Linda and Laurie Weingart (2014), “Breaking the Glass Ceiling with “No”: Gender Differences in Declining Requests for Non Promotable Tasks”, University of Pittsburgh, mimeo.
- Zinovyeva, Natalia and Manuel Bagues (2011), “Does Gender Matter for Academic Promotion? Evidence from a Randomized Natural Experiment,” IZA Discussion Paper 5537.

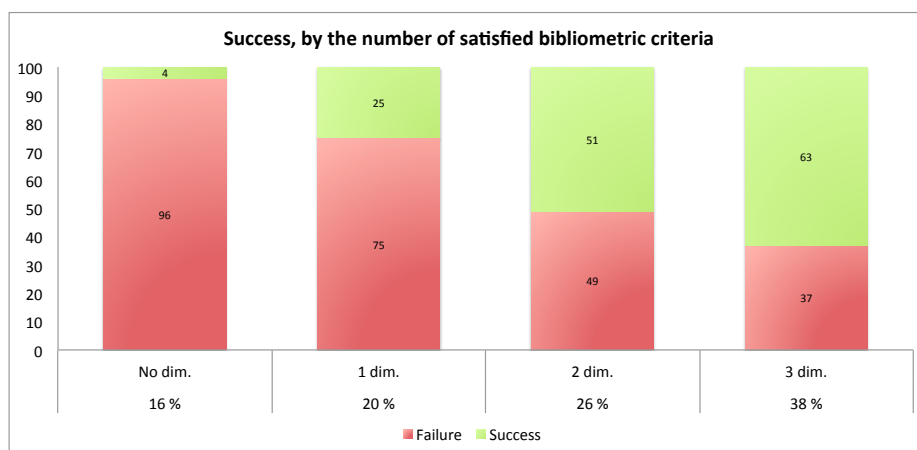
Zinovyeva, Natalia and Manuel Bagues (2014), “The Role of Connections in Academic Promotions,” *American Economic Journals: Applied Economics*, forthcoming.

Figure 1: Sample Individual Evaluation

DOE John

The candidate MARCO ROSSI has been Ricercatore universitario at the Università di XXX since 2006. His scientific work is concerned with the development of democracy, including a monograph on the role of public opinion in political thought and a series of contributions concerning English and Anglo-American thought and developments from the 17th through 19th centuries, with special reference to Edmund Burke. The candidate is a member of the "Re-Imagining Democracy in the Mediterranean, 1750-1860" project, based at the University of Oxford. The candidate has a significant number of international conference participations, among which those in which the English have invited him to speak about Burke are perhaps the most indicative of a strong international reputation. In terms of specific contributions, the "silent guest" metaphor is particularly significant in explaining how Burke plays out in the history of Italian political thought. The candidate scores above the median on two of the three indicators of impact and has substantial relevant teaching experience. On the basis of the application submitted, the candidate merits approval of the request for the abilitazione scientifica.

Figure 2: Success rate and bibliometric measures



Note: Each bar provides information on the success rate of a group of candidates. Candidates are classified in four groups, depending on the number of dimensions where their productivity is above the median in the corresponding category. The first bar provides information for candidates who are above the median in every dimension. The second, third and fourth provide information for candidates that excel in one, two and three dimensions respectively.

Table 1: Descriptive statistics – Evaluators

	1	2	3	4
	Mean	Std	Min	Max
Eligible valuers				
Based in Italy:				
- number	32	19	8	128
- share of women	0.20	0.15	0	0.73
Based abroad:				
- number	8	6	4	44
- share of women	0.12	0.15	0	0.75
Expected share of women in committee	0.18	0.14	0	0.69
Committee members				
Number of resignations per committee	0.38	0.67	0	5
Initial share of women in committee	0.18	0.20	0	1
Final share of women in committee	0.19	0.20	0	1

Note: The table includes information from the 184 fields where qualification exams were held. The pool of *Eligible evaluators based abroad* refers only to 162 fields where the number of foreign based eligible evaluators was above 4.

Table 2: Descriptive statistics – Candidates

	1	2	3	4	5	6	7
Initial set of candidates (N=69,020)							
	All exams		FP	AP	Male	Female	
	Mean	St.Dev.	Mean	Mean	Mean*	Mean*	p-value
Individual characteristics:							
Female	0.38	0.49	0.31	0.41			
Age	44	8	49	43	0.01	-0.02	0.000
Based in Italy	0.96	0.18	0.96	0.96	0.96	0.97	0.000
Permanent position:	0.55	0.50	0.74	0.47	0.55	0.56	0.190
- same field	0.75	0.43	0.77	0.74	0.74	0.78	0.000
Relative order of application	0.50	0.29	0.50	0.50	0.50	0.50	0.101
Quality indicators:							
CV length, pages	16	66	20	14	-0.02	0.03	0.000
All Publications:	64	67	89	53	0.04	-0.07	0.000
- Articles	36	51	52	29	0.07	-0.11	0.000
- Books	2	3	2	1	0.05	-0.08	0.000
- Book chapters	7	10	9	5	0.01	-0.02	0.000
- Patents	0.24	1.65	0.35	0.19	0.02	-0.04	0.000
- Other	20	32	25	17	-0.00	0.00	0.822
Average number of coauthors	5	6	5	5	-0.02	0.04	0.000
Single-authored	0.34	0.4	0.32	0.34	0.04	-0.06	0.000
First-authored	0.22	0.2	0.22	0.22	-0.00	0.00	0.793
Last-authored	0.12	0.16	0.15	0.11	0.02	-0.03	0.000
Total AIS (Sciences)	61	90	86	49	0.05	-0.09	0.000
A-journal articles (SSH)	4	8	7	4	0.04	-0.05	0.000
Outcome variables:							
Withdrawal	0.14	0.35	0.16	0.13	0.13	0.17	0.000
Failure	0.49	0.50	0.48	0.50	0.49	0.49	0.413
Qualified	0.37	0.48	0.36	0.37	0.38	0.35	0.000
Final set of candidates (N=59,150)							
Production previous 10 years:							
Social Sciences and Humanities							
- Articles	20	19	25	18	0.04	-0.06	0.000
- A-journal articles	3	4	3	2	0.04	-0.05	0.000
- Books	2	4	3	2	0.07	-0.10	0.000
Sciences							
- Articles	37	45	46	32	0.04	-0.08	0.000
- Citations	60	102	77	52	0.02	-0.04	0.000
- H-index	11	7	13	10	0.03	-0.06	0.000
Above median in all dimensions	0.38	0.48	0.42	0.36	0.40	0.34	0.000
Below median in all dimensions	0.16	0.36	0.13	0.17	0.15	0.17	0.000
Evaluations:							
Qualified	0.43	0.49	0.43	0.43	0.44	0.41	0.000
Unanimous decision	0.80	0.40	0.79	0.80	0.81	0.78	0.000
Individual votes	0.44	0.47	0.46	0.44	0.45	0.43	0.000
Length of collective evaluations	280	231	304	269	-0.01	0.02	0.000
Length of individual evaluations	176	277	203	164	-0.00	0.00	0.119

Note: Detailed productivity indicators for the 10 years prior to exam were provided by the evaluation agency only for the sample of final applicants. These indicators differ for researchers in Science and in the Social Sciences and Humanities. Individual evaluations are not available for 3% of the candidates.

(*) In columns 5-6 productivity indicators and age are normalized at the exam level. Column 7 reports the p-value for the t-test of difference in means between genders.

Table 3: Descriptive statistics – Share of women

Sample of candidates:	All	Permanent position in an Italian university:	
		Yes	No
AP exams			
Potential	-	0.45	-
Candidates	0.41	0.43	0.40
Qualified	0.39	0.40	0.37
FP exams			
Potential	-	0.35	-
Candidates	0.31	0.32	0.27
Qualified	0.29	0.30	0.22

Note: The set of potential candidates for Associate Professor and Full Professor positions includes respectively Associate Professors and Assistant Professors with a permanent position in an Italian university.

Table 4: Performance of male and female candidates

	1	2	3	4	5	6	7
	All			Category:		Field:	
				FP	AP	Sciences	SSH
Female	-0.022** (0.011)	-0.004 (0.005)	-0.006 (0.005)	-0.002 (0.008)	-0.007 (0.006)	0.001 (0.007)	-0.014* (0.008)
Quality controls	No	Yes	Yes	Yes	Yes	Yes	Yes
Individual controls	No	No	Yes	Yes	Yes	Yes	Yes
Exam FE	No	Yes	Yes	Yes	Yes	Yes	Yes
Adj. R-squared	0.001	0.374	0.395	0.386	0.400	0.426	0.344
N	59150	59150	59150	18061	41089	35856	23294

Note: OLS estimates. The dependent variable is an indicator for qualified candidates. Quality controls include normalized CV length, publications (by type), average number of coauthors, proportion of first and last-authored articles, total Article Influence Score, citations and H-index (in Sciences), A-journal articles and books (in Social Sciences and Humanities), dummies indicating whether candidates' observable productivity is above the median in the corresponding category and field, and the indicator for a permanent position in an Italian university in the same field. Individual controls include quadratic control for age, normalized application order, and indicators for researchers based abroad, exact position and university. Standard errors clustered at the field level are reported in parentheses. *** denotes significance at 1%, ** significance at 5% and * significance at 1%.

Table 5: Performance of male and female candidates - by evaluators' gender

	1	2	3	4	5
	All	FP	AP	Sciences	SSH
Female	0.005 (0.007)	0.006 (0.011)	0.005 (0.008)	0.007 (0.009)	0.001 (0.013)
Female*Share of women in committee	-0.056** (0.028)	-0.043 (0.046)	-0.063** (0.028)	-0.043 (0.032)	-0.055 (0.045)
Quality controls	Yes	Yes	Yes	Yes	Yes
Individual controls	Yes	Yes	Yes	Yes	Yes
Exam FE	Yes	Yes	Yes	Yes	Yes
Adj. R-squared	0.396	0.386	0.400	0.426	0.345
N	59150	18061	41089	35856	23294

Note: OLS estimates. Standard errors clustered at the field level are reported in parentheses.

*** denotes significance at 1%, ** significance at 5% and * significance at 1%.

Table 6: Randomization check

	1	2	3	4	5	6	7	8	9	10
Dependent variable:	CV length	Publ.	A-journal articles	Total AIS	Coauthors	Age	Appl. order	In Italy, same field	In Italy, other field	Based abroad
Female*Share of women in committee	0.073 (0.089)	-0.035 (0.085)	-0.092 (0.104)	0.082 (0.115)	-0.037 (0.070)	1.026* (0.589)	-0.034 (0.023)	0.051 (0.038)	0.018 (0.021)	-0.026 (0.017)
Female* Expected share of women	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Exam FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	69020	69020	27625	41395	69020	69020	69020	69020	69020	69020

Note: OLS estimates. The *Expected female share* among evaluators was computed using 1000 simulated draws from the pool of eligible evaluators, taking into account the constraint that committees cannot include more than one member from the same university. Standard errors clustered at the field level are reported in parentheses. *** denotes significance at 1%, ** significance at 5% and * significance at 1%.

Table 7: Does the gender of evaluators affect candidates' chances of success?

	1	2	3	4	5	6	7
Sample:	All			Category:		Field:	
				FP	AP	Sciences	SSH
			IV	IV	IV	IV	IV
Female	-0.015* (0.009)	0.000 (0.007)	0.005 (0.008)	-0.001 (0.012)	0.008 (0.010)	-0.000 (0.008)	0.010 (0.018)
Female*Share of women in committee	-0.073** (0.036)	-0.100*** (0.029)	-0.123*** (0.038)	-0.101 (0.063)	-0.134*** (0.040)	-0.113*** (0.038)	-0.125** (0.063)
Quality controls	No	Yes	Yes	Yes	Yes	Yes	Yes
Individual controls	No	Yes	Yes	Yes	Yes	Yes	Yes
Female*Expected share of women	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Exam FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adj. R-squared	0.081	0.277	0.277	0.280	0.275	0.268	0.287
N	69020	69020	69020	21594	47426	41395	27625

Note: The dependent variable is an indicator for candidates that qualified. Columns 1 and 2 provide OLS estimates. In columns 3-7 the share of female evaluators in the final committee has been instrumented using the outcome of the initial random draw. Quality controls include normalized CV length, publications (by type), average number of coauthors, proportion of first and last-authored articles, total Article Influence Score, A-journal articles, and the indicator for a permanent position in an Italian university in the same field. Individual controls include quadratic control for age, normalized application order, and indicators for researchers based abroad, exact position and university. Expected share of women in the committee is computed for an average of 1000 simulated draws from the pool of eligible evaluators, taking into account the constraint that committees cannot include more than one member from the same university. It is centered to have zero mean in the corresponding sample of candidates. Standard errors clustered at the field level are reported in parentheses. *** denotes significance at 1%, ** significance at 5% and * significance at 1%.

Table 8: Do women shy away from male evaluators?

	1	2	3	4	5	6	7
	All			Category:		Field:	
				FP	AP	Sciences	SSH
			IV	IV	IV	IV	IV
Female	0.045*** (0.005)	0.028*** (0.005)	0.028*** (0.006)	0.036*** (0.009)	0.024*** (0.007)	0.028*** (0.006)	0.027*** (0.013)
Female*Share of women in committee	0.012 (0.021)	0.019 (0.021)	0.023 (0.026)	0.024 (0.050)	0.022 (0.028)	0.023 (0.028)	0.013 (0.042)
Quality controls	No	Yes	Yes	Yes	Yes	Yes	Yes
Individual controls	No	Yes	Yes	Yes	Yes	Yes	Yes
Female*Expected share of women	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Exam FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adj. R-squared	0.046	0.109	0.109	0.140	0.091	0.098	0.125
N	69020	69020	69020	21594	47426	41395	27625

Note: The dependent variable is an indicator for candidates who withdrew their application before the evaluation took place. Columns 1 and 2 provide OLS estimates. In columns 3-7 the share of female evaluators in the final committee has been instrumented using the outcome of the initial random draw. Quality controls include normalized CV length, publications (by type), average number of coauthors, proportion of first and last-authored articles, total Article Influence Score, A-journal articles, and the indicator for a permanent position in an Italian university in the same field. Individual controls include quadratic control for age, normalized application order, and indicators for researchers based abroad, exact position and university. Expected share of women in the committee is computed for an average of 1000 simulated draws from the pool of eligible evaluators, taking into account the constraint that committees cannot include more than one member from the same university. It is centered to have zero mean in the corresponding sample of candidates. Standard errors clustered at the field level are reported in parentheses. *** denotes significance at 1%, ** significance at 5% and * significance at 1%.

Table 9: Evaluators' individual votes

	1	2	3
Female candidate	-0.007 (0.005)	0.002 (0.007)	
Female candidate*Share of women in committee		-0.047* (0.025)	
Female candidate*Female evaluator			0.006 (0.006)
Quality controls	Yes	Yes	Yes
Individual controls	Yes	Yes	Yes
Exam FE	Yes	Yes	Yes
Application FE			Yes
Adj. R-squared	0.407	0.407	0.897
N	274256	274256	274256

Note: Each observation corresponds to a candidate-evaluator pair. The dependent variable takes value one (zero) if the vote was positive (negative). Standard errors are clustered at the field level. Quality controls include normalized CV length, publications (by type), average number of coauthors, proportion of first and last-authored articles, total Article Influence Score, A-journal articles, and the indicator for a permanent position in an Italian university in the same field. Individual controls include quadratic control for age, normalized application order, and indicators for researchers based abroad, exact position and university.

Table 10: Length of Evaluation Reports

	1	2	3
Dependent variable:	Length of ...		
	collective report	individual report	
Female candidate	0.007* (0.004)	0.004 (0.004)	
Female candidate*Share women in committee	-0.006 (0.012)	-0.013 (0.012)	
Female candidate*Female evaluator			0.011 (0.020)
Quality controls	Yes	Yes	Yes
Individual controls	Yes	Yes	Yes
Exam FE	Yes	Yes	Yes
Candidate FE			Yes
Adj. R-squared	0.847	0.881	0.726
N	58691	294738	294738

Note: The left-hand side is the log number of words contained in the collective report (column 1) or in the individual evaluation report of each committee member (columns 2 and 3). Quality controls include normalized CV length, publications (by type), average number of coauthors, proportion of first and last-authored articles, total Article Influence Score, A-journal articles, and an indicator for a permanent position in an Italian university in the same field. Individual controls include quadratic control for age, normalized application order, and indicators for researchers based abroad, exact position and university.



Working Papers

www.cesifo.org/wp

Gender and Competition: Evidence from Academic Promotions in France

Clément Bosquet
Pierre-Philippe Combes
Cecilia García-Peñalosa

CESIFO WORKING PAPER NO. 4507

CATEGORY 4: LABOUR MARKETS

DECEMBER 2013

An electronic version of the paper may be downloaded

- from the SSRN website: www.SSRN.com
- from the RePEc website: www.RePEc.org
- from the CESifo website: www.CESifo-group.org/wp

CESifo

Center for Economic Studies & Ifo Institute

Gender and Competition: Evidence from Academic Promotions in France

Abstract

Differences in promotion across genders are still prevalent in many occupations. Recent work based on experimental evidence indicates that women participate less in or exert lower effort during contests. We exploit the unique features of the promotion system for French academics to look at women's attitudes towards competition in an actual labour market. Using data for academic economists over the period 1991-2008 we find that, conditional on entering the competition, there is no difference in promotions across the genders, which is difficult to reconcile with either discrimination or a poorer performance of women in contests. In contrast, women have a substantially lower probability than men to enter the promotion contest. Our data does not support that this gap is due to differences in costs or preferences concerning department prestige, indicating that women are less willing than men to take part in contests.

JEL-Code: J160, J700, I230.

Keywords: gender gaps, promotions, academic labour market.

Clément Bosquet
London School of Economics and
Political Science (SERC)
Houghton Street
UK – London WC2A 2AE
C.Bosquet@lse.ac.uk

Pierre-Philippe Combes
Aix-Marseille University
School of Economics
2 Rue de la Charité
France – 13236 Marseille Cedex 02
ppcombes@gmail.com

Cecilia García-Peñalosa
Aix-Marseille University
School of Economics
2 Rue de la Charité
France – 13236 Marseille Cedex 02
cecilia.garcia-penalosa@univ-amu.fr

November 26, 2013

This paper has benefited from the excellent research assistance of Philippe Donnay and Charles Laitong. We are grateful to Marco Manacorda, Esteban Aucejo, Javier Ortega, Felix Weinhardt, Stefano Bosi, Rodolphe Dos Santos Ferreira, Vêlayoudom Marimoutu, Michel Mougeot, Boris Racine, as well as participants at seminars at the LSE and Greqam for helpful suggestions and assistance. This project has been supported by the Agence Nationale de la Recherche, grant ANR-11-BSH1-0014.

1 Introduction

Despite the rapid increase in female educational attainment over the last decades, the labour market outcomes of men and women still differ in terms of wages and particularly in seniority. The literature on gender wage gaps is vast, but differences across genders in promotions have received much less attention.¹ These gaps can be large. For example, Bertrand and Hallock (2001) find that women account for only 2.5 percent of top executives in US firms. Differences in characteristics account for a large proportion of the gap, with the remaining fraction being usually attributed to preferences or discrimination. A recent literature has proposed an alternative explanation, namely that women tend to be less likely to enter competitions and tend to perform worse in competitions than men². Since promotions to top jobs tend to be highly competitive, such behaviour could explain why we observe fewer women in these jobs. Existing evidence is based on experiments conducted in laboratory conditions, and although the results present strong support for gender differences in attitudes during and towards contests, it is not obvious that such attitudes are also prevalent in actual labour markets.

This paper uses the particular features of the French academic system, namely the fact that promotions occur through national contests, to look at the effect of gender on attitudes towards competition. In the US and the UK academic systems or in private sector jobs only information on actual promotions is available, while the list of applicants tends not to be known.³ In France, academics are public servants and promotion occurs through a national contest or *concours*, with the lists of applicants being publicly available at the time of the *concours*. We can hence examine how gender affects both the likelihood that an academic applies for promotion and the probability of promotion conditional on having applied. As a result, we can ask whether the lower observed promotion rate of women is due to lower success in competitions or to their unwillingness to enter them.

Using data for academics has important advantages. Unlike in many private sector jobs, where a promotion is associated with longer hours and a requirement for greater

¹For an overview of work on gender wage gaps see Blau and Kahn (2000). See Lazear and Rosen (1990) for a theoretical analysis of gender gaps in promotion and Winter-Ebmer and Zweimuller (1997) and Goldin and Rouse (2000) for empirical evidence.

²See Gneezy, Niederle and Rustichini (2003), Niederle and Vesterlund (2007) and Gupta, Poulsen and Villeval (2013)

³An exception is the analysis of auditions for orchestra membership and promotion in Goldin and Rouse (2000).

availability outside normal working hours, academics have similar obligations and constraints at all hierarchical levels. Even if more senior academics tend to be involved in university administration and outside responsibilities such as participating in committees, seeking funding or performing editorial activities, these activities are to a large extent voluntary. They are not ‘required’ by the promotion and not performing them would not imply that the individual is demoted. Female associate professors should thus not feel more constrained in terms of combining career and family duties by becoming full professors, and there is hence no obvious reason why they would prefer not to be promoted. Male and female academics are also likely to have rather homogeneous labour market attachment, as argued by Kahn (1995), removing one of the reasons often branded to justify lower promotion rates for females. A further advantage of these data is that, unlike for most types of jobs, the key consideration in actual promotion decisions, an individual’s productivity (publications), can be observed and thus controlled for by the researcher.

In this context, we consider several potential causes of the unexplained component of promotions gaps. On the one hand, women may be less likely to be promoted conditional on having applied, and this could be due either to discrimination or under-performance during the *concours*. On the other, female academics may have a lower propensity to apply for promotion than males, which could be explained by the requirements of the contest being more costly for women, to their facing a different tradeoff between salaries and department prestige, or simply to an unwillingness to enter the contest. The special features of the French academic system, such as a national salary scale, the need to go through a national contest in order to be promoted, and the existence of several categories of academics with different requirements during the contest and upon promotion, allow us to test for these hypotheses.

We use data for academic economists in France over the period 1991 to 2008, and find lower promotion rates for women, which are partly but not completely explained by the age structure and publication records. We then consider separately the determinants of the likelihood to enter a promotion *concours* and the probability of being promoted conditional on having entered the *concours*. The latter is unaffected by gender, thus rejecting the idea that females underperform in contests, and lending no support to the presence of discrimination against women. In contrast, being a woman has a substantial

negative effect on the likelihood to enter the *concours*.

We explore several potential explanations for this difference. We find no evidence that either a different trade-off between income and department prestige nor the cost undertaking the *concours* and the mobility implications associated with it (see below) are the causes of the gender gap. The results seem to indicate that women are less willing than men to enter competitions for promotion, and this could be due either to different attitudes towards contests or to the expectation of being discriminated against. When we look at features of each year's *concours*, we find evidence consistent with both aspects.

Our contribution to the literature is twofold. First, we provide evidence on female attitudes towards competition in an actual labour market and hence add to the growing body that has addressed this question but which has so far been based on experimental evidence; see Gneezy et al. (2003), Niederle and Vesterlund (2007), Niederle and Vesterlund (2011) and Gupta et al. (2013). Our conclusions are partly supportive of what the experimental literature has found. In line with the results in Niederle and Vesterlund (2007), Niederle and Vesterlund (2011) and Gupta et al. (2013) we find that once we control for observed productivity women are less likely than men to enter the *concours* and that this difference does not depend on the visible costs associated with the contest. In contrast, we find no evidence that women are less likely to be promoted conditional on being a candidate, which can be interpreted as a lack of evidence for the under-performance in contests found by Gneezy et al. (2003).

The paper also contributes to the literature using data on academics to try to understand gender gaps in labour markets, which dates back to the seminal work of Cole and Cole (1973). Early on, empirical analyses identified both lower wages and lower promotion rates for female economists; see Johnson and Stafford (1974) and Farber (1977). More recent work, such as Ginther and Hayes (1999) and Ginther and Kahn (2004), indicates that in the US salary gaps are explained by differences in academic rank, while promotions to tenure and to full professor rank are still affected by gender even after controlling for research output and demographic characteristics. Some studies claim a decline in the promotion gap over time, while others find that it is large even in recent decades; see McDowell, Singell and Ziliak (2001) and Ginther and Kahn (2004). Evidence for the UK by Blackaby, Booth and Frank (2005) indicates that there are both gaps in promotions and in within-rank pay across genders, and their findings suggest that these

are partly due to differences in the outside offers received by men and by women, while Sabatier (2010) documents the existence of a promotion gap in France.

Most of this literature has considered the US and the UK, which have an academic labour market with much greater wage and promotion flexibility than those found in most European countries. Our paper focuses on a different promotion system, and thus we can examine whether promotion gaps also exist in a labour market that operates in an entirely different way, with salaries being fixed at the country-wide level and promotions being decided by national committees and not by the department where the individual is employed.

The paper is organised as follows. Section 2 starts by examining the possible reasons why women are less likely to be promoted, we then describe the French academic system and explain how the different hypotheses can be tested. Section 3 describes the data, an exhaustive panel of academic economists in France over the period 1991 to 2008. Our results are presented in section 4, while section 5 concludes.

2 Why are there so few female professors?

2.1 Discrimination, differences in payoffs or self-selection?

Although a substantial literature has examined the promotion gap across genders, a clear explanation is still lacking. Women may be less likely to be promoted either because they apply for promotion less often than men or because of a lower probability of being promoted conditional on being a candidate. These differences may in turn have three possible causes: discrimination, differences in costs of or rewards from promotion across genders, and different attitudes in and towards the promotion process itself. Let us consider them in turns.

The first possible explanation is simply that those making the promotion decisions discriminate against women. This discrimination may be due to tastes or statistical, and implies that a promotion committee would, for given research output and performance during the selection process, prefer to promote a man. Female candidates may also be less likely to succeed if they tend to under-perform in contests, as shown by Gneezy et al. (2003). As we will see below, the competitive nature of the promotion process for (most) academic promotions in France implies that if women ‘hold back’ during compe-

titions they may perform worse than men and hence will be promoted less often. Both discrimination and underperformance during the promotion process will result in female candidates having a lower likelihood of promotion than male candidates.

Alternatively, women may differ in the objective costs of and rewards from promotion and this can make them less willing to apply, which in turn result in different observed promotion rates. There can be various causes of differences in the propensity to seek promotion. First, there may be costs associated with promotion, and if these are higher for women than for men, they could result in the former applying for a promotion less often than men. In France, promotions tend to require moving to a different department and thus tend to imply moving to a different city (see below for details on the French academic system). If women's bargaining power in the household is lower, it may be harder for female than for male academics to impose the cost of moving on their families. As a result, the cost of a promotion is probably higher for women and hence they will be less likely to seek promotion.

Another reason why women may be less likely to apply for promotion is if the payoff is lower than for men. One possible cause of differences in payoff is that women may have different preferences over department prestige and income. Suppose that women have a lower marginal utility of income because, often, they are the second earner in the household. They may then be willing to be in a more prestigious department even if it implies lower wages. Since in the US and UK more prestigious departments also have tougher promotion thresholds (see McDowell et al. (2001)), this would result in a lower promotion probability for women.

Lastly, women may choose not to apply for promotion even if we observe no objective differences in the cost or payoff of promotion. This type of 'self-selection' out of the promotion race can be due to two reasons. On the one hand, female academics may expect to be discriminated against during the promotion process. Such a belief will reduce the expected payoff and hence the propensity to seek promotion. Note that this may happen even in the absence of actual discrimination. On the other hand, women may choose not to apply for promotion simply because they are less inclined to compete in tournaments than men, as indicated by experimental evidence; see Niederle and Vesterlund (2007), Niederle and Vesterlund (2011) and Gupta et al. (2013) as well as the discussion in Bertrand (2011). While certain aspects of an academic career, such as obtaining a

PhD, depend only on individual performance, promotion typically entails a competition between various candidates in which only one gets the price. If women are less inclined to enter competitions, then they will apply less often for promotion even if costs and benefits are the same.

Existing work has had difficulty in testing these different hypotheses. Part of the problem is the fact that, in most countries, there exists no record of who applied for promotion. It is hence impossible to know whether the higher probability of men to hold a senior position is due to a lower conditional probability of success of women or to their lower propensity to apply. Moreover, even if data on applications were available, the tradeoff that exists in the Anglo-Saxon academic system between prestige of the department and promotion thresholds makes it difficult to assess the possible causes of a lower application propensity. In contrast, the French system presents a number of features that will allow us to evaluate the various options. We next turn to a description of these.

2.2 The French academic system

The French academic system has a number of features that we intend to exploit to test the above hypotheses. There are two types of academic positions in France. The most common are university positions, where the individual is a professor with a substantial teaching load. There exist also a number of public research instances, of which the largest is the Cnrs, that have pure research positions.⁴ Researchers in this category are hired by the Cnrs, who pay their salaries, but are attached to a university and are hence located in its economics department just like the university professors are. Researchers have the possibility to undertake some teaching and they participate in department life in the same way as standard professors.

For all types of position there is an entry level category equivalent to assistant professor, termed ‘Rank B’, which includes the *maître de conférences* positions at the university and the *chargé de recherche* at the Cnrs. The individual can then be promoted to ‘Rank A’, the equivalent to full professor, a position denoted *directeur de recherche* at the Cnrs.

⁴Cnrs stands for *Centre National de la Recherche Scientifique*. There are other instances, such as the *Institut National de la Recherche Agronomique*, which accounts for about 5% of the academic economists in France. Unfortunately we have not been able to obtain the list of candidates for promotion for those and hence will not include them in our analysis.

Both rank A and rank B positions are tenured, which implies that an individual who does not get a promotion can spend her/his entire career in a ‘junior’, i.e. rank B, position. Because of this possibility, we will term the two types of positions ‘rank A’ and ‘rank B’, rather than junior and senior. The promotion from rank B to rank A entails a substantial salary increase and a much steeper slope for salaries over time. The salary scales are identical for university professors and researchers and set by the Ministry of Higher Education and Research.⁵

Promotions take place through a national contest, a *concours*, and are thus not decided by the department in which the individual holds her/his current position.⁶ Participation in this contest is public information, and at the end of it a list with the ranking of those that have undertaken it is published. The fact that departments do not make promotion decisions is important for our purposes. In a system in which there is a positive correlation between prestige and promotion threshold, women could choose to select into less prestigious department because promotion is easier in those, and hence the measured gap would underestimate the actual promotion gap. A second feature of the French system is that there is no relationship between department prestige and salaries. Academic salaries are determined according to a national scale based on rank, and rank is decided at the national level.⁷ There is thus no reason to prefer being employed in a less prestigious department since salaries are the same across universities and the threshold for promotion is set nationally.

Members of the national committees that decide on promotion are academic economists, drawn from various universities in the country and areas of expertise. Committees tend to be large and represent a wide spectrum of universities, not necessarily the most prestigious. Because members of these committees have to be of the full professor rank and because of the age distribution of the population of academics, there is a strong male dominance in these committees. Committees change regularly, every two to four years depending on the particular instance.

⁵Within each broad rank, A and B, there are subranks that affect salary. Subrank promotion is also decided by a national committee, although the cost here is minimal, with application for a promotion requiring filling a form and submitting a *vitae*. Promotion to a higher subrank does not involve change of department, the number of promotions is not fixed *ex ante*, and the list of candidates is not public knowledge.

⁶Some internal promotions exist for individuals that have undertaken substantial administrative tasks at the university, but they are rare. See Combes, Linnemer and Visser (2008) for more details.

⁷Some departments pay, out of their own funds, an extra salary on top of the one paid by the university/cnrs. This practice is, however, restricted to only a few members in a handful of departments.

The requirement of the *concours* differs across the two academic tracks. For university professors the contest, termed *concours d'agregation*, is biannual and entails four stages over, approximately, a 6-month period. It includes a research seminar, and three oral exams both in the candidate's field and in economics in general, one of which consisted of preparing in 24 hours a lecture on a topic randomly drawn by the applicant from a lengthy predetermined list. The *concours* hence takes time and requires substantial preparation outside the candidates field of expertise. In contrast the effort involved in the Cnrs promotion *concours* is minimal. The candidate simply declares him/herself a candidate for a position as *directeur de recherche* (DR2) and submits a vitae and research proposal to the committee. This *concours* takes place annually.

There is a second difference in terms of the costs involved in the two systems. For university professors, departments publish before the start of the contest the positions that they have available. At the end of the *concours*, the candidates choose, sequentially and starting with the highest ranked, which university to join. When promoted, the individual is usually not able to stay in the university where s/he held a rank B position and has to move to a different department.⁸ After three years s/he is allowed to move to another university, if the latter wishes to recruit her/him, including the university where s/he held the rank B position. For researchers, academics that are promoted can choose to stay at the university where they are or move to another department. The university does not need to have an open position for them since the researchers' salary is always paid by the Cnrs.

It is obvious that promotion is very costly for university professors. There is a cost that in principle is gender-neutral, the cost of preparing the various exams. The *agregation* also involves a substantial cost ex post both because of the geographical mobility involved in being promoted to full professor, but also because the individual faces considerable uncertainty about where s/he will be eventually recruited if s/he passes the contest since where s/he ends up depends on her/his ranking and hence on the entire pool of candidates. Since candidates seeking to become full professor are typically between 30 and 40 years of age, the process occurs at a moment in the life-cycle when family constraints are likely to be substantial. If women are less geographically mobile than men, then the cost is likely to be greater for them.

⁸In our sample, 80% of those promoted through the *agregation* have a new affiliation after their promotion.

There are hence two differences between the *agregation* and the Cnrs *concours*. First, the former involves substantial costs that do not exist for the later, notably the fact that successful candidates need to change department. The second difference is that performance during the contest only matters in the *agregation*, as in the Cnrs the committee bases the decision on the *vitae* without meeting the candidates.

2.3 Testing the different hypotheses

We can use the features of the French academic system to confront the various hypotheses about lower promotion rates for women. In order to do so, we will examine both the determinants of the likelihood to apply for promotion and of the probability of obtaining a promotion conditional on being a candidate. We start by considering the gender gap in the probability of success conditional on being a candidate for promotion, and examine two possible explanations for a difference across the sexes:

Under-performance in contests: If women tend to perform less well than men in contests, we expect to find a negative impact of being female on the conditional probability of promotion in the *agregation*. The absence of a gender gap in this probability can then be interpreted as no difference in the performance of women and men under competitive conditions.

Discrimination: Identifying discrimination is difficult. On the one hand, a negative impact of being female on the conditional probability of promotion can also be due to the presence of discrimination, and there is no feature of the data that would allow us to distinguish the effect of underperformance from that of discrimination. On the other, the absence of a gender effect could be the result of offsetting forces, with discrimination against women being compensated by better unobserved characteristics of women (not correlated to observed characteristics). As a result, it is not possible to rule out discrimination even in the absence of a significant gender effect.

Turning to the probability of being a candidate for promotion, we can test for two potential explanations for differences at this level by comparing the two tracks in the French academic system:

Higher cost of promotion: In the US system, it is not clear why women may have different costs of promotion from men. The nature of academia and the substantial flexibility associated with this type of jobs implies that jobs at the junior or senior level

are equally likely to be compatible with family life. As a result, there is no reason why women would prefer not to be promoted. In contrast, in the French system, the mobility required if promoted (of professors but not researchers) implies that there is a very specific cost: the need to move to another university and hence, in most cases, to another city. It is possible that this cost is, on average, higher for women than for men, implying that the latter are less likely to apply for promotion. To test this hypothesis we can exploit the difference between the implications of promotion for standard professors and for researchers. If women were not trying to get promoted because of the cost of mobility, we would expect to find a more negative effect of being female on the probability of applying for promotion for professors than for researchers.

Differences in preferences about prestige and salary: We can also exploit the difference between promotion implications for professors and researchers to test the hypothesis of different preferences over salary and prestige across genders. Since for university professors being promoted to a senior position requires changing department, promotion brings a higher salary at the risk of having to move to a less good department. If women are less willing to trade-off prestige for income, then those who hold a junior position in a top department may not want to apply for promotion. In contrast, for junior researchers promotion does not require a change of department, and there is no reason for women not to seek promotion even if they are in a top department. If women have a stronger preference for department prestige and/or a lower one for income, those with a university position will be less willing to apply for promotion if they are in a top department than if they are not. Because promotion does not require changing department for researchers, we would expect to find no differences in the probability of applying for promotion for female researchers across departments.

Alternatively, the impact of gender on the probability of promotion may be common across the two tracks, implying that the differences in the propensity to seek promotion are due to either beliefs or attitudes towards contests. Testing for these effects is difficult. However, the conditions under which the aggregation takes place can provide proxies for both a measure of expected discrimination and the competitive environment. We consider these in turns.

Expected discrimination: To examine whether women's belief that they will be discriminated against affects their propensity to seek promotion, we proxy expected discrim-

ination by the fraction of rank A professors that are women in a particular year, as a greater fraction could be interpreted by candidates as weaker discrimination. A positive effect of this variable interacted with being female would indicate that women’s decision to seek promotion responds to the observed promotion rate of women in the past.

Unwillingness to participate in contests: Women may also have different attitudes from men concerning contests, and, in particular, be unwilling to take part in them. If this is the case, they are likely not to respond to changes in the payoff from being a candidate. We hence consider the effect of changes in the likelihood of promotion, which we measure by the number of available positions of rank A in a particular year. A greater number of positions should increase a potential candidate’s expected payoff and hence make him/her more likely to enter the competition. A positive coefficient on this variable hence indicates that candidates respond to this ‘objective’ measure of the likelihood of promotion, while no effect implies that individuals’ decision to be a candidate is determined by considerations other than the probability of promotion, which in the case of women could be an unwillingness to enter competitions.⁹

2.4 Empirical specification

In line with the existing literature, we suppose that the probability of individual i being rank R at year t is given by

$$\Pr(R, i, t) = \Phi(X'_{it}\beta_j) \quad (1)$$

where the two states R are being rank A or not, the subscript $j = u, r$ denotes whether we are looking at data for university professors or researchers, and $\Phi(\cdot)$ denotes the logistic density function. The term $X'_{it}\beta$ in equation (1) is

$$\begin{aligned} X'_{it}\beta_j = & \beta_{j0} + \beta_{j0}^f \delta_i + \beta_{j1} Age_{it} + \beta_{j2} Age_{it}^2 + \beta_{j3} Pub_{it} \\ & + \beta_{j4} Pub_{it} \times Quantity_{it} + \beta_{j5} Pub_{it} \times Quality_{it} \end{aligned} \quad (2)$$

⁹A question that we have not addressed is whether the degree of competition is greater for university professors or for researchers. On the one hand, a much smaller number of researchers are promoted each year in the Cnrs than in the *concours d’agregation*. On the other, although the list of candidates and rankings are available on internet for both, the *agregation* solicits much greater interest from the academic community, with results at each stage being widely followed and discussed, which increases the pressure on candidates. It is hence difficult to argue that one or the other contest is more competitive.

implying that the probability of promotion is a function of age Age_{it} and its square, whether or not the individual has published in Econlit-classed journals (i.e. whether s/he is a ‘publisher’ measured by the dummy Pub_{it}), the number of publications and the average quality of these publications, denoted respectively $Quantity_{it}$ and $Quality_{it}$, both measured in logs (see below for the exact measurement). The variable δ_i is a dummy taking the value 1 for females, so that β_0^f measures the differences in promotion probability for men and women with the same characteristics. We would have liked to control for characteristics of the individual’s family life, such as whether s/he is married and the number and ages of children, but such data were not available.¹⁰

The probability of holding a rank A position is obviously a combination of the probability of applying for promotion and that of obtaining a promotion conditional on having applied. We hence model separately the two process. Consider first the probability that individual i applies for promotion (i.e. takes the *concours*) at time t , which we assume to be

$$\Pr(C, i, t) = \Phi(X'_{it}\gamma_j) \quad (3)$$

where the two states C are being a candidate for promotion or not and

$$\begin{aligned} X'_{it}\gamma_j &= \gamma_{j0} + \gamma_{j0}^f \delta_i + \gamma_{j1} Age_{it} + \gamma_{j2} Age_{it}^2 + \gamma_{j3} Pub_{it} \\ &+ \gamma_{j4} Pub_{it} \times Quantity_{it} + \gamma_{j5} Pub_{it} \times Quality_{it} \\ &+ \gamma_{j6} Dept_{it} \end{aligned} \quad (4)$$

The dummy $Dept_{it}$ has a value of 1 if the candidate is in a top department (see below for the definition) and will capture whether individuals in top departments are less likely to apply for promotion because of the mobility costs involved.

Lastly, the probability of success in a *concours* conditional on being a candidate is given by

$$\Pr(S, i, t | C_{it} = 1) = \Phi(X'_{it}\alpha_j) \quad (5)$$

where the two states S are succeeding or failing in the *concours* (i.e. being promoted or

¹⁰In the light of existing evidence, this is likely not to be a major omission; for example, Ginther and Kahn (2004) find that having children has only a weak effect of the promotion probabilities of female economists in the US and none on their productivity.

not), C_{it} is a dummy that takes the value one if the individual applied for promotion, and

$$\begin{aligned} X'_{it}\alpha_j = & \alpha_{j0} + \alpha_{j0}^f \delta_i + \alpha_{j1} Age_{it} + \alpha_{j2} Age_{it}^2 + \alpha_{j3} Pub_{it} \\ & + \alpha_{j4} Pub_{it} \times Quantity_{it} + \alpha_{j5} Pub_{it} \times Quality_{it} \\ & + \alpha_{j6} Dept_{it} \end{aligned} \quad (6)$$

Obviously, the qualities that lead to promotion conditional on applying are also those that make a potential candidate apply, hence it is not possible to run a selection model.

Equations (1) and (2) capture the mechanism usually examined in the literature on academic promotions. Assuming that the estimated coefficient $\hat{\beta}_{j0}^f$ is different from zero, we can then estimate the two other models to see to what extent the difference in the probability of being rank A is due to women being less likely to enter the *concours* or to them having a lower success rate in the *concours* than men. That is, we will compare estimates of γ_{j0}^f and α_{j0}^f .

3 The Data

Our sample consists of the entire population of French academic economists provided by the French Ministry of Higher Education and Research and the Cnrs for the years 1991 to 2008. For each individual we have information on age, rank, publication stock (see below) and department. We keep only individuals that are in departments larger than 4 full-time equivalent academics, which removes economists that are isolated in universities without real economics departments. Those (few) individuals for whom some of the individual characteristics (age or position for instance) are missing are also excluded.

In order to examine separately the determinants of the probability of applying for promotion and the likelihood to succeed in the *concours*, we use the list of candidates that applied to and those who succeeded in becoming a rank A university professor or a rank A Cnrs researcher (DR2). For professors we have the lists of candidates to promotion and actual promotions for the nine biannual *concours* taking place from 1992 to 2008, while for researchers we have the six annual *concours* between 2003 and 2008.

We define as potential candidates for promotion all academic economists of rank B the year prior to the *concours*. It is important to note that it is possible for Cnrs researchers to take the *agregation*, and some individuals in our dataset do so. In contrast, although it is in principle possible for rank B university professors to apply for a rank A Cnrs position, there are in our data no such individuals. This is probably due to the fact that few promotions are available each year (between 2 and 4) and they are perceived as being internal promotions for those already in the Cnrs. We hence consider as potential candidates for the *agregation* all academic economists in France of rank B, which yields 10,383 observations, while we have 681 observations for candidates susceptible to apply for Cnrs promotions.¹¹

We define two categories of department, somewhat equivalent to the division in the US between the top-50 and other departments. France has a substantial number of national academic publications in French, and hence we define prestigious departments as those that have the largest research output in international journals (see Appendix A for details). For this reason, we will term them ‘international’ departments and the rest ‘national’ departments. The international departments account for about one third of the population each year.

3.1 Measuring research output

Our key explanatory variable is an individual’s research output, and we measure the output of individual i at date t by her/his cumulative publication record between the first year for which we have an observation for that individual and date t .¹² We merge the data on publications with the list of French academics that includes individual characteristics (position, department, age, gender), merging by surname and initial, and then correct manually for those individuals having the same name and initial. Publication records are measured as weighted sums of publications. All publications come from the EconLit database, which includes more than 560,000 papers published in more than 1200 journals between 1969 and 2008. Three dimensions enter the weighted scheme of publications:

¹¹The difference in the number of observations is mainly due to the fact that the Cnrs accounts for a small fraction of academics. We have, on average, 114 rank B researchers each year and 1,040 rank B university professors. Hence, there are each year 114 potential candidates for promotion at the Cnrs and 1,154 potential candidates to promotion through the *agregation*, which include both rank B professors and rank B researchers.

¹²As an alternative, we have computed degressive publication scores, with older publications having a smaller weight than recent ones. Measuring research output this way does not change our results.

the quality of journals, the number of authors and the relative number of pages.

We measure the quality of publications using the journals weighting scheme proposed by Combes and Linnemer (2010). Two different degrees of convexity in the distribution of journals' weights have been proposed by Combes and Linnemer, and we use the most convex one (i.e. the one that most values quality), but our results are unchanged when we use the least convex one. We also consider the number of authors of each publication, and assume constant returns to scale at the coauthor level, so that we divide each publication by the number of authors, a standard practice in the literature. Lastly, since many journals now publish both notes and longer papers, we weight by the number of pages to capture the idea that longer articles contain more ideas. We consider an article's length relative to the average length in that journal, so that the ratio of the number of pages of article a over the average number of pages of articles published in the journal is used as weight. Because editors' policies can vary over time, the mean number of pages is calculated for each year separately.

The output of a researcher i at date t is then a weighted sum of her/his articles a published between the first year in which s/he published an article, t_0 , and date t , implying that

$$y_{it} = \sum_{a \in [t_0, t]} \frac{W(a)}{n(a)} \frac{p(a)}{\bar{p}} \quad (7)$$

where $p(a)$ is the number of pages of article a , $\bar{p}$ is the annual average number of pages of articles in the journal, $n(a)$ the number of authors of the article, and $W(a)$ the weighting scheme for journals. Then, each individual receives three scores: a dummy equal to 1 if s/he has at least on publication in an EconLit-listed journal, the quantity of single-author-equivalent published articles, and the average quality of her/his articles, defined as y_{it} divided by quantity, *i.e.* $\sum_{a \in [t_0, t]} 1/n(a)$.

3.2 Descriptive statistics

Table 1 gives the decomposition of our sample in terms of institutional affiliation and rank for the most recent year in our data, 2008. There were 2,105 (Cnrs and university) academic economists in France that year, and the vast majority of the population consists of university professors, with researchers accounting for only 9% of the total. Women account for 29% of observations, and they are over-represented amongst univer-

Table 1: Percentage rank A by gender in 2008

rank A	Total	%	Women	%	Men	%
Total sample						
0	1425	67.7	512	83.0	913	61.4
1	680	32.3	105	17.0	575	38.6
University professors (91% of population in 2008)						
0	1321	69.0	477	83.5	844	62.8
1	593	31.0	94	16.5	499	37.2
Cnrs Researchers (9% of population in 2008)						
0	104	54.5	35	76.1	69	47.6
1	87	45.5	11	23.9	76	52.4

sity professors and under-represented amongst researchers, where they account for only 24% of the population. This difference could be due to the fact that obtaining a position as a researcher tends to require a stronger publication record than for university positions and, as we will see below, women tend to have a weaker research output than men.

Slightly under a third of the population hold a rank A position, with the fraction being lower for university professors (31%) and higher for researchers (45.5%). Note that the data do not seem to indicate that women choose a career path that offers higher average promotion rates, which would counterbalance negative discrimination. Our data shows that feminisation is lower for researchers, which have a higher promotion rate, indicating that there is no selection of this type taking place. The gender promotion gap is large, 22 percentage points on average, and is smaller for university professors than for researchers (21 and 29 points, respectively).

Table 2 reports some descriptive statistics for the sample of potential candidates. We have 10,383 observations, including a total of 2,136 academics. The panel is unbalanced as individuals enter the pool of potential candidates (usually when they get their PhD) and exit it either because they are promoted or because they leave academia or the country over our sample period, with the average number of observations per individual being 4.9 years. The probability of being a candidate to the *aggregation* is 6% for women and 8% for men, while for researchers these figures are 10% and 19%, implying a much larger gap. The (unconditional) probabilities of being promoted on a given year are small, 3% to 4% for men, and 2% for women. The table indicates that some of these differences are

Table 2: Descriptive statistics of potential candidates

	Univ', panel 1992-2008				Cnrs, panel 2003-2008			
	Min.	Max.	Mean	Std. err.	Min.	Max.	Mean	St. err.
Women								
Prob. Candidate	0	1	0.062	0.241	0	1	0.103	0.305
Prob. Promotion	0	1	0.020	0.140	0	1	0.018	0.133
Age	27	68	41.6	8.9	29	64	42.8	9.1
Publisher	0	1	0.55	0.50	0	1	0.83	0.37
Quantity	0	15.0	1.02	1.55	0	15.0	3.14	3.22
Quality	0	62.5	0.42	2.83	0	16.6	1.01	2.31
Prob. Int. Dept.	0	1	0.32	0.47	0	1	0.58	0.49
Men								
Prob. Candidate	0	1	0.084	0.277	0	1	0.186	0.389
Prob. Promotion	0	1	0.033	0.179	0	1	0.044	0.205
Age	25	70	45.2	9.4	28	65	43.5	10.0
Publisher	0	1	0.57	0.50	0	1	0.91	0.29
Quantity	0	21.2	1.33	2.05	0	16.6	3.70	3.03
Quality	0	116.8	0.46	3.18	0	86.5	2.42	6.27
Prob. Int. Dept.	0	1	0.28	0.45	0	1	0.58	0.49

Int. Dept. = international department; see Appendix A for details. Productivity measures (quantity and quality) are in levels. We take their logs in the regression analysis. There are 10,383 observations from 2,136 individuals and 681 observations from 158 individual in the university and Cnrs samples respectively. These correspond to 687 women (3,328 observations) and 1,449 men (7,055 observations) in the university sample and 54 women (223 observations) and 104 men (458 observations) in the Cnrs sample.

likely to be explained by differences in observable characteristics. Women are on average 3 years younger than men in the aggregation sample (though not in the Cnrs). The main difference lies in publication records. In our larger sample, the probability of publishing in EconLit journals is 57% and 55% for men and women, respectively. This figure is not large, but it is important to bear in mind that our publication criteria is stringent, especially given the strong tradition in France to publish books and the large number of national journals, some of which are not in EconLit. The quality of publications is somewhat higher for men (0.46 compared to 0.42), while the quantity of publications is 30% higher for men. For the sample of researchers, we find a higher probability of being a publisher but a greater gender gap (91% for men but only 83% for women), and a particularly large difference in the quality of publications.¹³

¹³Obviously some of the differences in research output are due to the age structure of the two populations, but Bosquet and Combes (2013b) show that even when controlling for age women have worse publication records. Women are sometimes also found to be cited less than men. This is the case in political sciences, as shown by Maliniak, Powers and Walter (2013), while for publications in biomedical and exact sciences, Kelchtermans and Veugelers (2013) find that women have a lower probability of reaching top citations but once they do they are as likely as men to stay there. In contrast, the data we

The definition and construction of our two categories of departments is detailed in Appendix A and Table 10 gives the list of ‘international departments’. The last lines in the two panels of Table 2 indicate that the one aspect in which women seem to fare better than men is affiliation: in the larger sample 32% of women are in international departments, while only 28% of men are. When we focus only on researchers we find no difference in affiliation across the genders.

4 Results

4.1 The promotion of academic economists

We start by examining the determinants of being rank A for the entire population, which are reported in Table 3. In order to run regressions equivalent to those found in the literature on promotions in academia, where only outcomes are observed, we construct a sample that includes all rank A and rank B academics for each of the years that we will be using latter on to estimate the probability of being a candidate and of being promoted conditional of being a candidate. This gives us a sample of 16,077 observations. We can thus estimate the probability of *holding* a rank A position. All specifications include year fixed effects. In column (1), only time fixed effects and gender are included in the logit model, with the gender dummy being equal to 1 for women. The marginal effect on gender is significant at the 1% level and large, implying an odds ratio of 0.31.¹⁴ Including age reduces the effect on gender from -0.233 to -0.165, indicating that a large fraction of the difference in promotion is indeed due to the fact that the sample of women is younger than that of men. Column (3) includes our three measures of research output: whether or not the individual publishes, the quantity of publications and their quality. All three are highly significant and increase the probability of promotion.

Once we control for research output, the effect of gender falls to -0.053, about a quarter of the initial one, indicating that the lower promotion rate for women is to a large extent due to them having published less. The effect is nevertheless still strong: being a woman reduces the probability of promotion almost as much as having one rather than two

use exhibits no gender gap in citations once research output is controlled for; see Bosquet and Combes (2013a).

¹⁴This value is obtained from $\exp(-1.182)$, -1.182 being the coefficient associated to the -0.233 marginal effect.

Table 3: Likelihood to hold a rank A position: panel 1992-2008, marginal effects

	(1) logit	(2) logit	(3) logit	(4) logit	(5) probit	(6) OLS
Women	-0.233 ^a (0.006)	-0.165 ^a (0.007)	-0.053 ^a (0.007)	-0.049 ^a (0.007)	-0.048 ^a (0.007)	-0.044 ^a (0.007)
Age		0.015 ^a (0.002)	0.010 ^a (0.001)	0.010 ^a (0.001)	0.010 ^a (0.001)	0.007 ^a (0.001)
Age ²		-0.000 (0.000)	0.000 ^a (0.000)	0.000 ^a (0.000)	0.000 ^a (0.000)	0.000 ^a (0.000)
Publisher (Pub)			0.325 ^a (0.012)	0.331 ^a (0.012)	0.329 ^a (0.012)	0.341 ^a (0.010)
Pub*Quantity			0.148 ^a (0.004)	0.151 ^a (0.004)	0.152 ^a (0.004)	0.178 ^a (0.004)
Pub*Quality			0.040 ^a (0.002)	0.041 ^a (0.002)	0.041 ^a (0.002)	0.038 ^a (0.002)
Cnrs				-0.089 ^a (0.014)	-0.081 ^a (0.014)	-0.093 ^a (0.017)
Women*Cnrs				-0.042 ^c (0.025)	-0.039 (0.025)	-0.038 (0.026)
Time FE	Yes	Yes	Yes	Yes	Yes	Yes
pseudo-R ²	0.040	0.114	0.367	0.369	0.368	
Observations	16077	16077	16077	16077	16077	16077
log-likelihood	-9963	-9195	-6573	-6548	-6559	-6833

Standard error between brackets. ^a, ^b, ^c Significant at the 1%, 5% and 10% level, respectively.

single-authored publications. The effects we obtain are comparable to those found for the US. Ginther and Kahn (2004) find a raw gender gap of -0.213, which falls to -0.130 once age and publication records are included, indicating that differences in publications across genders explain a greater fraction of the gap in France than in the US.

Column (4) includes a dummy for being a researcher and its interaction with Women, and indicates that the negative effect of being a woman on seniority is stronger for the Cnrs, although this term is significant only at the 10% level. Columns (5) and (6) report the estimations of a probit and an OLS model, respectively, and indicate that the result that women are less likely to be promoted are robust to the use of those specifications. There is however no robust evidence of a difference between the two tracks, with the coefficient on Women*Cnrs being insignificant.

The result that women have lower promotion rates than men contrasts with evidence on affiliation. Table 11 in Appendix B examines the likelihood for an individual to be in an international department, using the same explanatory variables as above. Women are more likely to be in a prestigious department than men, and the effect of being

a woman on the probability of being in an international department is large.¹⁵ These results are surprising. As argued by Lazear and Rosen (1990), it is hard to understand why women would be discriminated against in promotions but not in other labour market experiences. In fact, in many instances the evidence tends to indicate that women have a lower probability of both being promoted and being hired, as for example in the case of top US orchestras; see Goldin and Rouse (2000). This does not seem to be the case for academic economists, and raises the question of whether the low promotion rate of female academics has a cause other than discrimination.

4.2 Decomposing outcomes

4.2.1 Gender differences in success in the *concours*

The limitation of our analysis so far is that it uses only outcomes and hence we do not know whether lower observed promotions are the result of a lower likelihood to apply for promotion or lower success in obtaining the promotions. We hence examine the two separate steps. We start by considering what determines success in the contest, conditional on being a candidate, and then move to the determinants of the decision to enter the contest.

The determinants of success in the promotion contests are assumed to be gender and the three variables measuring publications. We also include a dummy that captures whether the individual was, at the time of application, in an international department. This variable could measure unobserved ability –e.g. being a good teacher– or the positive effects that being in a more stimulating academic environment could have on the preparation of the *concours*. There are as well externalities due to having colleagues who are also preparing the *agregation* since candidates often work together and share the burden of preparing lectures on the various topics. Lastly, in the regression for the *agregation* we include a dummy for whether the candidate holds a position other than rank B university professor in economics (i.e. the candidate is in the Cnrs, in another field such as mathematics, comes from abroad, etc.).

¹⁵The positive effect of being female on the likelihood of being in an international department may have three causes. The first is positive discrimination when departments consciously try to increase their female faculty. A second is unobservable abilities, such as organisational skills or teaching ability, which could be, on average, higher amongst female than male academics, specially since fewer women than men that enter academia (see Petrongolo and Olivetti (2008) on selection and gender). Lastly, a strong female presence could be the result of joint offers made to couples by top departments.

Table 4: Likelihood to be promoted conditional on applying, marginal effects

	Agreg' 1992-2008		Cnrs 2003-2008		Agreg': selected candidates 1992-2008			
	All candidates		All candidates		"Admissibles"		Close to threshold	
	(1) logit	(2) logit	(3) logit	(4) logit	(5) logit	(6) logit	(7) logit	(8) logit
Women	-0.007 (0.032)	-0.018 (0.028)	-0.051 (0.078)	0.044 (0.110)	0.044 (0.056)	0.040 (0.057)	0.097 (0.082)	0.103 (0.086)
Age		-0.018 (0.011)		-0.071 ^b (0.031)		-0.033 (0.026)		-0.042 (0.039)
Age ²		-0.000 (0.000)		0.002 ^b (0.001)		0.001 (0.001)		0.002 (0.002)
Publisher (Pub)		0.155 ^a (0.047)		0.077 (0.209)		0.112 (0.086)		0.142 (0.130)
Pub*Quantity		0.091 ^a (0.019)		-0.009 (0.065)		0.032 (0.039)		-0.024 (0.060)
Pub*Quality		0.047 ^a (0.006)		0.046 ^b (0.022)		0.031 ^b (0.014)		0.024 (0.022)
Int. Dept.		0.090 ^a (0.030)		0.019 (0.088)		0.130 ^a (0.046)		0.117 (0.080)
Other		-0.030 (0.032)				0.008 (0.069)		0.008 (0.108)
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
pseudo-R ²	0.015	0.183	0.020	0.184	0.015	0.056	0.005	0.025
Observations	970	970	115	115	336	336	198	198
log-likelihood	-531	-441	-52	-43	-201	-192	-137	-134

Standard error between brackets. ^a, ^b, ^c Significant at the 1%, 5% and 10% level, respectively. Int. Dept. = international department; see Appendix A for details. Other: non-French Assistant Professors positions in economics, as well as Cnrs and Inra researchers, and assistant professors from other disciplines.

The results, reported in the first four columns of Table 4, indicate that research output is the key determinant of the probability of success both in the *agregation* and the Cnrs *concours*. In the former all three measures of research output have a significant coefficient, while for the Cnrs promotion only quality matters. This is consistent with the fact that there is a stronger selection for those joining the Cnrs than for those following the standard university track, implying that the former are more homogeneous in terms of research output. Being in an international department has a positive impact on the probability of passing the *agregation*, in line with our arguments before, while it has no impact on the probability of being promoted in the Cnrs contest.

Turning to our coefficient of interest, the impact of being female on the probability of success conditional on being a candidate for promotion, we find no significant difference across the genders. In fact, the probability of success in the *agregation* is 24.8% for men

and 24.1% for women, and hence not significantly different. The coefficient is insignificant both for the raw probabilities and also when we control for individual characteristics, for professors and researchers. Gender simply does not matter.

To interpret this result consider first the Cnrs *concours*. Recall that the *concours* consists of submitting a vitae and research proposal. The candidate does not ‘perform’ in front of the committee taking the promotion decision, hence the only reason why we would expect to find a negative coefficient on women would be discrimination. We can then interpret the insignificant coefficient as a lack of evidence of *ex post* discrimination against women in the contest. This does not mean that discrimination does not matter, since expectations of discrimination may have an *ex ante* effect on women’s choices and hence on observed outcomes, as discussed above.

In the case of the *agregation*, interpreting the coefficient is more complicated as it captures both potential biases in the committee and the actual performance of the candidate during the oral exams that we cannot observe. A negative and significant coefficient could then be due to either discrimination or under-performance by women, as in the experiments performed by Gneezy et al. (2003). Our results find no evidence of either discrimination against or under-performance by women. It is also possible that discrimination is offset by over-performance or that, more generally, selection is an issue. Since fewer women are candidates than men, as we will see below, it is possible that they are drawn from the top of the distribution and hence that their unobserved ability is higher than that of men. The insignificant coefficient could then be the result of opposing effects canceling out: higher unobserved ability of women and discrimination against them.

To try to control for this, we perform two further tests exploiting the fact that for the *agregation* contest we have information on which individuals passed each stage of the competition, as well as the final rank. At the various stages of the competition, a number of candidates are eliminated, and after the penultimate stage a list of candidates that are *admissibles* is provided by the jury. These candidates then undertake the final test, after which a ranking of all admissible candidates is provided and, if p positions are available that year, the top p candidates are promoted.

The first test focuses on admissible candidates only. Since at this stage a substantial number of candidates have been eliminated, if selection implies that the unobserved quality of male candidates is lower than that of female candidates, focusing on the ‘best’

should eliminate the (predominantly male) bottom tail of the distribution. Columns (5) and (6) of Table 4 present those regressions and indicate that there is no difference between men and women in either the raw rate of passing the last step of the context or the one obtained after controlling for output. Not surprisingly, only the quality of publications matters at this stage. To make our sample even more comparable, we construct from the final ranking of candidates a list of individuals close to the threshold. To do so we take, for each year, the n admissible candidates that did not get promoted and the n lowest-ranked candidates that did get promoted and rerun the same regression equations. If discrimination against women were taking place, it would be likely to appear in this reduced sample of candidates with similar observable (and probably close unobservable) characteristics. The last two columns of Table 4 show that, if anything, women are *positively* discriminated although the coefficient is not significant. In fact, none of the variables we consider at this stage are significant.

4.2.2 Likelihood to enter the *concours*

We turn now to the determinants of the likelihood to enter the *concours*. The results are reported in Table 5. The unconditional probability of applying is lower for women than for men both for university professors and researchers (column 1), and the negative impact of being female is even larger once we include individual characteristics, the marginal effect being 0.025 for professors and 0.092 for researchers. As expected, research output has a strong effect on the likelihood of being a candidate in either of the *concours*, with quality having a stronger effect for researchers. Being in an international department is significant only in the case of professors, where it tends to increase the probability of being a candidate. This may be due to the fact that, as we saw earlier, being in a top department has a positive effect on the probability of success; potential candidates may anticipate this and hence be more likely to apply if they are in those departments. An additional effect may come from peer pressure to pass the *concours*. Removing this variable from the analysis has no impact on the other estimated coefficients.

The results indicate that women have a lower likelihood to enter the contest for promotion. In terms of magnitude, being a woman is equivalent to decreasing the number of single-authored publications by around 0.6 for the university professors and 2 for the

Table 5: Likelihood to apply for a promotion: marginal effects

	Agreg', panel 1992-2008				Cnrs, panel 2003-2008			
	(1) logit	(2) logit	(3) probit	(4) OLS	(5) logit	(6) logit	(7) probit	(8) OLS
Women	-0.021 ^a (0.005)	-0.025 ^a (0.004)	-0.027 ^a (0.004)	-0.033 ^a (0.005)	-0.085 ^a (0.022)	-0.092 ^a (0.021)	-0.087 ^a (0.022)	-0.073 ^a (0.028)
Age		0.001 (0.002)	-0.001 (0.002)	-0.016 ^a (0.001)		0.100 ^a (0.011)	0.097 ^a (0.010)	0.064 ^a (0.007)
Age ²		-0.000 ^a (0.000)	-0.000 ^a (0.000)	0.000 ^a (0.000)		-0.002 ^a (0.000)	-0.002 ^a (0.000)	-0.001 ^a (0.000)
Publisher (Pub)		0.094 ^a (0.011)	0.094 ^a (0.011)	0.119 ^a (0.009)		0.033 (0.054)	0.029 (0.051)	0.075 (0.047)
Pub*Quantity		0.043 ^a (0.004)	0.043 ^a (0.004)	0.050 ^a (0.004)		0.050 ^b (0.020)	0.042 ^b (0.018)	0.044 ^b (0.018)
Pub*Quality		0.010 ^a (0.001)	0.010 ^a (0.002)	0.016 ^a (0.002)		0.040 ^a (0.007)	0.040 ^a (0.007)	0.036 ^a (0.007)
Int. Dept.		0.013 ^b (0.006)	0.012 ^b (0.006)	0.014 ^b (0.006)		-0.011 (0.028)	-0.006 (0.028)	-0.008 (0.029)
Cnrs		-0.087 ^a (0.002)	-0.086 ^a (0.002)	-0.130 ^a (0.009)				
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
pseudo-R ²	0.011	0.262	0.258		0.030	0.277	0.274	
Observations	10383	10383	10383	10383	681	681	681	681
log-likelihood	-2779	-2075	-2087	-297	-289	-215	-216	-213

Standard error between brackets. ^a, ^b, ^c Significant at the 1%, 5% and 10% level, respectively. Int. Dept. = international department; see Appendix A for details.

Cnrs researchers,¹⁶ or to decreasing the quality of publications by 87% and 81% of its standard deviation, for the university professors and the Cnrs researchers, respectively.¹⁷

In order to try to understand what lies behind the gender gap in seeking promotion, we examine whether the effect of gender differs between the two types of positions. Since the costs of the *agregation* contest are substantially larger than those of applying for promotion at the Cnrs, the coefficient on women should be higher for university professors than for researchers if differences across the genders in these costs were holding back potential female candidates. The first column of Table 6 hence runs a difference-in-differences regression where we have pooled together the data for the two *concours*. We interact all variables with a dummy taking the value 1 for the Cnrs *concours* to allow for different impacts across the two tracks, our coefficient of interest being women interacted

¹⁶The regression coefficients on gender and quantity in the regressions for the likelihood to apply for the *agregation* and to apply for a Cnrs promotion are, respectively, -0.455 , 0.740 , -1.007 and 0.501 .

¹⁷The coefficients on gender and quality in the *agregation* and the Cnrs promotion regressions are, respectively, -0.455 , 0.170 , -1.007 and 0.406 , and 3.07 is the standard deviation of the average quality of publications in our sample.

with this dummy.¹⁸ The coefficient is negative and significant at the 10% level, implying that the effect of gender on the likelihood to enter the promotion contest is twice as high in the research track. This result indicates that the time-cost of preparing the *agregation* and the implications in terms of mobility are not the main reason why women are less likely than men to apply for promotion.

As we discussed earlier, if women have a stronger relative preference for department prestige, they may choose not to apply for promotion in order to stay in an international department. Since promotion implies mobility for university professors, if females cared more about department quality they would be less willing to move –and hence to pass the *agregation*– whenever they are in an international than in a national department. In contrast, for researchers promotion does not require changing department and we expect to find no effect of department of origin on the likelihood of seeking a promotion.

We hence run again the regressions for the likelihood to apply for promotion and include an interaction between being a woman and being in a top department. If the tradeoff between rank (i.e. income) and department quality differed across the genders, this would be captured by a negative coefficient on being in a top department for female professors but not for female researchers.

The marginal effects obtained from these regressions are reported in the second and third columns of Table 6. Being in an international department does not affect the likelihood of female researchers applying for promotion, while it has a positive effect on whether females enter the *agregation*. For male university professors the department of origin seems not to matter. This indicates, first, that the effect on being in an international department we obtained earlier is driven by the effect on women in top departments. Second, note that when we consider together the two effects, that of women and that of women interacted with international department they cancel out. This seems to indicate that, for university professors, women in top departments are as likely as men to enter the *concours*. In contrast, women in national departments and those with a research position have a lower probability of being candidates for promotion.

The evidence so far indicates that there is a substantial difference in the likelihood that men and women academics apply for promotion. This gap appears both for professors and researchers despite the differences in the costs and implications of promotion, indicating

¹⁸To save space, other interacted terms are not reported but are available upon request.

Table 6: Differences across tracks and departments: likelihood to apply for a promotion, marginal effects

	All (1) logit	Agreg', panel 1992-2008 (2) logit	Cnrs, panel 2003-2008 (3) logit
Women	-0.024 ^a (0.005)	-0.035 ^a (0.005)	-0.085 ^a (0.031)
Age	0.004 ^c (0.002)	0.001 (0.002)	0.100 ^a (0.011)
Age ²	-0.000 ^a (0.000)	-0.000 ^a (0.000)	-0.002 ^a (0.000)
Publisher (Pub)	0.102 ^a (0.012)	0.094 ^a (0.011)	0.031 (0.054)
Pub*Quantity	0.046 ^a (0.004)	0.043 ^a (0.004)	0.050 ^b (0.020)
Pub*Quality	0.011 ^a (0.001)	0.010 ^a (0.001)	0.040 ^a (0.007)
Int. Dept.	0.015 ^b (0.006)	0.004 (0.006)	-0.006 (0.032)
DR2 Cnrs*Women	-0.025 ^c (0.013)		
Women*Int. Dept.		0.031 ^b (0.014)	-0.016 (0.055)
Cnrs		-0.087 ^a (0.002)	
Interacted terms	Yes	No	No
Time FE	Yes	Yes	Yes
pseudo-R ²	0.260	0.263	0.277
Observations	11064	10383	681
log-likelihood	-2318	-2072	-215

Standard error between brackets. ^a, ^b, ^c Significant at the 1%, 5% and 10% level, respectively. Int. Dept. = international department; see Appendix A for details. Interacted terms: all variables interacted with applying to the DR2 Cnrs contest.

that neither the direct costs associated with the *concours* nor differences in the way in which men and women trade-off income and the quality of their department are the causes of the gender gap in seeking promotion.

In order to gauge the importance of women's lower propensity to apply for promotion, we perform a Oaxaca decomposition. Given the difficulty of interpreting decomposition stemming from logit regressions, we perform the decomposition using the OLS regression in Table 5 column (4). The raw gender gap in the probability of being promoted to rank A in our sample amounts to 1.2 percentage points. Although this magnitude seems small, the probability of being promoted in a given year is only 3.3% for men, implying that being a woman reduces it by one third. As we have seen, this difference is driven by the decision to be a candidate, with an average male probability of being a candidate of 8.4%

and that for women being 2.2 percentage points lower, i.e. gender reduces the likelihood to apply for promotion by one quarter.

Table 7: Oaxaca decomposition of the likelihood to apply: Univ. only, OLS

	% of $\Delta \Pr(C)$ explained by
<u>Characteristics</u>	-86.8
Age	-115.1
Total outcome	39.0
<i>Publisher</i>	13.1
<i>Quantity</i>	28.9
<i>Quality</i>	-3.0
International Department	-3.5
Cnrs	-7.2
<u>Coefficient</u>	186.8
<u>Total gap</u>	100.0

Table 7 presents the Oaxaca decomposition for the probability of being a candidate, $\Pr(C)$, based on column (4) of Table 5, where the only coefficient that differs across the genders is the constant. The gender gap in the probability of being a candidate is driven by the direct effect of gender, which is partly offset by the fact that females have better characteristics in terms of age and affiliation. The fact of being younger has the strongest impact, with a magnitude slightly higher than that of the overall gap observed. Being in an international department also favors women, though to a much smaller extent. In contrast, their lower research output has a negative impact on their probability of being a candidate which amounts to 39% of the overall effect.¹⁹ The effect of research output is however dwarfed by the direct impact of being female, which is almost five times larger. Without the advantage in characteristics, the gender gap would be almost twice as large as the one we observe, as indicated by the fact that the coefficient on women explains 187% of the overall gap. That is, the direct effect of being female implies that the probability of applying for promotion for women is about half that of men, although this effect is partly offset by women being younger and in better departments.

¹⁹Interestingly, when we look at the three components of research output we see that women are less likely to be publishers and, especially, have fewer publications, but they exhibit a higher quality of publications than men.

4.3 Networks and the competitive environment

4.3.1 Network effects

One possible reason for differences in the likelihood of seeking a promotion is that there is some individual variable that we have so far ignored and which has an impact on the actual (or perceived) probability of success. A candidate explanation are research networks or groups of coauthors. The idea that networks are important in obtaining jobs and achieving promotions is widespread in the literature, and the issue has been addressed for promotions in academia; see McDowell and Smith (1992) for the US, Combes et al. (2008) for France and Zinovyevay and Bagues (2012) for Spain. Coauthor networks have been shown to differ across genders, with females having fewer coauthors and a lower fraction of male coauthors; see McDowell and Smith (1992) and Boschini and Sjögren (2007). If women have smaller or less efficient networks, then this may affect the expected outcome and hence the payoff from entering the competition for promotion. Potential reasons for this effect are that a candidate who has a member of her/his network in a promotion committee has a higher likelihood of success, but also that with a larger network the candidate's work may be better known and cited, or that this could provide extra information about how to best prepare for a *concours*. As a result, the gap that we find between men and women could be due to differences in coauthors and networks, and hence including these variables would have an effect on the coefficient on gender.

In order to test this hypothesis, we construct for each individual two measures of networks. Our measures are based on coauthorship, obviously an imperfect measure of actual networks, but one that is quantifiable with the EconLit data. Our first measure is the size of an individual's network, defined as the total number of different coauthors the researcher has had over his/her publishing lifetime; the second is the fraction of network members that are men. More than in other contexts, academic networks are highly endogenous, with research output and affiliation being both causes and consequences of an individual's network size, and patterns of network formation have been shown to differ across the genders; see Boschini and Sjögren (2007). Moreover, being a candidate for promotion can increase networks, especially in a system where individuals are often candidates several years in a row, since other candidates that are met during sessions aimed at the preparation of the *agregation concours* may become coauthors. As a result,

any resulting effect has to be interpreted with care, indicating correlation rather than as having causal implications.

Table 8: Network effects: likelihood to apply for a promotion, marginal effects

	Agreg', panel 1992-2008			Cnrs, panel 2003-2008		
	(1) logit	(2) probit	(3) OLS	(4) logit	(5) probit	(6) OLS
Women	-0.024 ^a (0.004)	-0.026 ^a (0.004)	-0.032 ^a (0.005)	-0.091 ^a (0.021)	-0.088 ^a (0.022)	-0.066 ^b (0.028)
Age	0.000 (0.002)	-0.002 (0.002)	-0.017 ^a (0.001)	0.097 ^a (0.011)	0.095 ^a (0.010)	0.061 ^a (0.007)
Age ²	-0.000 ^a (0.000)	-0.000 ^a (0.000)	0.000 ^a (0.000)	-0.002 ^a (0.000)	-0.002 ^a (0.000)	-0.001 ^a (0.000)
Publisher (Pub)	0.086 ^a (0.012)	0.084 ^a (0.012)	0.104 ^a (0.010)	-0.048 (0.060)	-0.057 (0.057)	0.013 (0.052)
Pub*Quantity	0.042 ^a (0.004)	0.042 ^a (0.004)	0.051 ^a (0.004)	0.041 ^b (0.020)	0.034 ^c (0.019)	0.030 (0.019)
Pub*Quality	0.009 ^a (0.001)	0.009 ^a (0.002)	0.015 ^a (0.002)	0.032 ^a (0.007)	0.031 ^a (0.008)	0.030 ^a (0.008)
Int. Dept.	0.012 ^b (0.006)	0.012 ^b (0.006)	0.013 ^b (0.006)	-0.010 (0.027)	-0.007 (0.027)	-0.010 (0.029)
Cnrs	-0.087 ^a (0.002)	-0.086 ^a (0.002)	-0.131 ^a (0.009)			
Pub*Network size	0.004 (0.005)	0.006 (0.005)	0.001 (0.007)	0.039 ^b (0.018)	0.040 ^b (0.018)	0.045 ^b (0.022)
% Men in Network	0.014 ^b (0.006)	0.015 ^b (0.007)	0.024 ^a (0.008)	0.057 (0.044)	0.058 (0.043)	0.046 (0.041)
Time FE	Yes	Yes	Yes	Yes	Yes	Yes
pseudo-R ²	0.263	0.259		0.291	0.290	
Observations	10383	10383	10383	681	681	681
log-likelihood	-2072	-2083	-291	-211	-211	-209

Standard error between brackets. ^a, ^b, ^c Significant at the 1%, 5% and 10% level, respectively. Int. Dept. = international department; see Appendix A for details.

Table 8 reports the regressions for the likelihood of entering the promotion contest for the two tracks, to which we have added network variables. Both variables have positive coefficients although only one of them is significant, the proportion of men in the case of university professors and the network's size for researchers. Despite the significance of these effects, the coefficients on gender barely change as compared to our earlier specifications. The only exception is the OLS regression for researchers where the coefficient increases from -0.073 (see Table 5) to -0.066. That is, gender gaps in networks do not seem to explain the lower propensity of women to seek promotion. The reason is probably that differences in network size across the genders are minor in France. In the larger sample, the average network size is 0.88 for men and 0.82 for women, while

researchers have a large number of coauthors, averaging 2.9 for women and 3.2 for men. These differences across the genders are not statistically significant. There is a greater gap in the fraction of men in the network, which are 70% for females in the *agregation* sample and 80% for males, and although it is significant the effect of this variable is moderate and hence its inclusion barely affects the coefficient on gender.

4.3.2 The competition's environment

Our final test consists of examining whether the environment under which the competition takes place affects men and women differently. We consider two variables. The first is the proportion of women amongst rank A professors in the year of the competition, which can act as a proxy for ‘perceived discrimination’. This variable has increased steadily over the period, going from 8.9% in 1992 to 15.2% in 2008. The other variable is the number of available positions as professor, which is known before the decision to be a candidate is taken, and ranges between 15 and 33, fluctuating substantially from year to year and showing no clear trend.

Since the effects of these variables can be identified only over the time dimension of the data we consider only the *concours d’agregation*. For the Cnrs we have a shorter time span, and hence little change in the fraction of women DRs (which goes from 10.6 to 10.9), while there is too little variability in the number of positions, as either 2, 3 or 4 are available every year. Because we are exploiting the time dimension of the data we cannot include time fixed effects as well as the variables of interest for the entire population. We hence consider two specifications. One includes time fixed effects and our two variables of interest interacted with the female dummy, which allows us to see whether effects differ across the genders but not to identify the impact on men. Alternatively, we substitute the fixed effects by a time trend and estimate the impact of these variables for both men and women.

Table 9 presents the estimated regressions. Column (1) indicates that the fraction of women amongst rank A professors has a greater effect on women than on men, and our specification with a time trend implies that this variable has no significant effect on men. This is consistent with the idea that perceived discrimination is a potential cause for females’ lower propensity to apply for promotion. Nevertheless, the coefficient on being a woman remains negative and highly significant.

Table 9: The competition's environment: likelihood to apply for a promotion, *agregation*, panel 1992-2008, marginal effects

	(1) logit	(2) logit	(3) logit	(4) logit
Women	-0.093 ^a (0.030)	-0.082 ^a (0.031)	-0.090 ^a (0.031)	-0.079 ^b (0.032)
Age	0.000 (0.002)	0.000 (0.002)	0.000 (0.002)	0.000 (0.002)
Age ²	-0.000 ^a (0.000)	-0.000 ^a (0.000)	-0.000 ^a (0.000)	-0.000 ^a (0.000)
Publisher (Pub)	0.093 ^a (0.011)	0.093 ^a (0.011)	0.093 ^a (0.011)	0.093 ^a (0.011)
Pub*Quantity	0.043 ^a (0.004)	0.043 ^a (0.004)	0.043 ^a (0.004)	0.043 ^a (0.004)
Pub*Quality	0.010 ^a (0.001)	0.010 ^a (0.001)	0.010 ^a (0.001)	0.010 ^a (0.001)
Int. Dept.	0.013 ^b (0.006)	0.013 ^b (0.006)	0.006 (0.006)	0.006 (0.006)
Cnrs	-0.087 ^a (0.002)	-0.087 ^a (0.002)	-0.087 ^a (0.002)	-0.087 ^a (0.002)
Time trend		-0.012 ^a (0.004)		-0.012 ^a (0.004)
% Women rank A		0.016 (0.011)		0.016 (0.011)
Women*% Women rank A	0.007 ^c (0.004)	0.006 ^c (0.003)	0.006 ^c (0.004)	0.006 ^c (0.003)
Men*Avail.Pos.		0.002 ^a (0.001)		0.002 ^a (0.001)
Women*Avail.Pos.	-0.000 (0.001)	0.002 ^c (0.001)	-0.001 (0.001)	0.001 (0.001)
Women*Int.Dept*Avail.Pos.			0.001 ^b (0.000)	0.001 ^b (0.000)
Time FE	Yes	No	Yes	No
pseudo-R ²	0.263	0.262	0.264	0.263
Observations	10383	10383	10383	10383
log-likelihood	-2072	-2075	-2070	-2073

Standard error between brackets. ^a, ^b, ^c Significant at the 1%, 5% and 10% level, respectively. Int. Dept. = international department; see Appendix A for details.

The number of available positions has no differential effect across the genders, as seen in column (1), and the coefficient on being female remains negative, large and highly significant. In column (2) we try to identify the effect for both men and women. For ease of interpretation, we interact this variable with a dummy that takes the value 1 for males (Men*Avail.Pos) as well as with the female dummy (Women*Avail.Pos), implying that the coefficient on this last variable captures the overall impact of the number of positions on women's decisions. We can see that a greater number of available positions increases the likelihood of being a candidate, as expected, capturing the idea that more

positions imply a greater probability of being promoted and hence encourages individuals to apply. The coefficient is the same across the genders, although the standard error is much greater for females, with the coefficient being significant only at the 10% level. A possible interpretation is that women are, on average, influenced by the number of positions just as much as men, but that there is greater variability in their reactions.

To further explore this aspect we ask whether the large variability in this effect stems from the fact that not all groups of women behave in the same way. A possible division is between women who are in a national and those in an international department. Women in different types of departments may react differently towards competition because of unobserved characteristics; in particular, some of them may be more ‘driven’ which makes them both manage to get hired by a better department and be more responsive to the competition’s environment. Columns (3) and (4) indicate that there is indeed a difference in the coefficients on the number of available positions for the two groups. Women in international departments respond to the number of positions available, with the coefficient being of the same magnitude as for men (since the overall coefficient for this group is the sum of those on Women*Avail.Pos. and $\text{Women*Int.Dept.*Avail.Pos.}$). In contrast, for women in national departments the decision to apply for promotion seems to be unaffected by how competitive the promotion contest is likely to be. A possible interpretation of this effect is that, contrary to their international department counterparts, women in national departments find the idea of taking part in the contest so unattractive that variations in the expected success rate do not affect their decision.

5 Conclusions

This paper has used data for promotions amongst academic economists in France to look at the attitudes of women during and towards contests. We have information both on who obtained and who was a candidate for promotion, which allows us to test different hypotheses about the reasons that cause women’s lower promotion rates.

On the one hand, women may be less likely to be promoted conditional on having applied, and this could be due either to discrimination or under-performance during the promotion contest. On the other, female academics may have a lower propensity to apply for promotion than males. This could in turn be due to the requirements of the

contest being more costly for women, to their facing a different tradeoff between salaries and department prestige, or to an unwillingness to enter the contest. The features of the French academic system, such as a national salary scale, the need to go through a national contest in order to be promoted, and the existence of several categories of academics with different requirements during the contest and upon promotion, allow us to test for these hypotheses.

We find that there is no difference between men and women in their probability of success, but that the latter are less likely than the former to enter contests. Moreover, our results indicate that neither the differential cost of promotion nor the tradeoff between income and department quality are behind these differences.

There are two possible explanations for our findings. One is that although women are not discriminated against during the contest they believe they will be, and hence decide not to enter the competition for promotion. The data points at this as a partial explanation of gender differences, with women being more likely to apply the more females there are amongst top-rank academics. The alternative explanation is that women are less willing than men to enter contests, in line with the experimental evidence provided by see Gneezy et al. (2003), amongst others. When we try to assess the impact of the competition's environment, we find that, unlike for men and women in top departments, for women in other departments the likelihood to apply for promotion is unaffected by the number of available positions in a particular year.

Our results have two main implications. On the one hand, they provide a link between laboratory evidence and behaviour observed in actual markets, and indicate how patterns of behaviour, namely women's lower propensity to take part in contests, can be observed in both contexts and what their consequences are for observed labour market outcomes. On the other, they raise the question of what type of policy intervention can help increase female promotion rates. The evidence in Gneezy et al. (2003) indicates that differences in contest participation seem to be partly driven by differences across genders in preferences for taking part in competitions, which can be the result of women being less confident than men. Building confidence is a process that is difficult and probably starts in early childhood, but changes in the way in which deciding whether to enter a competition occurs could facilitate women's probabilities of climbing up the rank ladder. For example, a system of mentoring whereby junior faculty are assigned a mentor that 'proposes' them

as candidates for promotion may incite more women to apply. Alternatively, creating a system in which the default is that an individual will be considered for promotion after x years and s/he has to opt out instead of opting in, could also be a way of overcoming the differences in confidence across the genders.

Appendix

A Definition of international departments

Academia in France is organised around ‘research centers’, with a university potentially having several research centers in economics. We have defined ‘departments’ either by an economics department when it is the single affiliation where economists are found in a given university (which corresponds to the majority of cases), or by the aggregation of all research centres where there are economists in the university. We performed robustness checks by looking at research centres rather than departments, and obtained consistent results (available upon request). However, our notion of slightly aggregated economics departments better matches the reality of French academic research and hence have preferred to focus our discussion of the results on this concept.

We define ‘international departments’ as departments with the highest research output, measured by both the total stock and average stock of publications per member of the department in EconLit journals. Also, we wish to take into account both “normal” quality and top quality research, two measures that we denote CLm and CLh, respectively, following Bosquet and Combes (2013*b*). In order to do so, we apply the following procedure at each point in time. We first calculate the total research output of a department (in terms of the CLm and CLh measures) as the sum of research outputs of its members. We also calculate the two scores per member of the department. Hence, we obtain 4 scores per department: CLm total, CLh total, average CLm and average CLh. Departments are then ranked with respect to these 4 scores and their final score is the average of these rankings.

The cutoff point for top departments is arbitrary. We choose our cutoff at each point in time so as to have 35% of the total population in the top department category. Since departments may change their status over time, we considered as international departments for the entire sample period those departments which are more often than not in the top department category.

Table 10: List of international departments

Cepremap	1991	2004
Crest	1993	2008
Ecole Polytechnique	1991	2008
Ehess-Ens (Delta)	1991	2004
Enpc (Ceras, Latts)	1991	2004
Hec	1995	2008
University Aix Marseille 2-3	1991	2008
University Cergy	2005	2008
University Cergy-Paris 10	1991	2004
University Clermont 1	1991	2008
University Nancy 2-Strasbourg	2006	2008
University Paris 1	1991	2004
Paris School of Economics	2005	2008
University Paris 10	2005	2008
University Paris 9	1991	2008
Toulouse School of Economics	1991	2008

The last two columns give the first and last data at which the departments are observed in the sample, respectively. Note that some departments have been merged, as for instance the Cepremap, the Ehess-Ens and the University Paris 1 to form the Paris School of Economics from 2005.

B Likelihood to be in an international department and to hold a rank A position by category

Table 11: Likelihood to be in an international department and to hold a rank A position by category, 1992-2008, marginal effects

	International department					Rank A	
	All (1) logit	All (2) logit	All (3) logit	Univ. (4) logit	Cnrs (5) logit	Univ. (6) logit	Cnrs (7) logit
Women	0.020 ^b (0.008)	0.022 ^b (0.009)	0.068 ^a (0.009)	0.071 ^a (0.009)	0.079 ^b (0.034)	-0.064 ^a (0.007)	-0.120 ^a (0.023)
Age		-0.010 ^a (0.002)	-0.009 ^a (0.002)	-0.008 ^a (0.002)	-0.026 ^a (0.007)	0.011 ^a (0.002)	0.059 ^a (0.007)
Age ²		0.000 ^a (0.000)	0.000 ^a (0.000)	0.000 ^a (0.000)	0.001 ^a (0.000)	0.000 ^c (0.000)	-0.001 ^a (0.000)
Publisher (Pub)			0.192 ^a (0.012)	0.196 ^a (0.013)	0.041 (0.054)	0.324 ^a (0.013)	0.036 (0.039)
Pub*Quantity			0.051 ^a (0.005)	0.049 ^a (0.005)	0.083 ^a (0.017)	0.147 ^a (0.004)	0.145 ^a (0.012)
Pub*Quality			0.048 ^a (0.002)	0.047 ^a (0.002)	0.058 ^a (0.007)	0.041 ^a (0.002)	0.047 ^a (0.005)
Cnrs			0.185 ^a (0.024)				
Women*Cnrs			0.018 (0.031)				
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
pseudo-R ²	0.012	0.014	0.082	0.056	0.118	0.347	0.487
Observations	16077	16077	16077	14848	1229	14848	1229
log-likelihood	-10022	-9999	-9315	-8578	-721	-6212	-432

Standard error between brackets. ^a, ^b, ^c Significant at the 1%, 5% and 10% level, respectively.

References

- Bertrand, M. (2011), New Perspectives on Gender, *in* O. Ashenfelter and D. Card, eds, ‘Handbook of Labor Economics’, Elsevier, pp. 1543–1590.
- Bertrand, M. and Hallock, K. F. (2001), ‘The Gender Gap in Top Corporate Jobs’, *Industrial and Labor Relations Review* **55**(1), 3–21.
- Blackaby, D., Booth, A. L. and Frank, J. (2005), ‘Outside Offers and the Gender Pay Gap: Empirical Evidence from the UK Academic Labour Market’, *Economic Journal* **115**(501), 81–107.
- Blau, F. and Kahn, L. M. (2000), ‘Gender Differences in Pay’, *Journal of Economic Perspectives* **14**(4), 75–99.
- Boschini, A. and Sjögren, A. (2007), ‘Is Team Formation Gender Neutral? Evidence from Coauthorship Patterns’, *Journal of Labor Economics* **25**(2), 325–365.
- Bosquet, C. and Combes, P.-P. (2013a), ‘Are academics who publish more also more cited? Individual determinants of publication and citation records’, *Scientometrics* (DOI 10.1007/s11192-013-0996-6).
- Bosquet, C. and Combes, P.-P. (2013b), ‘Do Large Departments make Academics more Productive? Agglomeration and Peer Effects in Research’, *CEPR Discussion paper* (9401).
- Cole, J. and Cole, S. (1973), *Social Stratification in Science*, Chicago: The Univeristy of Chicago Press.
- Combes, P.-P. and Linnemer, L. (2010), ‘Inferring Missing Citations. A Quantitative Multi-Criteria Ranking of all Journals in Economics’, *GREQAM Working Paper* .
- Combes, P.-P., Linnemer, L. and Visser, M. (2008), ‘Publish or Peer-Rich? The Role of Skills and Networks in Hiring Economics Professors’, *Labour Economics* **15**(3), 423–441.
- Farber, S. (1977), ‘The Earnings and Promotion of Women Faculty: Comment’, *American Economic Review* **67**(2), 199–206.

- Ginther, D. K. and Hayes, K. J. (1999), ‘Gender Differences in Salary and Promotion for Faculty in the Humanities’, *American Economic Review Papers and Proceedings* **89**(2), 397–402.
- Ginther, D. K. and Kahn, S. (2004), ‘Women in Economics: Moving Up or Falling Off the Academic Career Ladder?’, *Journal of Economic Perspectives* **18**(3), 193–214.
- Gneezy, U., Niederle, M. and Rustichini, A. (2003), ‘Performance in Competitive Environments: Gender Differences’, *Quarterly Journal of Economics* **118**(3), 1049–1074.
- Goldin, C. and Rouse, C. (2000), ‘Orchestrating Impartiality: The Impact of ”Blind” Auditions on Female Musicians’, *American Economic Review* **90**(4), 715–742.
- Gupta, N. D., Poulsen, A. and Villeval, M. C. (2013), ‘Gender Matching and Competitiveness: Experimental Evidence’, *Economic Inquiry* **51**(1), 816–835.
- Johnson, G. E. and Stafford, F. P. (1974), ‘Lifetime Earnings in a Professional Labor Market: Academic Economists’, *Journal of Political Economy* **82**(3), 549–69.
- Kahn, S. (1995), ‘Gender Differences in Academic Career Paths of Economists’, *Journal of Economic Perspectives* **9**(4), 193–204.
- Kelchtermans, S. and Veugelers, R. (2013), ‘Top Research Productivity and Its Persistence: Gender as a Double-Edged Sword’, *The Review of Economics and Statistics* **95**(1), 273–285.
- Lazear, E. P. and Rosen, S. (1990), ‘Male-Female Wage Differentials in Job Ladders’, *Journal of Labor Economics* **8**(1), S106–S123.
- Maliniak, D., Powers, R. and Walter, B. F. (2013), ‘The Gender Citation Gap in International Relations’, *International Organization*, (Available on CJO 2013 doi:10.1017/S0020818313000209).
- McDowell, J. M., Singell, L. D. and Ziliak, J. P. (2001), ‘Gender and Promotion in the Economics Profession’, *Industrial and Labor Relations Review* **54**(2), 224–244.
- McDowell, J. M. and Smith, J. K. (1992), ‘The Effect of Gender-sorting on Propensity to coauthor: Implications for Academic Promotion’, *Economic Inquiry* **30**(1), 68–82.

- Niederle, M. and Vesterlund, L. (2007), ‘Do Women Shy Away from Competition? Do Men Compete Too Much?’, *Quarterly Journal of Economics* **122**(3), 1067–1101.
- Niederle, M. and Vesterlund, L. (2011), ‘Gender and Competition’, *Annual Review of Economics* **3**(1), 601–630.
- Petrongolo, B. and Olivetti, C. (2008), ‘Unequal Pay or Unequal Employment? A Cross-Country Analysis of Gender Gaps’, *Journal of Labor Economics* **26**(4), 621–654.
- Sabatier, M. (2010), ‘Do Female Researchers Face a Glass Ceiling in France? A Hazard Model of Promotions’, *Applied Economics* **42**(16), 2053–2062.
- Winter-Ebmer, R. and Zweimuller, J. (1997), ‘Unequal Assignments and Unequal Promotion in Job Ladders’, *Journal of Labor Economics* **15**(1), 43–71.
- Zinovyevay, N. and Bagues, M. (2012), ‘The Role of Connections in Academic Promotions’. mimeo.

Is the Clock Still Ticking? An Evaluation of the Consequences of Stopping the Tenure Clock

Authors: Colleen Flaherty, Lisa M. Leslie and Amit Kramer

<https://doi.org/10.1177/001979391306600101>

Abstract

Using a longitudinal administrative data set from a large research university, the authors empirically evaluate the consequences of using stop the clock (STC) policies for the career success of tenure track faculty. STC policies were introduced approximately 40 years ago, yet surprisingly little has been written about how they affect career outcomes. The prevalence of the ideal worker norm in academia raises the possibility of negative consequences as evaluators may treat STC policy use as a signal that the faculty member lacks sufficient commitment to his or her academic role. Consistent with this possibility, the authors find that faculty members who stop their clock for family reasons incur a salary penalty relative to faculty members who do not stop their clock, which cannot be explained by differences in productivity. Alternatively, faculty members who use the policy are not at a promotion disadvantage as compared with nonusers, and they actually have higher promotion rates.

Research output at US economics departments

Authors: Michael A. McPherson, Myungsup Kim, Megan Dorman and Nishelli Perera

Pages 889-892 | Published online: 07 Feb 2013

<https://doi.org/10.1080/13504851.2012.761331>

Abstract

Using curricula vitae and EconLit, we examine the publication records of nearly 2000 academic economists. We consider the probationary period and the years between tenure and promotion to professor. Faculty tenured at top programmes average 3.44 articles in top-ten journals (and 8.75 total articles). The quality of faculty members' publications decreases with programme ranking; the quantity of publications does not differ much among top-100 programmes. Those promoted to professor generate fewer top-ten and total articles. There is some evidence that females produce fewer top-ten and total journal articles than males at many programmes.

IZA DP No. 6990

**Where Do New Ph.D. Economists Go?
Evidence from Recent Initial Job Placements**

Jihui Chen
Qihong Liu
Sherrilyn Billger

November 2012

Where Do New Ph.D. Economists Go? Evidence from Recent Initial Job Placements

Jihui Chen

Illinois State University

Qihong Liu

University of Oklahoma

Sherrilyn Billger

*Illinois State University
and IZA*

Discussion Paper No. 6990
November 2012

IZA

P.O. Box 7240
53072 Bonn
Germany

Phone: +49-228-3894-0
Fax: +49-228-3894-180
E-mail: iza@iza.org

Any opinions expressed here are those of the author(s) and not those of IZA. Research published in this series may include views on policy, but the institute itself takes no institutional policy positions. The IZA research network is committed to the IZA Guiding Principles of Research Integrity.

The Institute for the Study of Labor (IZA) in Bonn is a local and virtual international research center and a place of communication between science, politics and business. IZA is an independent nonprofit organization supported by Deutsche Post Foundation. The center is associated with the University of Bonn and offers a stimulating research environment through its international network, workshops and conferences, data service, project support, research visits and doctoral program. IZA engages in (i) original and internationally competitive research in all fields of labor economics, (ii) development of policy concepts, and (iii) dissemination of research results and concepts to the interested public.

IZA Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

ABSTRACT

Where Do New Ph.D. Economists Go? Evidence from Recent Initial Job Placements^{*}

We use data from the 2007-2008 Ph.D. economist job market to investigate initial job placement in terms of job location, job type, and job rank. Our results suggest gender differences in all three dimensions of job placement. Relative to their male counterparts, female candidates are less (more) likely to be placed into academic (government or private sector) jobs and, on average, are placed into worse ranked jobs. Foreign female candidates are also more likely than foreign males to stay in the U.S. When foreign students are placed outside the U.S., they are more likely to be in academia than in government or private sector, while the opposite holds when foreign students are placed in the U.S., which is largely consistent with a stylized theory model. Our results also reveal various country/region heterogeneities in the type, location, and rank of job placements.

JEL Classification: A11, A23, J44

Keywords: Ph.D. labor market, job type, job location, job rank

Corresponding author:

Sherrilyn M. Billger
Department of Economics
Illinois State University
Normal, IL 61790
USA
E-mail: smbillg@ilstu.edu

^{*} We thank Leah Brooks, Greg Burge, Jen Graves, John Ham, Cory Koedel, Catherine Mooney, Rati Ram, Adam Rennhoff, and session participants at the 2008 Southern Economic Association Annual Conference and Illinois State University Applied Econometrics Workshop participants for helpful comments and suggestions. We also thank Meredith Willinger, Yoonho Choi, Chris Denly, Hojin Jung, and Golaleh Moshrefi for their excellent research assistance, and University of Oklahoma Honors Research Assistant Program for financial support. Finally, we would like to thank placement coordinators and other individuals at various institutions who exchanged correspondences with us during the course of this project. The usual caveat applies.

1 Introduction

The extensive literature on the economist labor market focuses on initial placement or job conditions across all tenure lengths.¹ These studies show that job outcomes vary with demographic characteristics (e.g., gender and nationality), academic characteristics (e.g., grades, GRE scores, advisor reputation, and research portfolio), and program characteristics (e.g., the rank and size of a doctoral program). Our paper is similar to the existing literature in that we investigate the relationship between job placement and various individual and program characteristics. On the other hand, our paper differs from existing studies in several ways.

First, due to the labor-intensive data collection process, existing studies usually focus on a handful of top-ranked programs or use data collected from the AEA Membership directory or surveys with low response rates.² In recent years, the information on job candidates and placement outcomes has become more accessible to researchers, as many programs advertise their candidates and placements on the internet. We have therefore compiled a comprehensive sample of 578 candidates from 57 top economics programs in the 2007-2008 job market,³ allowing us to conduct an analysis generally free of selection bias.

Second, in addition to gender (which is extensively studied in the literature), we explore source country heterogeneity in job placement outcomes. Existing studies usually use a single country dummy (i.e., the U.S.) as a demographic characteristic, with a few exceptions (Krueger and Wu, 2000; Grove, Dutkowsky and Grodner, 2007) who use a combination of country and continent dummies. Our analysis provides a more careful exploration across many countries.

Third, few existing studies are concerned with whether the candidate is placed within or outside the U.S. Instead, studies analyzing job location usually consider regions within the U.S. The American labor market has become more saturated in recent years, partially fueled by a weak economy. This saturation, coupled with the increased prestige of many international universities, has led a number of Ph.D.s (particularly foreign-born) to pursue employment outside the U.S. Anecdotal evidence reveals this trend. For example, Siegfried and Stock (2004) document that the percentage of foreign candidates has been increasing over time, from about 33% in the 1976-1977 cohort to 62% in the 2000-2001 cohort. They also report the percentage of U.S. placements but they include the U.S. candidates while we only consider foreign candidates in the job location analysis. This is also seen in our data, where about 71% (409 out of 578) of the candidates are foreigners and 37% (146 out of 395) of those who succeed in initial job placements are placed outside the U.S. Therefore,

¹Coupe (2004) and Ehrenberg (2004) provide excellent surveys of this literature. There are also studies focusing on the success of completing the Economics Ph.D. (Siegfried and Stock 2001; Stock and Siegfried 2006a; Grove, Dutkowsky and Grodner 2007).

²For example, Krueger and Wu (2000), Oyer (2006), Athey et al. (2007), and Grove and Wu (2007) use a handful of top programs. McMillen and Singell (1994) and McDowell, Singell and Ziliak (1999, 2001) use AEA membership directory data. Most other studies use survey data, including Barbezat (1992), Neumark and Gardecki (1998), Siegfried and Stock (1999, 2004), Stock and Siegfried (2001), and Gallet et al. (2005).

³Source: <http://consusrankings.com/2008/04/28/us-news-best-graduate-economics-programs-2005/>.

it is increasingly important to document the patterns of job placements in terms of geographic location. The significance of this is twofold. First, it measures supply in the labor market of U.S. and foreign economists. Second, a candidate's investment in human capital should adjust to better match their future job location. Currently, the prominent job search manuals (e.g., Cawley, 2011) are almost exclusively based on the U.S. market, but it appears that the global market may be considered. To more carefully explore this phenomenon, we also offer a stylized theory model to link job type and job location.

In this study, we pose the following research question: where do new Ph.D. economists go? In particular, we examine three important elements of initial job placements: job location (within or outside the U.S.), job type (academia, government, or private sector) and job rank.

Our results reveal significant gender differences and source country heterogeneity in job location. Among foreign candidates, females are more likely to remain in the U.S. relative to their male counterparts. In the job location analysis, we introduce a set of country/region dummies to account for the considerable country/region disparities arising in the sample (more details can be found in Section 5.1). For example, *ceteris paribus*, candidates from Argentina and India are most likely to work in the U.S. upon graduation, while Korean students are the least likely to stay.

In addition, we show evidence of gender difference in the placement outcomes.⁴ For example, in our baseline model, we find that females are 7.6% less likely to be placed into academic jobs relative to male candidates. On the other hand, female candidates are more likely to be placed into government jobs (2%) and private sector jobs (5.4%), but these gender differences are insignificant. Our findings are in line with previous studies (e.g., McDowell, Singell and Ziliak, 1999, 2001; Broder, 1993) which show that gender difference has diminished over time. Our results also indicate that female candidates' placements are on average almost 9 ranks worse relative to male candidates in our baseline model, but this difference is insignificant.⁵ Our results further reveal significant country/region heterogeneity in both job type and job rank. We also find significant gender differences in job rank for candidates from selected countries.

Finally, to offer some theoretical considerations for the main results, we develop a stylized model to link job type and job location. Focusing on foreign candidates, we divide all permanent jobs into two categories: academic and non-academic (i.e., government and private sector jobs are combined), to be consistent with the theoretical model. In foreign job markets, non-academic employers have a relatively large pool of local labor supply, but academic jobs rely more on U.S.-educated Ph.D. economists. Under this assumption, candidates would face less (more) competition in academia

⁴Note that this gender difference does not necessarily reveal gender discrimination. For example, Barbezat (1992) documents significant gender differences in their preferences for jobs types. On the other hand, McMillen and Singell (1994) find no significant gender difference in sector choice probabilities (employee preference), but the marginal effect of several productive attributes is smaller for women than for men (employer discrimination).

⁵Our job rank analysis follows Krueger and Wu (2000) closely. Other studies have adopted more crude measure of job rank to compare gender differences. For example, McMillen and Singell (1994) divide placements into top 50 economics department and other departments. Barbezat (1992) divide employment into top 15, 16-30, and other economics departments, respectively.

(nonacademia) in the foreign job market than in the U.S. market. Thus, we expect to see that, for foreign candidates, foreign placements are more likely to be academic than U.S. placements. This is confirmed in our empirical results where we find that U.S. placements are less likely to be academic than foreign placements, by 2.7%–7.5% depending on the model specification. This correlation becomes significant and much stronger (7%–13.4%) when Korean candidates are excluded. We see that those who stay in the U.S. are much more likely not to take academic jobs.⁶

The remainder of the paper is organized as follows. Section 2 offers a review of the related literature. We discuss our data in Section 3 and provide descriptive statistics results in Section 4. Section 5 contains the main estimation results on job type, job location and job rank as well as a stylized theory linking job type and job location. We conclude in Section 6. A list of variables and proof of Proposition 1 are provided in Appendix A.1 and A.2, respectively.

2 Literature review

Our paper is most closely related to studies on initial job placement of new Ph.D. economists (Siegfried and Stock 1999, 2004; Stock and Siegfried 2001, 2006). Several studies analyze the type of initial job placements and/or starting salary. Barbezat (1992) is among the first to study the labor market for new economists. She surveys about 600 candidates (with 291 usable responses) in the top 46 Economics programs in the 1988-1989 market. Her data set consists of both CVs and questionnaires. In particular, her survey includes candidates' preferences across different job types which she uses together with other individual and program characteristics to predict candidates' job types. While our analysis also relies on candidates' CVs, much has changed since the Barbezat study, allowing us to offer an updated understanding of recent doctoral education outcomes. Siegfried and Stock (1999) conduct an interesting survey of economics Ph.D. graduates to explore initial job placement and job satisfaction, highlighting differences across employment sector. They document different placement rates among various sub-fields and background characteristics, such as whether the candidate completed a master's degree prior to starting a Ph.D. program (we also include this variable in our analysis). Stock and Siegfried (2001) survey the 1996-1997 cohort of Ph.D. economists and investigate job type and salary. In an updated version, Siegfried and Stock (2004) provide an array of summary statistics as a snap-shot of the 2001-2002 graduating class. They highlight the influx of international workers, salary differences across sectors, and job satisfaction distributions by sector and Ph.D. program ranking. McMillen and Singell (2001) use the AEA Membership Directory and focus on the role of gender in job search. They conduct the analysis using subsamples of each gender, and conclude that pre-market differences in attributes do not fully explain the post-market gender discrepancies in job placement. Their findings of gender differences could stem from either the employer side (discrimination) or the candidate side (candidates'

⁶More details are provided in Section 5.3.2 where we also offer an explanation of why the Korean candidates stand in sharp contrast with others in terms of the relationship between job type and job location.

preferences).⁷

Other studies are concerned with mentorship configurations in terms of gender and their impact on job placement outcome. For example, Neumark and Gardecki (1998) survey various programs for their female graduate students. They find that neither the number of female faculty in the department nor the gender of dissertation chair affects the quality of female students' first job placement. Hilmer and Hilmer (2007) extend this analysis by including male graduate students and analyzing research productivity during the doctoral program, in addition to placement quality. They find that female students working with male advisors are more likely to accept research-oriented jobs than male students with male advisors upon graduation.

Aside from initial job placements, intermediate job search outcomes are also analyzed in this literature. Stock and Alston (2000) include both new and previous Ph.D.s in the 1995-1996 market and investigate the impact of various individual and program characteristics on the number of interviews a candidate receives and the probability of receiving at least one offer. Gallet, List and Orazem (2005) compare employers' and candidates' search strategies between boom (i.e., 1987) and bust (i.e., 1997) job markets. Using survey data of two cohorts (boom and bust), they confirm that in weak job markets, academic employers raise their hiring standards, candidates from top programs are more likely to apply for lower-ranked academic positions while those from low programs are more likely to apply for nonacademic positions.

Another strand of the literature analyzes later career outcomes, not just the initial job placement. Much of the focus is on gender gap.⁸ Kahn (1995) reports statistics on female economists from various aspects such as grades, Ph.D. admission and graduation, job placement, publications and salary. McDowell, Singell and Ziliak (1999, 2001) use panel data of AEA members in 4 selected years between 1960 and 1990. They document evidence of a glass ceiling for women in economics; women are less likely to be promoted, or are promoted more slowly than men, after controlling for research productivity, life-cycle attributes and fields of specialization. However, this gender difference is diminishing over time.⁹ Gender differences in promotion also exist in other academic markets. Ginther and Hayes (2003) identify significant gender differences in promotion for humanities professors which are responsible for the gender differences in salary.¹⁰ In contrast, Blackaby and Frank (2000) and Blackaby, Booth and Frank (2005) find both gender difference in promotions and within-rank gender pay gaps. Their UK survey data include outside offers, which they argue explain the gender differences in pay and promotion. In another study, Kolpin and Singell (1996)

⁷Results in Barbezat (1992) Table 6 indicates significant gender differences in how they value certain workplace characteristics.

⁸Some studies focus on race, in particular, the status of minorities in the Economics profession. See, for example, Collins (2000).

⁹This is in the same spirit as Broder (1993), who uses data on grants and grant applications of the NSF Economics program and finds significant gender difference in older cohorts but not at the assistant professor level, both in terms of salary and other job characteristics.

¹⁰In a recent paper, Krause, Rinne and Zimmermann (2012) conduct a randomized experiment among applicants to a research institution and analyze the impacts of anonymous job applications on interview invitations.

investigate gender difference in initial job placements and its subsequent impacts on departmental research productivity. Using data on economists in 4 years between 1970s and 1980s, they find that higher-ranked departments are less likely to hire females, despite the fact that women in 1970s cohort were more productive than their male counterparts at comparable institutions. Moreover, departments hiring fewer women in 1970s subsequently declined in publications rank relative to other departments, suggesting that employment discrimination is costly to the employer.

Other studies investigate the determinants of placement quality. Using a sample consisting of 1,029 graduate students who enrolled in the top five economics programs in the 1990s, Athey et al. (2007) find that, conditional on first-year grades, neither demographic characteristics, nor GRE scores, nor an additional master’s degree seems to affect the quality of job placement. Using data on applicants to a top 5 Economics program in 3 selected years, Krueger and Wu (2000) examine the relationship between admission to graduate program and the subsequent outcome of job placement, and argue that subjective ratings of admission committee and objective GRE scores are effective signals of professional success. Still other studies analyze the determinants of research productivity. Buchmueller, Dominitz and Hansen (1999) use the AEA COGEE survey and regress research productivity on initial job placement (rank and type) and pre-market individual and program characteristics. Oyer (2006) finds that better initial placements lead to greater professional productivity in the long run. Grove and Wu (2007) identify both subjective and objective selection criteria in top-five economics doctoral programs. Their main contributions include the use of the identity of letter writers and admission committee ranking as effective signals of economics talent. They conclude that superior math skills (e.g., quantitative GRE scores) and the prominence of letter writers are highly correlated with applicants’ subsequent success in doctoral program and future academic career.

3 Data

To capture most candidates on the Ph.D. economist job market, we include 57 top U.S. economics programs in our sample.¹¹ Due to the nature of the job search process, our data collection involved two steps. First, in November 2007, we obtained a total of 617 candidates’ CVs from each program’s placement web page.¹² From these CVs, we extracted the candidate and program characteristics variables, which are described in Appendix A.1. The second step started in fall 2008 when we collected candidates’ placement information.

Focusing on initial job placements in the 2007-2008 market, we employ two criteria during this two-step process: (1) it should be the candidate’s first time on the job market, and (2) it should be

¹¹We selected 57 of the top graduate economics programs based on the *US News* 2005 ranking. We include economics programs in business schools as well.

¹²It is possible that a listed job market candidate may later choose not to proceed and stay in the program one more year. Our data seem to confirm this possibility. It is also possible but less likely that a late candidate may enter into the market after mid-November for jobs starting the following year.

his/her first job placement. Some programs list the placement outcomes online. For the remaining candidates, we conducted extensive online searches through which the placement outcomes for most candidates placement outcomes were confirmed. In cases where the internet search failed, we directly contacted their program placement directors/coordinators, advisors, faculty members and sometimes other students in the same department for a definitive answer. We were able to confirm all placement outcomes in our final sample.

Our final sample includes a total of 578 candidates. While this is not an exhaustive list of all the job candidates in the year 2007-08, the sample is representative of candidates (including both new Ph.D.s and ABDs) who participated in the initial job search, since 57 of the top programs are included. That particular job market may be considered as successful overall in that only 19 candidates were not placed as of fall 2008. More details are provided in Section 4.

4 Descriptive analysis

Our data include the following information for each candidate: (1) demographic information including gender and citizenship; (2) academic information including characteristics of the candidate's Ph.D. program and advisor(s), whether s/he has a previous master's degree, teaching awards or had journal publication or revise and resubmit; and (3) placement information including whether the candidate is placed successfully and if yes, the type, location and rank of job placement. Now we turn to descriptive statistics of each category. Refer to Appendix A.1 for variable definitions.

Demographic characteristics

Table 1 lists the gender composition of the sample. About 1/3 of the candidates (199/578) are female and this ratio is roughly the same across program tiers. Figure 1 indicates the female ratio by country/region. Taiwan has the highest female ratio (61%) among Asian nations with more than 10 students, followed by China (61%) and India (52%). Of the European countries with over 10 students, Italy has the highest ratio of female students (55%) followed by Russia (30%). A few Latin American countries also have high female ratios, such as Colombia and Venezuela.

<<Figure 1 here>>

We also compare the ratios of international candidates between doctoral programs and initial job placements. In Table 1, the fraction of foreign students is inversely U-shaped in program ranking, with Tier 1 and Tier 4 admitting a relatively lower percentage of foreign students than the middle tiers (tier 2 and 3) do.

<<Table 1 here>>

Academic characteristics

Table 2 reports the summary statistics for academic characteristics. All variables except *size* are dummy variables. About half of the candidates have a master's degree prior to entering the doctoral program. As expected, only 4% of the candidates have a top advisor or co-advisor. 1% of the candidates have either a journal publication or revise and resubmit (R&R) at the top four economics journals, and the fraction rises slightly for those who have either a publication or R&R at the economics journals ranked top 5 through 50 (5% and 3%, respectively). In the sample, about 26% of the candidates earned at least one teaching award. Finally, the average program size, defined as the number of candidates on the job market, ranges between 1 and 32.

<<Table 2 here>>

Placement characteristics

Stayus is defined as the candidate accepts a job in the U.S. while *Return* refers to the case where the candidate accepts a job in his/her home continent.¹³ As almost all U.S. students take a job in the U.S., we thus exclude American students when reporting the results in Table 3, as well as those who fail to find a job, leaving us a subsample of 395 students.

<<Table 3 here>>

Table 3 reports the fraction of students staying and the fraction returning, separated by gender, program tier, and continent respectively. Foreign students usually either stay in the U.S. or return to home continent/country, so we expect opposite trends between the stay and return ratios. About 63% of foreign students accepted jobs in the U.S., and this ratio varies slightly across gender (female 69% vs. male 60%) and program ranking (tiers 1 and 4 slightly higher than tiers 2 and 3). Approximately 23% of foreign female candidates return to their home continents, compared to 31% for foreign male students, while the return ratios are higher for tiers 2 and 3 than for the other two tiers.

When broken down by continents, we find that Middle East has the highest ratio of staying in the U.S., or 81%, with other continents in the vicinity of 60%, except for Africa. In terms of return-to-home-continent ratios, Mideast, Latin America, and Africa rank the lowest three, in such order. Figures 2 and 3 show the distribution of these ratios by country. Darker color indicates the higher stay/return ratio. Among countries with at least 20 candidates, South Korea has the lowest stay ratio (38%), followed by Russia (43%), China (61%), Turkey (72%) and India (74%).

<<Figure 2 here>>

<<Figure 3 here>>

¹³These two outcomes coincide with each other for U.S. candidates. For candidates from other continents, they are not perfectly correlated since a foreign student can be placed into a country outside U.S. and his/her home continent.

Table 4 reports the types of the initial jobs taken by the Ph.D.s in our sample. Note that we now include U.S. candidates. As one would expect, a considerable number of students (309 or about 55%) were placed into (tenure-track) academic positions, followed by the private sector (125 or about 22%). 14% of candidates found government jobs (e.g., Federal Agencies and regional central banks) and the remaining 48 candidates (9%) found temporary jobs (e.g., postdoc and visiting positions). Across gender, female candidates are less likely than males to be placed into academic jobs (49% vs. 59%) and the result is reversed for the private sector (27% vs. 20%). There is little difference across program tiers, but more prominent differences exist across continents.

<<Table 4 here>>

5 Estimation results

In the previous section, we discuss descriptive statistics on the composition of candidates by home nation/continent, gender, program ranking, and their placement in terms of location (stay or return) and type. Next, we present a variety of regression results to explore how various candidate characteristics affect their placement outcomes.

5.1 Job location

The U.S. job market for economists is obviously impacted by the health of the overall economy. Decades ago, we might have expected most international students to try to stay in the U.S., but that is no longer obvious. Some other economies (e.g. China) have been booming and job candidates are increasingly attracted there. We therefore explore the factors that contribute to whether an international candidate stays in the U.S. or takes a job in another country. Approximately one third of our sample are candidates from the U.S. and they are placed within the U.S. with very few exceptions. Including Americans in the location analysis would significantly affect the results, so we exclude them from this analysis. We also exclude the candidates without jobs for obvious reasons. However, we include those placed to temporary jobs as they may indicate candidates' location preferences.¹⁴ The final subsample used for this location analysis includes 395 foreign candidates.

We first explore whether international candidates stay in the U.S. The dependent variable *Stayus* is equal to 1 if the candidate takes a U.S. job and 0 otherwise. Due to extensive heterogeneity across countries within a given continent, a seemingly better choice is to use selective country dummies. We use two criteria in selecting these countries. The first criteria is the number of candidates from that country in our sample. The second is the GDP of the candidate's home country/region. There

¹⁴The results are qualitatively the same if we exclude temporary jobs.

is a lot of overlap in these two criteria and in the end we choose the following countries/regions: *China, India, Korea, Russia, Turkey, Japan, Italy, Taiwan, Argentina.*

Probit estimation results are presented in Table 5. Model (1) is the baseline specification, and models (2) and (3) include interaction terms between gender and country, and between gender and program ranking, respectively. In all three models, female international candidates are 5.1-6.5% more likely to stay in the U.S. than their male counterparts (statistically insignificant), depending on the model specification. The next set of variables are country/region dummies. Some of the patterns we observed in the descriptive statistics results remain. For example, compared to the reference group, candidates from Argentina and India are most likely to work in the U.S. upon graduation, while Korean students are the least likely to stay. In addition, we have some evidence that candidates from China and Turkey are more likely to stay in the U.S. while those from Italy are less likely to stay, relative to the reference group. Moving on to other candidate characteristics, candidates are about 10% less likely to stay if they had obtained a separate master's degree prior to joining the doctoral program. Relative to candidates from Tier 1 programs, those from lower-ranked programs are at least 28% less likely to stay in the U.S.¹⁵ The marginal effects of program quality are robust across specifications, suggesting that program quality is an important determinant of job location, perhaps through access (or lack thereof) to job networks. For each additional student in the program, the likelihood of staying in the U.S. goes down by about 1% (statistically insignificant).

<<Table 5 here>>

We expect dissertation advisors to impact candidates' research productivity during graduate training and their eventual job placements. Although we do not have access to the content of recommendation letters, we do know who candidates' advisors are, since this information is usually listed on the CV. Grove and Wu (2007) adopt a similar proxy for letter writers. The underlying assumption is that the more prominent the letter writer is, the better job placement will be. Interestingly, as indicated in Table 5, a candidate with a top advisor is about 30% less likely to stay in the U.S. Overall, these results remain qualitatively the same with the inclusion of interactions between gender and region/program ranking.

5.2 Job type

Next, we explore the likelihood of securing one of the four job types: academia, government, private sector, or temporary. Recall that visiting positions and postdocs are included in temporary jobs, not academic jobs. We estimate these probabilities using a multinomial logit model. The dependent variable equals 1 if the candidate takes a job in academia, 2 if in government, 3 if in private sector,

¹⁵Our results suggest that candidates from Tier 4 programs are more likely to stay in the U.S. than those from Tier 2 and 3 programs. That is, the relationship between job location and program ranking is U-shaped. One may speculate that foreign placements are better fits for candidates in the two middle tiers than tier 1 and 2.

and 4 if it is a temporary position. This analysis uses a sample of 559 candidates who are placed successfully (including American students).

Table 6 reports the results from multinomial logit regressions, incorporating candidate characteristics such as gender, home country/region, and program ranking, among other variables. For brevity, we do not report the estimates for temporary jobs in Table 6. Examining gender, we see that females are 7.6% less likely than males to obtain academic jobs. In contrast, women are more likely to be placed in the government and private sector but the differences are statistically insignificant.¹⁶

<<Table 6 here>>

Moving on to country/region dummies, there is little economical or statistical difference between American students and those from the reference group countries. However, there is substantial variation among international students. Relative to the reference group of all other countries, candidates from China and Korea are less likely to work in academia, while those from Japan and Taiwan are more likely. In columns (2) and (3), those from China and Korea more likely to find private sector jobs, and those from China, India, Russia, and Taiwan are less like to work for the government.

The impact of having an additional Master degree on job type is insignificant, as is that of program tier. Candidates from larger programs are more likely to work for the government and less likely in academia. Our results suggest that candidates with a top-four journal publication are about 32.3% more likely to be placed into academia, followed by government jobs, and are least likely to be placed into the private sector. The same pattern holds for candidates with a publication in journals ranked 5 through 50, but to a lesser extent. Having a top-four R&R reduces a candidate’s likelihood of being placed into private sector, while having a R&R in journals ranked 5 through 50 reduces the likelihood of being placed into government jobs. One might believe that top advisors help their students secure university jobs, but we find no significant effect.

5.3 Linking job type and job location

5.3.1 A theoretical framework

In this section, we offer a stylized model to analyze the link between job type and job location. The key feature of the model is the following. In the foreign market, employers can draw from local labor supply for nonacademic jobs, but need to rely more on U.S.-educated Ph.Ds in academia. That is, there is more local competition for foreign nonacademic jobs which decreases their attractiveness

¹⁶In addition to this baseline specification, we also introduce interaction terms to investigate the gender difference within a specific country and within a program tier. Our main results remain qualitatively unchanged. Results are available from the authors upon request.

and channels more candidates with a U.S.-granted Ph.D. toward foreign academic jobs. Therefore, foreign placements are more likely to be in academia than U.S. placements.

We employ a multi-dimensional Hotelling model to capture this differentiation in terms of job type (academic vs. nonacademic) and job location (U.S. or foreign).¹⁷ Agents in the Hotelling models are typically interpreted as firms and consumers, with firms choosing prices to maximize their profits and consumers choosing which firm to buy from in order to maximize their utilities. Here we interpret firms as employers and consumers as candidates. Without loss of generality, one can interpret the prices firms charge as the opposite of the salaries employers pay to candidates.

Note that employers (firms) are differentiated from each other on two dimensions: job type (academic vs. nonacademic, on the x -axis) and job location (within vs. outside the U.S., on the y -axis). There are four possible combinations or four types of employers: U.S. academic, foreign academic, U.S. nonacademic and foreign nonacademic. As indicated in Figure 4, going clockwise, firm 1 (i.e., U.S. academic) is located at $(0,0)$, firm 2 (i.e., foreign nonacademic) at $(0,1)$, firm 3 (i.e., foreign nonacademic) at $(1,1)$, and firm 4 (i.e., U.S. nonacademic) at $(1,0)$. We restrict each employer (firm) to offer only one job (product). The implication for the empirical analysis in Section 5.3.2 is that we will consider only permanent jobs and exclude temporary jobs.

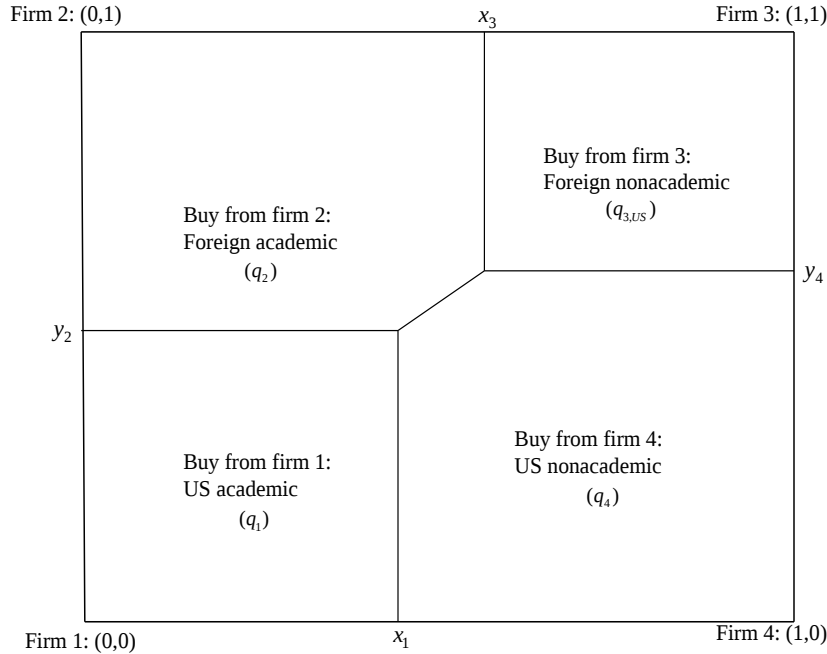


Figure 4: Job types and job location

A continuum of consumers of measure one is uniformly distributed within the square and each consumer is represented by the location (x, y) where x represents job type and y job location. For simplicity, consumer distributions on the two dimensions are independent of each other. On the

¹⁷For studies employing multi-dimensional Hotelling models, see, for example, Gans and King (2006) and Liu and Shuai (2012).

x -axis (i.e., job type), a location closer to 0 means that the consumer is better matched to an academic job, either because of her own preference or because the academic employer prefers her qualifications over other candidates'. As one moves on the x -axis towards 1, the consumer becomes better matched to a nonacademic job. On the y -axis (i.e., job location), a location closer to 0 implies a better match to a U.S. placement while a location closer to 1 implies the opposite. To be exact, a consumer located at (0,0) has the strongest preference for a U.S. academic job, while a consumer located at (1,1) for a foreign nonacademic job. Any consumer away from the four corners will have mismatch with any of the 4 firms, and the mismatch is characterized by a disutility (transport cost in the Hotelling literature) which increases with the distance between the consumer's location and the firm's location on each dimension. In particular, consider a representative consumer located at (x, y) . If she buys from firm i located at (a, b) , she will enjoy an indirect direct utility of

$$U_i = V - p_i - t_1(x - a)^2 - t_2(y - b)^2,$$

where p_i is the price firm i charges and t_j measures consumers' intensity of preference on dimension $j = 1, 2$. For simplicity, we assume that $t_1 = t_2 = t$. Each consumer buys at most one unit from the firm which maximizes her utility. We assume that V is sufficiently large so all consumers buy in the equilibrium (covered market). All firms have constant marginal cost which we normalize to zero. Firms choose prices simultaneously to maximize their respective profits.

Next, we introduce additional firm asymmetry beyond their locations. Specifically, we assume that firms 1, 2 and 4 sell only to consumers in the unit square. However, firm 3 (foreign nonacademic employer) can sell to another market and the demand from that market is represented by $q(p_3)$. The idea is that while foreign academic employers need to rely on the U.S.-granted Ph.D. supply, foreign nonacademic employers can draw from local labor supply who are likely willing to accept lower salary (equivalent to buying at higher price in our model). To simplify the analysis, we assume that $q(p_3) = a - p_3$ with $a > 1$.¹⁸

Our theory gives a testable hypothesis which is summarized in the next Proposition.

Proposition 1 *Foreign placements are more likely to be academic than the U.S. placements.*

Proof. See Appendix A.2. ■

Why are foreign placements more likely to be academic than U.S. placements? Candidates who receive a Ph.D. from the U.S. face more competition for foreign nonacademic jobs (due to the local labor supply), which reduces the attractiveness of nonacademic jobs, giving candidates more incentive to switch to either foreign academic jobs or US nonacademic jobs. Correspondingly, foreign placements are more likely to be academic, relative to U.S. placements. An alternative

¹⁸The intuition is more general than the specific choice of linear demand function. All we need is to allow firm 3 to have different demand elasticities than the other three firms, and the reason is because it also serves a foreign market with stronger demand (equivalent to lower salary).

interpretation is that the criteria for success in academic jobs such as research publication is more universal, and candidates with a Ph.D. from the U.S. have advantages over local candidates. For nonacademic jobs, this advantage is likely less or may even be in the opposite direction. Therefore, foreign placements are more likely to be academic.

5.3.2 Empirics

The above stylized model suggests a link between job type and job location for foreign candidates. Relative to foreign candidates who stay in the U.S., those who leave the U.S. are more likely to be placed into academic jobs. Next, we test this prediction empirically. Since almost all U.S. candidates stay in the U.S., we exclude American students from this analysis. To be consistent with the theoretical model, we consider only permanent jobs and exclude temporary jobs.

Among foreign candidates who stay in the U.S., 128 and 90 are placed into academic and nonacademic jobs, respectively, implying a ratio of $\frac{128}{128+90} \approx 59\%$ for academic placement in the U.S. For those who leave the U.S., the ratio of academic placement is $\frac{87}{87+51} \approx 63\%$. We can see that relative to those who stay in the U.S., those who leave are more likely to be placed into academic jobs. However, the difference is small and insignificant. We then investigate our data more closely for heterogeneity across countries and find that Korean candidates behave drastically differently from those in other countries. The difference in the ratio of academic placements increases significantly if Korean candidates are removed.

Next, we test our hypothesis more formally and offer an explanation for why Korean candidates behave differently from the prediction in Proposition 1. We use Probit estimation where the dependent variable *academicjob* is defined as 1 if a foreign candidate is placed into academia and 0 otherwise. The variable *stayus* is used as an explanatory variable.¹⁹ The results are provided in Table 7. In model (1), for all foreign candidates, those who stay are 2.7% less likely to end up in academia relative to those who leave the U.S., but this difference is insignificant. From our previous descriptive statistics and estimation results, we observe significant heterogeneity across countries. Taking this into account, we introduce country/region dummies in model (2) and both country and country-stay interaction terms in model (3). In model (2), compared to candidates from the regions in the reference group, those who stay in the U.S. are 3.4% less likely to be placed into academia jobs than those who leave, but the difference remains insignificant. Moving to specification (3), the interaction term is positive and significant only for *koreastay*. For Korean candidates, the sum of the coefficients of *stayus* and *koreastay* is $-0.075 + 0.404 > 0$, suggesting that Korean candidates who stay in the U.S. are more likely to be placed into academia.²⁰ Why are the results different for

¹⁹There is potential endogeneity problem using *stayus* as the explanatory variable as unobserved heterogeneity in the candidates may affect both decisions to choose between academic and non-academic and to choose between stay and return. However, with only a single year data, it is difficult to find good instrument variables to account for this endogeneity issue.

²⁰This can be easily confirmed by looking at the raw data of Korean candidates. 11 of the 13 who stay in the U.S. find jobs in academia and only 2 go to the private sector. For those who leave the U.S., 3, 8 and 11 are placed to

Korean students? Anecdotally, we asked some Korean Ph.D. economists we know and were told that for some of the Korean economics Ph.D. students, they were financially sponsored by various entities in Korea and are required to return to Korea after graduation, usually to work for their sponsors. We then exclude Korean candidates and re-run models (1)-(3). The results are listed in (4)-(6) of Table 7. We can see that now the coefficient for *stayus* is negative and significant in models (4) and (5) but remains insignificant in model (6).²¹ Taken together, our estimation results largely support the prediction of our stylized model.

<<Table 7 here>>

5.4 Job rank

Next, we regress job rank on candidates' demographic and academic characteristics. We use the Kalaitzidakis, Mamuneas, and Stengos (2003) ranking (their Table 3) to rank academic jobs from 1 to 200 (lower ranks connote better jobs). Following Krueger and Wu (2000), we give the rank of 40 to top government jobs (e.g., IMF, World Bank, and Federal Reserve Banks) and the rank of 120 to top consulting jobs (e.g., Abt, NBER, and Rand). All remaining jobs are assigned the rank of 300. Placement in a business school is assigned five ranks better, or minus five (Krueger and Wu, 2000). Table 8 reports the estimation results from a Tobit model where the dependent variable is *jobrank*. In these regressions, we focus on candidates who have found a permanent job in one of the three sectors, and exclude those with temporary positions. These criteria lead to a subsample of 511 observations, including Americans and international students.

<<Table 8 here>>

From Table 8, our results in model (1) suggest that relative to male candidates, female candidates' placement worsens by 8.704 ranks but this difference is insignificant. The gender difference increases to 29.502 and becomes significant for the reference group in model (2) where we include gender-country interactions. The estimates for program tiers have the right sign and relative magnitude, i.e., lower tier programs are related to worse placement rank in both models. Comparing models (1) and (2), the most striking feature is probably the sharply different results in their estimates for country dummies. In particular, almost no country/region dummy has a significant coefficient in model (1) while almost all are significant in model (2). In model (1) for example, our results suggest that candidates from China are likely to be placed into worse ranked jobs relative to the reference groups. Note that this comparison considers both male and female in both groups.

academia, government, and private sector, respectively.

²¹We also run multinomial logit estimation similar to that in Section 5.2 but with variable *stayus* and country-stay interaction terms. The results are qualitatively similar to the probit estimation results here. The estimate for *stayus* is always negative but may be insignificant. The country-stay interaction terms are all negative except for Korea. Details are available upon request.

In contrast, results in model (2) provide separate comparisons for male and female candidates. Relative to other male candidates, Chinese male candidates’ placement is worse by 96.775 ranks. On the other hand, relative to female candidates in the reference groups, Chinese female candidates’ placement improves by $128.458 - 96.775 = 31.683$ ranks. The results for other countries are similar, the country fixed effects have different signs for males and females. This is confirmed in model (4) and (5), where we divide the sample into two subsamples by gender and re-run model (1) separately for each sample. We can see that the country fixed effects have mostly different signs between model (4) (female) and model (5) (male). On the other hand, the similarity between results in model (1) and (3) suggests that there is little gender heterogeneity across program tiers.

Summary of main results

In previous subsections, we have presented results on the relationship between job placements and various candidate characteristics. To ease interpretation, we now report key results in Table 9.

	Job location (Stay)	Job type			Job rank ^(a)
		Academic	Government	Private	
Female	+	-	+	+	+
China	+	-	-	+	+
India	+	-	-	+	-
Korea	-	-	+	+	+
U.S.	N/A	-	+	+	+
Russia	-	+	-	+	+
Argentina	+	-	-	+	-
Program rank ^(b)	U-shaped	Non-monotonic	+	Non-monotonic	+
Top advisor	-	+	-	-	+

(a): A larger number means a worse ranked job; (b) Program rank takes the value of tiers 1 – 4.

Table 9. Summary of Main Findings

6 Conclusion

Using data on new Ph.D. economists in 2007-2008, we explore the relationship between candidate demographic/academic characteristics and job placement, as captured by type, location, and rank. Demographic characteristics include gender, nationality, and academic characteristics. Our results suggest that female candidates are less likely to be placed into academic jobs but more likely into government and private sector jobs. On average, female candidates are also placed into worse ranked jobs. We also find significant gender differences in job rank for candidates from some countries. In terms of job location, foreign female candidates are more likely to stay in the U.S. relative to foreign male candidates.

Our results identify significant source country heterogeneity in almost all dimensions of job placement. For example, candidates from Korea are more likely to leave the U.S., while candidates from Argentina and India are more likely to stay in the U.S. We also find that having a top advisor increases the probabilities of academic placement and foreign placement, as well as improving job rank. Relative to tier 1 candidates, those from other tiers are less likely to stay in the U.S. and, not surprisingly, are placed into worse ranked jobs on average.

We also investigate the link between job location and job type. Comparing placements within the U.S. and outside for international students, which placements are more likely to be academic? Our empirical results suggest that, relative to U.S. placements, foreign placements are more likely to be in academia than nonacademia. We develop a stylized theoretical model with the following key feature: the relevant foreign market (for which Ph.D.s from American universities are good match) draws candidates from the U.S. for the academic jobs, but can draw from local talent for nonacademic jobs.

This paper contributes to the literature on the link between job characteristics and individual characteristics. In particular, we reveal new relationships between job location and job type. Thus, our analysis offers a new perspective for potential employers and graduate students and is useful for the profession. We recommend further investigation of how job location (within and outside the U.S.) affects the profession, in particular, how differences in American and other markets impact job placement.

It should be noted that our analysis is based upon placement outcomes from a single-year. Our sample is not fully exhaustive of that particular job market, though a collection of 57 programs is more comprehensive than existing studies. In future work, a more thorough appraisal of initial job placements may include (1) the evaluation of recommendation letters and of teaching abilities and (2) multiple years of data to reflect fluctuations in this job market over time. Furthermore, having multiple years of data would also alleviate endogeneity in linking job type and location by introducing valid instrument variables (IVs).²²

A Appendix

A.1 Variable definitions

(1) Demographic characteristics.

- Gender: We define a dummy variable, *female*, as 1 if the candidate is a female and 0 otherwise.
- Country/region: We define a set of country/region dummy variables as 1 if the candidate

²²For example, the previous year's GDP of each placement country might be appropriate to use as an IV in a multi-year analysis. One would assume that the demand for new Ph.D. economists in a country, to some extent, depends on its GDP from the previous year, especially in the U.S. where these graduates are produced.

comes from that country/region and 0 otherwise. They include *Argentina, China, India, Italy, Japan, Korea, Russia, Taiwan, Turkey, US*.

(2) Academic characteristics.

- Additional Master Degree: We define a dummy variable, *addmaster*, as 1 if the candidate earned a master’s degree prior to entering doctoral training and 0 otherwise.
- Advisor: To control for the effectiveness of recommendation letters, we refer to Tom Coupe’s index of top economists.²³ We define a dummy variable, *topadvisor*, as 1 if the advisor and/or the co-chair is ranked top-50 economists worldwide and 0 otherwise. The underlying assumption is that the more prominent the letter writer is, the recommendation letter will likely yield better job placement (Grove and Wu, 2007).
- Top Publications: We use the journal ranking by Kalaitzidakis, Mamuneas, and Stengos (2003) to define two top journal variables.²⁴ First, following Grove and Wu (2007), we use *top4journal*, defined as 1 if the candidate has at least one of the top four economics journals (i.e., *American Economic Review (AER)* (excluding *Papers and Proceedings*), *Quarterly Journal of Economics (QJE)*, *Journal of Political Economy (JPE)*, and *Econometrica*), and 0 otherwise. There are only 8 candidates with at least one top four journal publication at the time of application in the sample. Second, we define *top5-50j*, which is defined as one if the candidate has at least one publication in a journal ranked between 5 and 50. A total of 29 candidates have at least such a publication.
- Revise and Resubmit (R&R): To be consistent with the variables used for publications, we define two similar quality indicators for R&Rs, using the Kalaitzidakis, Mamuneas, and Stengos (2003) ranking. First, we define a dummy variable, *top4rr*, as one if the candidate has at least one revision and resubmission in a top four journal, and zero otherwise. Second, we define a dummy variable, *top5-50rr*, as one if the candidate has at least one R&R in a journal ranked between 5 and 50. At the time of job application in the sample, 3 of them have a top-four R&R, and 18 have a R&R in journals ranked 5 through 50.
- Teaching Awards: This variable measures the number of teaching awards that the candidate receives while in the Ph.D. program.
- Program Ranking: Following Buchmueller et al. (1999), we define four dummy variables for graduate program tiers. Tier 1 (*tier1*), Tier 2 (*tier2*), Tier 3 (*tier3*) and Tier 4 (*tier4*) refer

²³The list of top 1000 economists is available upon request (Source: <http://student.ulb.ac.be/~tcoupe/update/top1000p.html>. Accessed February, 2009).

²⁴The full list of the journal ranking is available upon request.

to programs ranked 1 to 6, 7 to 15, 16 to 30, and beyond 30, respectively.²⁵ The variable *tier4* is omitted as the reference category in the analysis.

- **Program Size:** We define *size* as the total number of candidates in the program on the same job market. This variable indicates competition within the program. For instance, in larger programs, candidates face more competition from their peers with the same degree credentials. On the other hand, a larger program is more likely to offer candidates a wider network in the job market, likely facilitating their job search.

(3) Placement Characteristics

- **Job Location:** We define two variables regarding job location. *Stayus* is defined as 1 if the student accepts a job in the U.S. and 0 otherwise. *Return* is defined as 1 if the student accepts a job in his/her home continent and 0 otherwise. We only include foreign candidates when considering job location.
- **Job Type:** We divide all jobs into four types: Academia, Government, Private Sector, and Temporary Job (i.e., a visiting position or a postdoc position). *Academia*, *Gov't*, *Private*, *Temp* is defined as one if a person is placed in academia, Government, Private Sector, and Temporary Job, respectively, and zero otherwise.

In Section 5.3.2, *academicjob* is defined as 1 if a foreign candidate is placed into academia and 0 if s/he is placed into the government or the private sector.

- **Job Ranking:** We use the Kalaitzidakis, Mamuneas, and Stengos (2003) ranking to rank academic jobs from 1 to 200 (the lower the number, the better the job is ranked). A placement in a business school is given an extra five ranks, or minus five. Using a methodology similar to the one used in Krueger and Wu (2000), we give the rank of 40 to top government jobs (e.g., IMF, World Bank, and Federal Reserve Banks) and the rank of 120 to top consulting jobs (e.g., Abt, NBER, and Rand). For all the remaining jobs, we give the rank of 300. As discussed in Krueger and Wu (2000), such a ranking system is highly subjective but as we will show later, our findings are robust to different ranking systems. Therefore, candidates with unranked jobs are treated as censored observations, with the censoring point at the rank of 300.

A.2 Proof of Proposition 1

Without the local market which the foreign nonacademic firm also serves, all firms would be symmetric and would choose the same price in a symmetric pure strategy equilibrium. However,

²⁵Some institutions offer economics programs both in the business school and the college of arts and sciences. We treat them as different programs but give them the same rank. The list of all programs included in our sample is available upon request.

asymmetry arises with the introduction of the foreign local market. Suppose that consumers in firm 3 (i.e., foreign nonacademic)'s foreign market are willing to pay higher prices (equivalent to accepting jobs with lower salaries). This means that firm 3 should charge the highest price in the equilibrium. In a partially symmetric (i.e., $p_2 = p_4$) pure strategy equilibrium, it follows that $p_3 > p_2 = p_4 > p_1$. Next, we derive each firm's demand conditional on all firms' equilibrium prices. Notice that in our model, each firm competes directly with two neighboring firms, but competes indirectly with only one non-neighboring firm.

Consider firm 1 for example. First, it competes with firm 4, a neighboring rival. The marginal consumer, who is indifferent between buying from either firm, is characterized by

$$U_1 = U_4 \Rightarrow x = x_1 = \frac{p_4 - p_1 + t}{2t}.$$

Second, firm 1 also competes with firm 2, another neighboring rival, and the corresponding marginal consumer is represented by

$$U_1 = U_2 \Rightarrow y = y_2 = \frac{p_2 - p_1 + t}{2t}.$$

Similarly, firm 3 competes with two neighboring rivals, firms 2 and 4, and the marginal consumers are characterized by

$$x_3 = \frac{p_3 - p_2 + t}{2t}, \quad y_4 = \frac{p_3 - p_4 + t}{2t}.$$

With $p_2 = p_4$, it can be easily verified that $x_1 - x_3 = y_2 - y_4$. There are then two possibilities for which we consider separately: $x_1 \leq x_3$ in the first case and $x_1 > x_3$ in the second case. We verify that no equilibrium exists in the second case. In the first case, from Figure 4, the firms' demand functions are

$$\begin{aligned} q_1 &= x_1 y_1, \\ q_2 &= x_3(1 - y_2) - \frac{1}{2}(x_3 - x_1)(y_4 - y_2), \\ q_3 &= q_{3,US} + q_{3,foreign}, \quad q_{3,US} = x_3 y_4, \quad q_{3,foreign} = (a - p_3), \\ q_4 &= (1 - x_1)y_4 - \frac{1}{2}(x_3 - x_1)(y_4 - y_2). \end{aligned}$$

Firm $i = 1, \dots, 4$'s profit is

$$\pi_i = p_i q_i,$$

and the first-order condition (FOC) is given by

$$\frac{\partial \pi_i}{\partial p_i} = 0.$$

Imposing a partial symmetry in prices ($p_4 = p_2$), we then solve firm 1's FOC to obtain

$$p_1 = \frac{p_2 + t}{3}.$$

and firm 2's to obtain

$$p_2 = \frac{-31t + 9p_3 + 6\sqrt{35t^2 - 9tp_3 - p_3^2}}{13}.$$

Substituting p_1 and p_2 into firm 3's FOC, we have

$$\frac{\partial \pi_3}{\partial p_3} = -\frac{-396t^2 - 72tp_3 + (54t + 51p_3)\sqrt{35t^2 - 9tp_3 - p_3^2} - 21p_3^2 - 169at^2 + 338p_3t^2}{169t^2} = 0.$$

Absence from a closed-form solution for p_3 , we thus rely on numerical results. We first normalize $t = 1$ and then restrict $a \in (1, 3]$ so that firm 3's foreign market is stronger than the U.S. market (lower salary) but not too strong so that firm 3 still makes sales in the U.S. market. The results are qualitatively similar for any $a \in (1, 3]$. Let $a = 2$ and the corresponding equilibrium prices are

$$p_1 \approx 0.5322, \quad p_2 = p_4 \approx 0.5966, \quad p_3 \approx 0.9035.$$

Under these prices, the $x_1 < x_3$ constraint is satisfied since

$$0 < x_1 = y_2 < x_3 = y_4 < 1,$$

Firms' demand are

$$q_1 = 0.2833, \quad q_2 = 0.2983, \quad q_{3,US} = 0.1201, \quad q_4 = 0.2983.$$

In our theory model, q_1 and q_4 are the measures of the U.S. academic placement and the U.S. nonacademic placement, respectively. The ratio of academic jobs in the U.S. placements is $\frac{q_1}{q_1 + q_4} \approx 49\%$. Similarly, the ratio of academic jobs in foreign placements is $\frac{q_2}{q_2 + q_{3,US}} \approx 71\%$. ■

References

- [1] Athey, S., L. Katz, A. Krueger, S. Levitt, and J. Poterba (2007). "What Does Performance in Graduate School Predict? Graduate economics Education and Student Outcomes," *AEA Papers and Proceedings*, 97(2), 512-518.
- [2] Barbezat, D. (1992). "The Market for New Ph.D. Economists." *Journal of Economic Education*, 23(3), 262-276.
- [3] Blackaby, D., A. Booth, and J. Frank (2005). "Outside Offers and the Gender Pay Gap: Empirical Evidence from the UK Academic Labour Market," *Economic Journal*, 115, 81-107.

- [4] Blackaby, D. and J. Frank (2000). "Ethnic and other minority representation in UK academic economics." *Economic Journal*, 110, 293-311.
- [5] Broder, I. (1993). "Professional Achievements and Gender Differences among Academic Economists." *Economic Inquiry*, 31(1), 116-127.
- [6] Buchmueller, T., J. Dominitz, and W. Hansen (1999). "Graduate Training and the Early Career Productivity of Ph.D. economists," *Economics of Education Review*, 14, 65-77.
- [7] Cawley, J. (2011). "A Guide (and Advice) for Economists on the U. S. Junior Academic Job Market."
- [8] Collins, S. (2000). "Minority groups in the economics profession." *Journal of Economic Perspectives*, 14(2), 133-148.
- [9] Ehrenberg, R. (2004). "Econometric studies of higher education." *Journal of Econometrics* 121, 19-37.
- [10] Coupe, T. (2004). "What Do We Know about Ourselves? On the economics of economics," *Kyklos*, 57, 197-216.
- [11] Gallet, C., J. List and P. Orazem (2005). "Cyclicalities and the Labor Market for Economists," *Southern Economic Journal*, 72(2), 284-304.
- [12] Gans, J. and S. King (2006) "Paying for Loyalty: Product Bundling in Oligopoly," *Journal of Industrial Economics*, 54(1), 43-62.
- [13] Ginther, D., and K. Hayes (2003). "Gender Differences in Salary and Promotion for Faculty in the Humanities 1977-95," *Journal of Human Resources*, 38(1), 34-73.
- [14] Grove, W., D. Dutkowsky and A. Grodner (2007). "Survive Then Thrive: Determinants of Success in the economics Ph.D. Program," *Economic Inquiry*, 45(4), 864-871.
- [15] Grove, W. and S. Wu (2007), "The Search for Economics Talent: Doctoral Completion and Research productivity," *AEA Papers and Proceedings*, 97(2), 506-511.
- [16] Hilmer, C., and M. Hilmer. (2007). "Women Helping Women, Men Helping Women? Same-Gender. Mentoring, Initial Job Placements, and Early Career Publishing," *AEA Papers and Proceedings* 97(2), 422-426.
- [17] Kahn, S. (1995). "Women in the Economics Profession." *Journal of Economic Perspectives* 9(4), 193-206.
- [18] Kalaitzidakis, P., T. Mamuneas, and T. Stengos (2003). "Rankings of Academic Journals and Institutions," Working Papers 2003-8, University of Guelph, Department of Economics.

- [19] Kolpin, V., and L. Singell, Jr. (1996). "The Gender Composition and Scholarly Performance of economics Departments: A Test for Employment Discrimination," *Industrial and Labor Relations Review*, 49(3), 408-423.
- [20] Krause A., U. Rinne and K. Zimmermann (2012). "Anonymous job applications of fresh Ph.D. economists." *Economics Letters*, 117, 441-444.
- [21] Krueger A. and S. Wu (2000). "Forecasting Job Placements of Economics Graduate Students," *Journal of Economic Education*, winter, 81-94.
- [22] Liu, Q. and J. Shuai (2012). "Multi-dimensional product differentiation," Working paper.
- [23] McDowell, J., L. Singell, Jr., and J. Ziliak (1999). "Cracks in the Glass Ceiling: Gender and Promotion in the economics Profession," *AEA Papers and Proceedings*, 89(2), 392-396.
- [24] McDowell, J., L. Singell, Jr., and J. Ziliak (2001). "Gender and Promotion in the economics Profession," *Industrial and Labor Relations Review*, 54(2), 224-244.
- [25] McMillen, D., and L. Singell, Jr. (1994). "Gender Differences in First Jobs for economists," *Southern Economic Journal*, 60(3), 701-704.
- [26] Neumark, D. and R. Gardecki (1998). "Women helping women? Role model and mentoring effects on female PhD students in economics." *Journal of Human Resources*, 33(1), 220-246.
- [27] Oyer, P. (2006). "Initial Labor Market Conditions and Long-Term Outcomes for economists," *Journal of Economic Perspectives*, 20(3), 143-160.
- [28] Siegfried, J. and W. Stock (1999). "The Labor Market for New Ph.D. economists," *Journal of Economic Perspectives*, 13(3), 115-134.
- [29] Siegfried, J. and W. Stock (2001). "So you want to earn a Ph.D. in Economics: How Long Do You Think it Will Take?" *Journal of Human Resources*, 36(2), 364-378.
- [30] Siegfried, J. and W. Stock (2004). "The Market for New Ph.D. economists in 2002," *AEA Papers and Proceedings*, 94(2), 272-285.
- [31] Stock, W. and R. Alston (2000). "Effect of graduate-program rank on success in the job market." *Journal of Economic Education*, 31(4), 389-401.
- [32] Stock, W. and J. Siegfried (2001). "So you want to earn a Ph.D. in Economics: How much do you think you'll make?" *Economic Inquiry*, 39(2), 320-335.
- [33] Stock, W. and J. Siegfried (2006a). "Time-to-Degree for the Economics Ph.D. Class of 2001-2002," *AEA Papers and Proceedings*.
- [34] Stock, W. and J. Siegfried (2006b). "Where Are They Now? Tracking the Ph.D. Class of 1997." *Southern Economic Journal*, 73(2), 473-488.

Figure 1. Percent PhD Students Who are Female

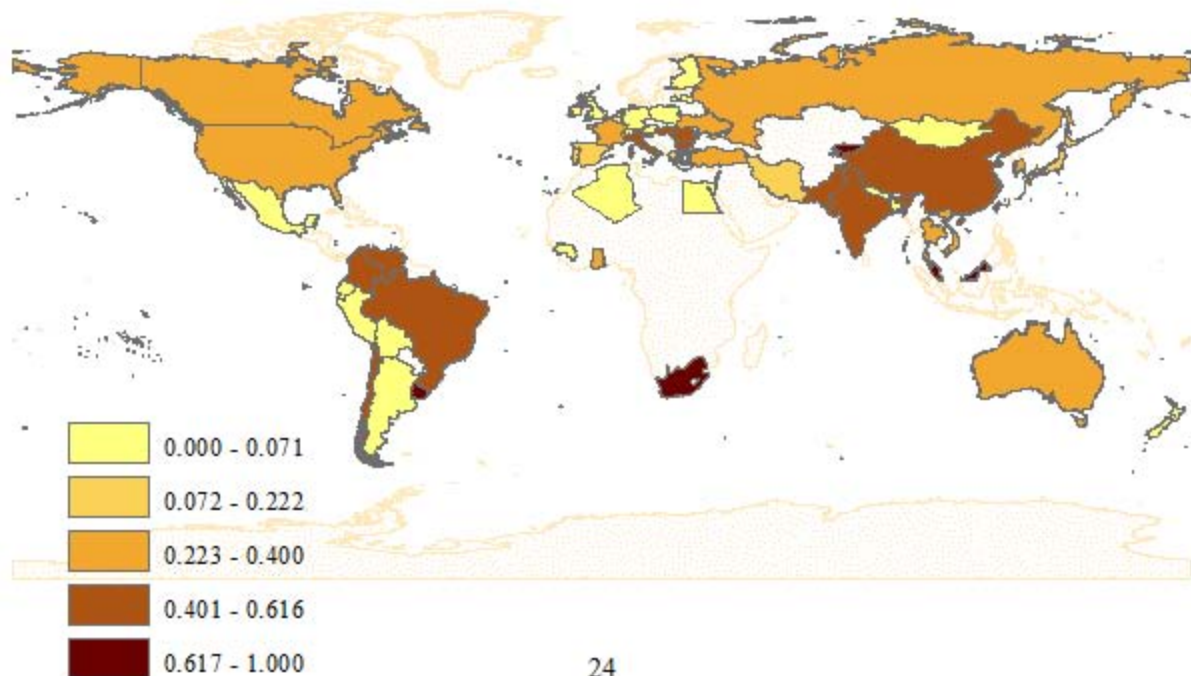


Figure 2. Percent PhD Students Who Stay in U.S.

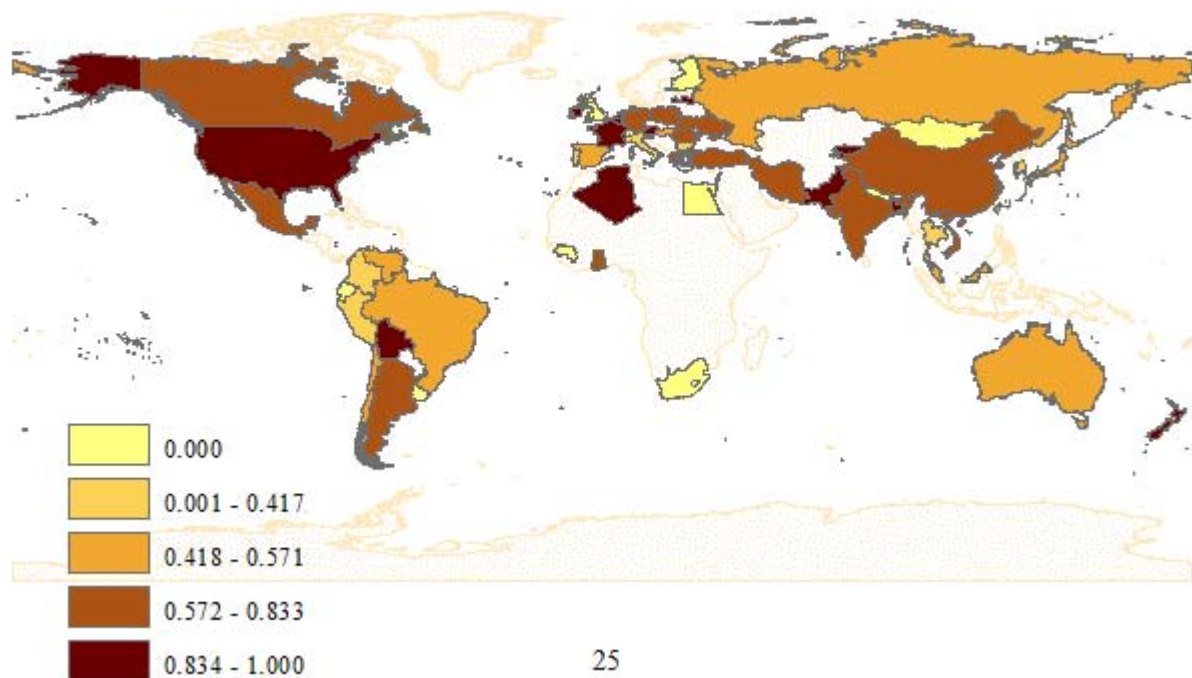


Figure 3. Percent PhD Students Who Return to Home Country

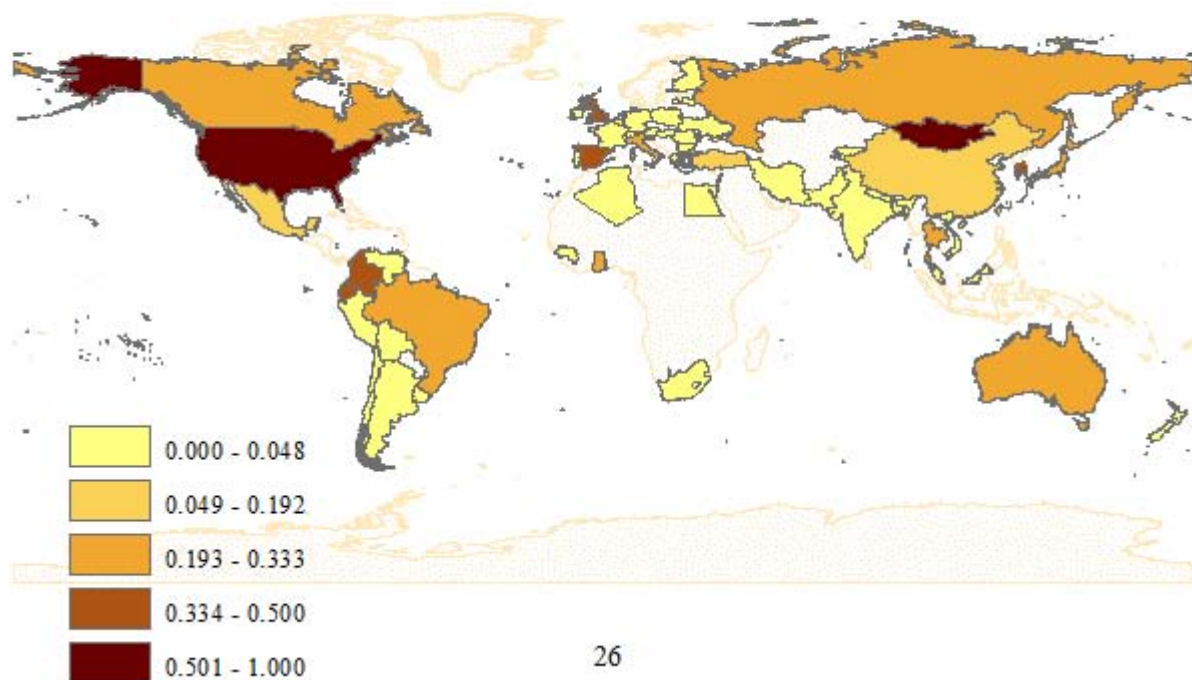


Table 1. Female and Foreign student Ratios by Program Ranking

Program Ranking	Total	Female	Female Ratio	Foreign	Foreign Ratio
Tier1	136	46	34%	83	61%
Tier2	115	40	35%	90	78%
Tier3	174	60	34%	131	75%
Tier4	153	53	35%	105	69%
Total	578	199	34%	409	71%

Table 2. Summary Statistics of Academic Characteristics

Variable	Obs	Mean	Std. Dev.	Min	Max
Addional master degree	578	0.51	0.50	0	1
Topadvisor	578	0.04	0.21	0	1
top4j	578	0.01	0.12	0	1
top5-50j	578	0.05	0.22	0	1
top4rr	578	0.01	0.07	0	1
top5-50rr	578	0.03	0.17	0	1
Teaching awards	578	0.26	0.61	0	4
Size	578	14.19	7.52	1	32

Table 3. Foreign Students' Stay In the U.S. and Return-to-Home Continent Ratio

	Total	Stay	ratio	return	ratio
overall	395	249	63%	113	29%
by gender					
male	254	152	60%	80	31%
female	141	97	69%	33	23%
by program tier					
tier1	83	63	76%	20	24%
tier2	88	52	59%	27	31%
tier3	122	70	57%	42	34%
tier4	102	64	63%	24	24%
by continent					
Africa	7	3	43%	1	14%
Asia	194	117	60%	53	27%
Australia	8	5	63%	2	25%
Europe	111	64	58%	41	37%
Latin Americ	45	26	58%	6	13%
Mideast	36	29	81%	3	8%
Canada	8	5	63%	7	88%

Note: the sum of Stay and Return may not add up to Total. For example, candidates may be placed in a third continent other than their home continent or the U.S.

In addition, Canadian candidates who return to home continent may be placed in either Canada or the U.S.

Table 4. The Job Type of All Candidates

Job Types	Academia		Gov't		Private		Temp		Total
<i>Overall</i>	<i>Count</i>	<i>Percent</i>	<i>Count</i>	<i>Percent</i>	<i>Count</i>	<i>Percent</i>	<i>Count</i>	<i>Percent</i>	
	309	55%	77	14%	125	22%	48	9%	559
<i>Gender</i>									
Female	94	49%	29	15%	52	27%	17	9%	192
Male	215	59%	48	13%	73	20%	31	8%	367
<i>Program Tier</i>									
Tier1	76	56%	21	16%	30	22%	8	6%	135
Tier2	61	54%	14	13%	27	24%	10	9%	112
Tier3	93	57%	22	13%	35	21%	14	9%	164
Tier4	79	53%	20	14%	33	22%	16	11%	148
<i>Continent</i>									
Africa	2	29%	2	29%	2	29%	1	14%	7
Asia	97	52%	20	11%	54	29%	15	8%	186
Australia	4	50%	3	38%	0	0%	1	13%	8
Europe	66	61%	12	11%	19	18%	11	10%	108
Latin America	22	51%	5	12%	11	26%	5	12%	43
Mideast	19	54%	6	17%	4	11%	6	17%	35
North America	99	58%	29	17%	35	20%	9	5%	172

Table 5. Marginal Effects of Probit Analysis on Job Location
(Foreign Candidates Only)

VARIABLES	(1)	(2)	(3)
Female	0.063 (0.052)	0.051 (0.101)	0.065 (0.054)
China	0.132* (0.074)	0.055 (0.109)	0.132* (0.074)
India	0.150* (0.083)	0.251*** (0.095)	0.150* (0.083)
Korea	-0.227** (0.102)	-0.214* (0.113)	-0.227** (0.102)
Russia	-0.100 (0.118)	-0.126 (0.124)	-0.100 (0.119)
Turkey	0.170 (0.105)	0.246*** (0.077)	0.170 (0.105)
Japan	-0.121 (0.144)	-0.152 (0.167)	-0.121 (0.144)
Italy	-0.217 (0.144)	-0.431** (0.167)	-0.217 (0.143)
Taiwan	-0.046 (0.155)	-0.157 (0.220)	-0.047 (0.156)
Argentina	0.308*** (0.044)	0.306*** (0.044)	0.307*** (0.046)
Additional master degree	-0.101* (0.052)	-0.101* (0.056)	-0.101* (0.052)
Ph.D. tier 2	-0.301*** (0.082)	-0.304*** (0.083)	-0.297*** (0.099)
Ph.D. tier 3	-0.279*** (0.084)	-0.276*** (0.087)	-0.279*** (0.084)
Ph.D. tier 4	-0.276*** (0.096)	-0.277*** (0.098)	-0.276*** (0.096)
Size	-0.007 (0.004)	-0.007 (0.004)	-0.007 (0.004)
Teaching awards	0.006 (0.042)	-0.004 (0.045)	0.006 (0.042)
top4j	-0.028 (0.183)	-0.070 (0.174)	-0.028 (0.183)
top5-50j	-0.149 (0.111)	-0.121 (0.113)	-0.149 (0.111)
top5-50rr	0.161 (0.128)	0.131 (0.139)	0.162 (0.129)
Topadvisor	-0.299** (0.143)	-0.293* (0.159)	-0.299** (0.144)
Female*County Dummy	N	Y	N
Female*Program Tier	N	N	Y
Observations	394	394	394
Pseudo R-squared	0.0998	0.113	0.0998

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table 6. Marginal Effects of Multinomial Logit Analysis on Job Type
(All Candidates)

Placement Type	(1) Academic	(2) Government	(3) Private
Female	-0.076* (0.045)	0.020 (0.016)	0.054 (0.047)
U.S.	-0.019 (0.061)	0.004 (0.018)	0.026 (0.055)
China	-0.124* (0.065)	-0.034** (0.016)	0.167** (0.069)
India	-0.027 (0.075)	-0.040** (0.019)	0.070 (0.073)
Korea	-0.209** (0.100)	0.019 (0.034)	0.200** (0.096)
Russia	0.033 (0.091)	-0.100*** (0.011)	0.093 (0.091)
Turkey	0.013 (0.109)	0.015 (0.045)	-0.034 (0.113)
Japan	0.152* (0.086)	-0.034 (0.023)	-0.117 (0.089)
Italy	-0.060 (0.147)	0.004 (0.054)	0.076 (0.140)
Taiwan	0.182* (0.094)	-0.077*** (0.008)	-0.105 (0.095)
Argentina	-0.044 (0.178)	-0.024 (0.039)	0.088 (0.170)
Additional master degree	0.002 (0.040)	0.007 (0.015)	-0.009 (0.035)
Ph.D. tier 2	-0.024 (0.061)	0.003 (0.024)	0.020 (0.062)
Ph.D. tier 3	0.017 (0.049)	0.004 (0.013)	-0.025 (0.049)
Ph.D. tier 4	-0.048 (0.064)	0.018 (0.024)	0.028 (0.065)
Size	-0.005* (0.003)	0.002** (0.001)	0.004 (0.003)
Teaching awards	0.009 (0.044)	-0.007 (0.011)	-0.003 (0.039)
top4j	0.323*** (0.021)	-0.068*** (0.007)	-0.237*** (0.022)
top4rr	0.075 (0.311)	0.156 (0.307)	-0.215*** (0.020)
top5-50j	0.165*** (0.064)	-0.018 (0.026)	-0.145*** (0.056)
top5-50rr	0.037 (0.088)	-0.039* (0.022)	0.026 (0.087)
topadvisor	0.043 (0.069)	-0.008 (0.036)	-0.034 (0.080)
Observations	559	559	559
Pseudo R-squared	0.0691	0.0691	0.0691

Standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1

Table 7. Marginal Effects of Probit Analysis on Job Location and Job Type
(Foreign Candidates Only)

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)
	All Foreign Candidates			Korean Candidates Excluded		
Stayus	-0.027 (0.066)	-0.034 (0.071)	-0.075 (0.093)	-0.134** (0.058)	-0.122** (0.061)	-0.070 (0.090)
Female	-0.092* (0.051)	-0.083 (0.057)	-0.062 (0.055)	-0.098* (0.052)	-0.084 (0.060)	-0.073 (0.059)
China		-0.104 (0.070)	0.073 (0.138)		-0.087 (0.070)	0.075 (0.135)
India		0.033 (0.089)	0.049 (0.185)		0.033 (0.086)	0.046 (0.182)
Korea		-0.206** (0.099)	-0.545*** (0.090)		—	—
Russia		0.063 (0.097)	0.123 (0.133)		0.041 (0.101)	0.114 (0.132)
Turkey		-0.038 (0.124)	-0.203 (0.275)		-0.043 (0.123)	-0.191 (0.281)
Japan		0.213* (0.109)	0.148 (0.198)		0.194* (0.116)	0.142 (0.189)
Italy		-0.113 (0.170)	-0.124 (0.236)		-0.114 (0.168)	-0.133 (0.232)
Taiwan		0.285** (0.112)	0.415*** (0.027)		0.269** (0.106)	0.394*** (0.029)
Argentina		-0.039 (0.202)	0.425*** (0.027)		-0.015 (0.198)	0.404*** (0.029)
China*Stayus			-0.243 (0.168)			-0.236 (0.174)
India*Stayus			-0.037 (0.249)			-0.033 (0.246)
Russia*Stayus			-0.152 (0.304)			-0.154 (0.306)
Turkey*Stayus			0.193 (0.197)			0.173 (0.196)
Korea*Stayus			0.404*** (0.027)			—
Japan*Stayus			0.130 (0.292)			0.118 (0.285)
Italy*Stayus			0.019 (0.237)			0.044 (0.218)
Taiwan*Stayus			-0.647*** (0.025)			-0.673*** (0.025)
Argentina*Stayus			-0.678*** (0.023)			-0.707*** (0.024)
Additional master degree	-0.017 (0.051)	-0.046 (0.051)	-0.029 (0.049)	0.023 (0.052)	-0.003 (0.052)	-0.005 (0.050)
Ph.D. tier 2	0.088 (0.080)	0.094 (0.078)	0.093 (0.085)	0.090 (0.083)	0.100 (0.083)	0.100 (0.083)
Ph.D. tier 3	0.103** (0.048)	0.114** (0.050)	0.143*** (0.054)	0.126** (0.052)	0.134** (0.053)	0.138*** (0.051)

Ph.D. tier 4	0.073 (0.084)	0.082 (0.082)	0.124 (0.085)	0.105 (0.085)	0.117 (0.082)	0.131 (0.083)
Size	-0.001 (0.003)	-0.001 (0.004)	-0.001 (0.004)	-0.000 (0.004)	0.001 (0.004)	0.001 (0.004)
Teaching awards	0.029 (0.066)	0.026 (0.068)	0.021 (0.065)	0.039 (0.068)	0.042 (0.070)	0.038 (0.068)
top5-50j	0.265** (0.105)	0.259** (0.105)	0.228** (0.105)	0.235** (0.106)	0.243** (0.100)	0.218** (0.098)
top5-50rr	0.139 (0.135)	0.161 (0.126)	0.140 (0.126)	0.100 (0.140)	0.124 (0.129)	0.133 (0.124)
Topadvisor	0.177 (0.123)	0.140 (0.125)	0.148 (0.128)	0.185 (0.119)	0.165 (0.115)	0.157 (0.118)
Observations	350	350	350	314	314	314
Pseudo R-squared	0.0283	0.0592	0.122	0.0480	0.0661	0.0777
Robust standard errors in parentheses						
*** p<0.01, ** p<0.05, * p<0.1						

Table 8. Tobit Analysis on Job Rank (All Candidates)

VARIABLES	(1) Full Sample	(2) Full Sample	(3) Full Sample	(4) Female Only	(5) Male Only
Female	8.704 (28.298)	29.502*** (6.542)	17.083 (33.185)		
U.S.	32.257 (34.515)	31.773*** (4.642)	32.590 (34.751)	57.261 (82.412)	21.022 (31.815)
China	22.148 (41.894)	96.775*** (9.128)	22.764 (41.624)	-10.041 (49.672)	82.291 (63.826)
India	-47.377 (41.289)	-92.105*** (10.100)	-47.601 (41.554)	12.014 (66.238)	-96.185* (57.611)
Korea	166.841*** (60.868)	188.096*** (8.542)	167.455*** (60.712)	135.180* (75.643)	178.083** (74.595)
Russia	16.646 (54.567)	42.754*** (12.426)	15.203 (53.824)	-55.416 (119.504)	30.407 (49.450)
Turkey	-69.737 (48.864)	-155.086*** (6.501)	-70.030 (48.277)	1,427.906 (0.000)	-162.910*** (41.858)
Japan	-12.572 (68.464)	42.929*** (5.851)	-14.570 (69.237)	-439.821*** (61.065)	30.703 (60.676)
Italy	127.638 (105.681)	1,379.374*** (46.814)	127.904 (107.391)	3.782 (128.384)	1,348.202 (0.000)
Taiwan	29.150 (61.958)	2.458 (13.959)	25.265 (62.974)	79.190 (97.583)	-9.458 (91.256)
Argentina	-46.654 (60.508)	-41.137*** (5.732)	-51.894 (57.306)		-49.672 (60.266)
Female*Italy		-1,366.203*** (46.814)			
Female*Taiwan		52.041*** (19.550)			
Female*Japan		-488.283*** (16.362)			
Female*China		-128.458*** (12.800)			
Female*India		80.043*** (12.292)			

Female*Korea		-72.804*** (17.397)			
Female*Russia		-77.030*** (16.483)			
Female*Turkey		1,481.327 (0.000)			
Additional master degree	9.900 (26.401)	10.056* (5.875)	10.117 (26.223)	14.190 (52.630)	8.559 (29.846)
Ph.D. tier 2	40.794 (27.714)	39.747*** (3.918)	54.672 (33.736)	-9.447 (43.091)	62.306* (34.843)
Ph.D. tier 3	75.729** (32.889)	65.588*** (5.715)	75.438** (32.750)	15.751 (50.665)	87.680* (46.929)
Ph.D. tier 4	250.914*** (42.007)	253.526*** (4.589)	250.431*** (41.877)	208.535*** (65.236)	269.271*** (55.830)
Size	2.455 (2.010)	2.638*** (0.348)	2.435 (1.997)	0.540 (3.600)	3.441 (2.226)
Teaching awards	3.765 (16.394)	5.800** (2.607)	3.356 (16.549)	22.357 (52.313)	5.065 (17.729)
top4j	-160.698*** (32.006)	-133.473*** (7.817)	-159.989*** (31.384)	71.660 (155.175)	-194.401*** (26.263)
top4rr	1,563.154 (0.000)	1,442.473 (0.000)	1,566.313 (0.000)	1,345.840 (0.000)	1,489.679 (0.000)
top5-50j	34.417 (47.903)	26.836*** (5.259)	35.495 (48.400)	39.317 (120.671)	36.717 (50.958)
top5-50rr	-151.065*** (48.772)	-132.030*** (6.079)	-150.473*** (49.235)	-60.773 (122.268)	-145.740*** (51.746)
Topadvisor	-2.018 (52.242)	-23.209*** (4.010)	-2.194 (52.237)	-62.631 (70.939)	9.253 (69.489)
Female*Program Tier Dummy	N	N	Y	N	N
Constant	152.001*** (48.601)	142.442*** (6.826)	149.650*** (49.993)	219.857*** (73.417)	121.569** (54.851)
Observations	511	511	511	175	336
Pseudo R-squared	0.0236	0.0304	0.0237	0.0272	0.0338

Robust standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1

Gender Differences in Academic Promotion and Mobility at a Major Australian University

Author: Shulamit Kahn

First published: 08 August 2012

<https://doi.org/10.1111/j.1475-4932.2012.00828.x>

Abstract

This paper analyses gender differences in faculty promotion and mobility at the University of New South Wales (UNSW) using personnel records for 1999–2010. It finds female lecturers less likely than men to be promoted or to leave UNSW and female associate professors more likely to be promoted. These results are consistent with a flipping statistical discrimination model. Leave-taking did not account for gender differences. Instead, taking maternity leave indicates higher likelihood of staying at UNSW and being promoted. The paper finds that gender differences narrowed over time but not always as predicted by gender-related policy changes.

The Effect of Time Use in Explaining Male–Female Productivity Differences Among Economists

Authors: Colleen Manchester and Debra Barbezat

First published: 19 December 2012

<https://doi.org/10.1111/irel.12011>

Abstract

This study examines the opportunity cost of non-research responsibilities on research output. Using a sample of early career faculty members who received their Ph.D. in economics, we consider the effect of two dimensions of time use on research output: (1) time allocation, or how time is divided between research and other duties and (2) time concentration, or how research time is distributed during the academic year relative to summer months. This second dimension has not been used in prior studies on research productivity; however, the inherent delays in the publication process as well as start-up costs may imply that concentrating research time exclusively in the summer months reduces research output. We find that both dimensions of time use are significant predictors of peer-reviewed publications and that time concentration is a significant predictor of submissions. We find gender differences in both dimensions of time use, which are attributable to gender differences in employment at research institutions and on-going childcare responsibilities.



The Ivory Ceiling of Service Work

Service work continues to pull women associate professors away from research. What can be done?

By Joya Misra, Jennifer Hickey Lundquist, Elissa Holmes, and Stephanie Agiomavritis

How does a successful associate professor with a distinguished publication record, a visible leadership role among women scientists on campus, and prestigious grant funding for interdisciplinary initiatives in graduate and undergraduate training as well as research feel about seeking promotion to full professor? In the course of our research, we commonly heard, “It feels like the first time in my life that I’m hitting up against the glass ceiling.” Compared with earlier cohorts, women are earning more doctorates, taking more academic jobs, and earning tenure more frequently. But when it comes to promotions to full professorship, our research confirms growing scholarship that women may hit a glass ceiling near the top of the ivory tower.

Men still hold more than three-quarters of full professorships in the United States, and women’s share of full professorships has increased only marginally over the last several decades. Women are less likely ever to be promoted to full professor than men, and their promotions take longer. A 2006 report of the Modern Language Association’s Committee on the Status of Women in the Profession, [Standing Still: The Associate Professor Survey](#), showed that women professors in the association were less likely to be promoted than their male counterparts, and it took women from one to three and a half years longer than men to advance to full professorships, with women at doctoral universities lagging farthest behind.

National data across all disciplines confirm these findings. Indeed, University of Pennsylvania education researcher Laura Perna’s [analysis](#) of [National Study of Postsecondary Faculty](#) data reveals that, nationwide, women at four-year colleges and universities are 10 percent less likely than men to attain promotion to full professor, even after controlling for productivity (career-refereed publications), educational background, institution type, race, ethnicity, and nationality. Perna’s findings suggest that gender differences cannot simply be attributed to men being more productive researchers or having more experience than women. Instead, they point to entrenched institutional practices that may disadvantage women.

What explains the stubborn ivory ceiling of academia? A [study](#) of tenured and tenure-track women faculty members in science, technology, engineering, and math (STEM) fields by Georgia Institute of Technology sociologist Mary Frank Fox and cultural studies and literature professor Carol Colatrella identified ambiguity in institutions’ criteria for promotion to full professor. While most faculty members interviewed could identify the expectations for tenure, fewer could identify the expectations for promotion to full professor. This lack of clarity opens the door for promotion based on vague criteria rather than straightforward expectations.

A related explanation may be found in the way men and women professors divide their time at work—and how colleagues value the division of labor. A variety of studies show that men focus more on research than do women. While men are not necessarily more productive than women, they are more protective of their research time. Tenured women, on the other hand, devote more time to teaching, mentoring, and service, and particularly to activities that may be seen as building bridges around the university. Yet, these pursuits hold less value in promotion cases in many institutions, especially at research-intensive universities.

Faculty Work-Life Study

Our own [research](#)—surveys and focus groups with 350 faculty members at the University of Massachusetts Amherst in 2008–09—explores evidence of a glass ceiling among faculty at a research-intensive university. What we found mirrors other studies: by cohort, women are less likely to be promoted than men, and when they are promoted, the process takes longer. Why might this be? Could women's service to the university, an issue often raised in our focus groups, limit chances for promotion? Our survey asked whether associate professors had served in major administrative positions in their departments—for example, directing undergraduate or graduate programs or working as associate chair or chair. We also examined compensation in the form of course releases and extra pay for assuming administrative roles. We saw few differences in how men and women were compensated. We did, however, find substantial differences by gender in *when* professors and associate professors took on major service roles. Three-quarters of women associate professors, compared with half of their male counterparts, had played such roles. And a third of the women had served as undergraduate directors, compared with 17 percent of the men. Because undergraduate directors spent more time teaching and working with undergraduates—tasks that research universities tend to undervalue—gendered norms may contribute to women associate professors spending more time in devalued roles.

Such norms, however, do not hold for service as chair. Chairing a department is generally viewed as *more* prestigious, reflecting a high level of respect. Yet, not one of the male associate professors who responded to the survey had served as chair, compared with 15 percent of the women. The reverse is true among full professors—35 percent of men have chaired, while only 14 percent of women have done so. Thus, women may be asked, and may agree, to chair their departments earlier in their careers than men.

Does serving in these administrative positions slow progress to full professorship? There were no differences in time to promotion for men and women who served as chairs, associate chairs, or graduate directors. But women who served as undergraduate directors were promoted significantly more slowly; women associate professors who served as undergraduate directors took twelve years (rather than the usual seven) after receiving tenure to be promoted to full professor. By contrast, male undergraduate directors saw no slowing in time to promotion. While, of course, other factors may contribute to being selected as undergraduate director *and* being slow to promotion, there is a clear association. As one female associate professor noted, associate professorship is “the midcareer service gully that we find ourselves taking an extended stay in.”

The Gendered Gully of Service

Our survey asked faculty members at all ranks to summarize their time spent on research, teaching, mentoring or advising, and service. We saw some small gendered differences between lecturers, assistant professors, and full professors—for the most part, men and women in these ranks spent their time similarly, although women spent a little more time on mentoring and service.

However, when we focused on associate professors, we saw remarkable differences. Although associate professors of both sexes worked similar amounts of time overall—about sixty-four hours a week—the distribution of work time varied considerably. Men spent seven and a half hours more a week on their research than did women. Even if these differences in research time occurred only during semesters, not during summer or holiday breaks, this would mean that men spent in excess of two hundred more hours on their research each year than women. On the other hand, women associate professors taught an hour more each week than men, mentored an additional two hours a week, and spent nearly five hours more a week on service. This translates to women spending roughly 220 more hours on teaching, mentoring, and service over two semesters than men at that rank.

Another way to consider this would be to look at the percentage of time spent on research, teaching, mentoring, and service. On average, male associate professors spent 37 percent of their time on research, while women associate professors spent 25 percent of their time on research. While women associate professors spent 27 percent of their time on service, men spent 20 percent of their time on service. This dramatic difference suggests that men focus more on their research, which earns greater prestige and potential for promotion. One associate professor survey respondent reported difficulty balancing research, teaching, and service, commenting, “In reality, only research matters when it comes to tenure and promotion, but service and teaching require lots of time.” Our survey also asked faculty members to differentiate between service to the university and service to the profession. Men and women associate professors spent about the same amount of time on service to the profession (5.4 hours a week), but women spent much more time on service to the university (11.6 hours versus 7 hours). While service to the profession often means more prestigious external visibility, service to the university includes a range of activities necessary to keep the university going but less likely to increase one’s faculty stature.

We found similar patterns when zeroing in on STEM fields, despite the fact that faculty members in these fields, overall, have lower teaching loads and therefore more time to spend on research. On average, STEM men and women associate professors spent less time teaching than associate professors in other disciplines. But STEM men spent significantly more time on research (42 percent) than STEM women (27 percent), while STEM women put in more time mentoring (21 percent compared with 15 percent for STEM men) and performing service (25 percent for women compared with 20 percent for men). Thus, gendered patterns of work distribution are more pronounced when we focus on STEM faculty.

Not Just for the Love of Service

Perhaps women associate professors *enjoy* service more, while men *prefer* research? None of our data fits this interpretation. Overwhelmingly, both men and women faculty members express a preference for research. Most associate professors viewed service as impinging on their time, while they expressed frustration at how service is distributed. As one associate professor noted, “Having good judgment, being thorough, means more work. . . . The reward for good work is more work.” Another argued, “There are faculty [who] earn twice as much as I do, but they are making more work for everyone, since they don’t share in the work that needs to be done.” Sylvia Hurtado and Linda DeAngelo, both research directors with the Higher Education Research Institute at the University of California, Los Angeles, conducted a study further suggesting that a gendered time allocation leads to lower levels of job satisfaction and higher turnover risk among senior women as compared with senior men.

Why do associate professors, both men and women, spend so much time on service activities? In one focus group exchange for our study, an associate professor noted, “Because departments try to shield junior faculty from service, and full professors are usually in a better position to say ‘no’ when asked, associate professors often carry disproportionately heavy service loads compared to their junior and senior colleagues.” Another said, “Some departments have lost a lot of faculty. So their full professor ranks are thin. Even if these departments are hiring assistant professors, the associates are doing most of the service to protect the junior faculty.”

Men and women faculty members consistently noted that when faced with multiple demands, they sacrificed research first. As one associate professor said in response to an open-ended survey question, many faculty members believe they “have to fit research into ‘spare time’ that isn’t consumed by committee meetings, teaching undergrads, and mentoring grad students.” Another associate professor said in a focus group, to wide agreement among the participants, “You do your vocational work on the weekend because the administrative work takes up the week.” In one focus group, a participant said that emphasizing service is “actually very counterproductive for the university. Tenured faculty are seasoned researchers. If they are putting all their time into admin[istration] rather than research it is really terrible for the university.” Another added, “Academia is a whole series of bait and switch. You go to grad school because you are good in college classes and then have to switch and write a dissertation. . . . When you get good, you are asked to do service—something else I have never been trained to do.”

Nevertheless, the work trends differed by gender, suggesting that women felt particularly pressured by the demands of service, mentoring, and teaching. Focus-group participants directly related the pressure on associate professors to do service and the imbalance of service to difficulties in attaining promotion. At the same time, they voiced concern about how service might limit their ability to be promoted. In one focus group, we heard the following exchange among three participants. “There’s a contradiction between the pressure for service at the associate level and the devaluing of service for promotion to full,” one said. “People who do a lot service for their departments and schools have difficulty going up for full because they just don’t get enough time to do their research,” said another. A third responded: “The criteri[on] for promotion is research. Associate professors have time for everything but research.” As a participant noted in another focus group, “Something has to give at some point. [I wish] there was a recognition of these different aspects, different balance for different people . . . and valuing them for that.”

All in all, associate professors, especially the women, seemed to view service primarily as an imposition taking time away from the valued work—research, which earns promotion. One associate professor argued, “As associate faculty . . . we become penalized for doing this service monster.” Another said, “There is a price to pay . . . for the administrative load.” A third chimed in, “Yes, \$13,000,” which was the university’s salary bump related to promotion, and laughter broke out around the room.

If faculty members recognize that service is undervalued and likely to lead away from promotion, why do they do it? Many devote time to service because they see it as vital to the running of the university and believe students will suffer if it is not done. At the same time, faculty members voice bitterness at colleagues who do not share the burden. “I have colleagues who have not seen service in ten years,” one said. A sense of being part of a larger enterprise also led some to do more service than they would have liked. One associate professor noted, “If I set limits, I know it means [other faculty and graduate students] will do that extra work. I feel guilty if I say no.”

Changing the Service Culture

What policies might help alleviate the stress associate professors, and particularly women associate professors, experience? We recommend two policy changes that would require university investment and three that would require shifting the culture of work.

First, at many universities, economic hard times have led to fewer full-time faculty appointments and greater reliance on non-tenure-track instructors, who may also be less involved in faculty governance and leadership. For universities to thrive, they must replace lost tenure-line faculty members, increase tenure-line faculties along with student enrollment, and ensure that all faculty members are involved in and compensated for governance activities. Secondly, we urge that greater resources be focused on mentoring to support promotions to full professorships, particularly for women faculty. We believe that workshops that emphasize the “pathways” to full professorship may be particularly instructive.

Cultural changes also matter. Deans and department chairs or heads need to examine teaching, advising, mentoring, and service responsibilities to ensure that all faculty members pull their weight and are rewarded accordingly. Department chairs should review service, teaching, and mentoring expectations with their department members and ensure that women do not disproportionately carry their departments’ service burdens.

We also believe that cultural changes are needed to stress the value of the work of the professoriate more broadly. Too many faculty members and administrators devalue the importance of “institutional housekeeping,” even though it is crucial for the institution’s ongoing health. Universities need to recognize, reward, and publicize their faculty’s service, mentoring, and teaching accomplishments, in addition to their research accomplishments, and ensure that promotions recognize the wide range of contributions faculty make.

To return to the successful woman colleague with whom we started, what explains the glass ceiling she seems to be hitting? For her, and for countless other women, the ceiling exists not simply because of her gender, but because she is engaged in a broader range of activities, focusing not only on her own research but also on interdisciplinary training and mentoring grants and on building a community of scholars. Clearly, a promotion system that recognizes these contributions—in addition to the value of her research—will support truly healthy, sustainable universities.

Joya Misra is professor in the Department of Sociology and the Center for Public Policy and Administration at the University of Massachusetts Amherst. She studies labor market inequalities and work-family policies. Her e-mail address is misra@soc.umass.edu. Jennifer Hickes Lundquist is associate professor of sociology at the University of Massachusetts Amherst. She is a social demographer who studies racial, ethnic, and gender inequality. Her e-mail address is lundquist@soc.umass.edu. Elissa Holmes is a master’s student in the public policy and administration program at the University of Massachusetts Amherst, and Stephanie Agiomavritis graduated from the university in 2010 with a bachelor’s degree in sociology.

Comments

Many faculty members can identify with the choices presented in this highly informative article. Another pattern is for those who want to duck service is to demonstrate a clear lack of competence and/or the responsibility needed to accomplish a task, which, as the article states, is needed "to keep the university going." This means not showing up for a meeting, oversleeping and missing an exam for students one is supposed to proctor, or never contributing to faculty governance or faculty presentation forums. The excuses can be as lame as car trouble, needy kids, or forgetting the day of the week (all valid excuses at times!) to being so deeply involved in research that one lost all track of time. The pay-off for the (often male) faculty member is that they escape service tasks because they have clearly demonstrated that the university would be better run by others. It works--but it's too risky for most women faculty members to pull off.

L.O.

An unintended consequence of well-intentioned policies of ensuring participation of women on all committees within a university or department is that -- when women are in the minority -- there are constant requests to all of them, but especially those known to be diligent in committee work, to be on more committees. In my own experience, it is indeed the Associate Professors -- knowledgeable enough to be good committee members, close enough to tenure to be disinclined to say no, and not yet able to say no with impunity -- who get the most of these requests, consistent with the findings in this article. It's a tradeoff -- ensuring presence of women in "important" committee roles = more burden on female faculty.

N.M.

[◀ Defending Academic Freedom in the Age of Garcetti](#)

[up](#)

[A Midcareer Feminist Reflection ▶](#)

Gender salary differences in economics departments in Japan

Authors: Ana Maria Takahashi and Shingo Takahashi

<https://doi.org/10.1016/j.econedurev.2011.06.002>Get rights and content

Abstract

By using unique survey data, we conduct a detailed study of the gender salary gap within economics departments in Japan. Despite the presence of rigid pay scales emphasizing age and experience, there is a 7% gender salary gap after controlling for rank and detailed personal, job, institutional and human capital characteristics. This gender salary gap exists within ranks. We find no gender promotion differences. In addition, we find a concentration of the salary gap in public universities and in research oriented universities. Our results show no evidence that the gender salary gap is reducing over time, and reject the hypothesis that females' choice between household work and market activities is responsible for the gender salary gap.

Highlights

► We conduct a study of the gender [salary](#) gap within economics departments in [Japan](#). ► Despite the presence of rigid pay scales there is a 7% gender salary gap. ► This gender salary gap exists within ranks. We find no gender promotion differences. ► There is no evidence that the gender salary gap is reducing over time. ► The choice between housework and market activities is not responsible for the gap.

Introduction

Labor market discrimination is a situation in which an otherwise identical person is treated differently by virtue of that person's gender or race. A large part of the literature on labor market discrimination focuses on the problem of gender salary gap. This paper seeks to explore the issues of gender salary gap among Japanese academic economists.

There is substantial literature that examines the gender salary gap within the academic profession in the US and the UK Broder, 1993, McNabb and Wass, 1997, Ward, 2001, Ginther and Hayes, 2003, Blackaby et al., 2005. Most of these studies found that much of the gender salary gap stems from the gender differences in rank attainment. By using data from Scottish universities, Ward (2001) found that female academics receive 7% lower salary than males, after controlling for various characteristics but without controlling for rank. When rank was controlled for, the estimated salary gap reduced to 3.2% and became statistically insignificant. By using data from the Surveys of Doctorate Recipients in the US, Ginther and Hayes (2003) found negligible gender salary differences within rank, but substantial gender differences in rank attainment.

In contrast to the abundance of literature in the US and the UK, there have been few studies about the gender salary gap within Japanese academia despite a growing public interest in gender equality in Japan. In 1999, the Japanese government enacted *the Basic Law for a Gender-equal Society*. Consequently, in 2000, the Association of National Universities set out an *Action Plan* stipulating that national universities should increase the proportion of female academics to 20% by 2010. In 2008, the Ministry of Education, Sports, Science and Technology (MEXT) announced that it would provide 6 million yen in support to selected universities for each female academic hired. Despite such an interest in achieving gender equality, evidence regarding the presence or the absence of a gender salary gap within Japanese academia is not well established. Fujimura (2002) estimated a salary equation using 648 Japanese academics from the 1992 Carnegie International Survey of the Academic Profession. He found a negative female coefficient for national universities, and a negative but statistically insignificant female coefficient for private universities. However, his model had limited control variables, lacking department dummies (except for the medical department) and other important institutional, job and personal characteristics. Senoo, Matsushige, and Umezaka (2002) also estimated a salary equation for a sample of Japanese academics. They found a negative but insignificant female coefficient. However, their sample was too small (49 academics) and the regression contained too few controls (age and gender only).

Academic salary in Japan is determined by rigid pay scales where salaries are largely determined by experience, age and educational attainment. One might expect that the existence of these pay scales would ensure gender equality in salary. Nevertheless, whether there is a gender salary gap is still an empirical question.

We conduct one of the first and the most detailed study of gender salary differences within Japanese academia by using an original data set that we collected via a mail survey administered in 2008. Our study focuses on academic economists. The data contain detailed personal, job, institutional and productivity characteristics. We first establish that in fact there exists a 7% gender salary gap within Japanese economics departments. To illustrate the reasons for this, we then examine whether the gap is concentrated among particular groups of faculty. In doing so, we focus on the following characteristics: rank, experience, type of university and research orientation. In addition, we examine if the gender salary gap is caused by the females' choice between household work and market activities.

This paper is organized as follows: Section 2 outlines the background information. Section 3 describes the methodology. Section 4 presents the data, and Section 5 presents the descriptive statistics. Section 6 estimates the gender salary gap, and Section 7 disaggregates pay disparities by characteristics. Section 8 discusses the results, and Section 9 concludes.

Section snippets

Compensation scheme in Japanese universities

There are three types of universities in Japan: national, public and private. National universities are established and funded by the central government. Public universities are established by local governments, and funded by both the local and the central governments. Private universities are established by private entities and are financially self-supporting. The total annual salary of academics includes: monthly salary for 12 months, bonuses and allowances. Bonuses are typically provided

Empirical methodology

In the literature, two methods are commonly used to investigate the gender salary gap. On the one hand, there is the Oaxaca–Blinder decomposition Oaxaca, 1973, Blinder, 1973 which requires running separate earning regressions for males and females. On the other hand, one can estimate a salary equation for the pooled sample with a female dummy included. As will be noted in Section 4, our data contain 337 academics out of which 58 are female. We, therefore, decided that the number of females is

Data

The data were obtained from a survey we administered via postal questionnaire. The questionnaire can be found in the online supplemental material. Our survey method follows. First, from the website of the Ministry of Education, Culture, Sports, Science and Technology (MEXT) we obtained an official list of all four-year universities in Japan: 87 national, 76 public and 584 private universities. We decided to exclude women's universities since we were interested in the gender salary gap in

Variables and descriptive statistics

Variables are summarized in Table 2. The dependent variable is the log of annual salary (including the 12-month salary, bonuses and allowances). We did not specifically ask respondents to exclude earnings from other institutions (such as income from teaching in other universities), assuming that respondents will report income from the current institution only. Nonetheless, there is a possibility that such income has been included in the reported annual salary. In Section 6.2, we discuss the

Estimation results

Table 3 reports two sets of OLS estimates, each using different publication measurements. We present the results for each model separately, with rank variables included and excluded. OLS 1 uses the most detailed publication measures, with publications in the US/Europe separated from publications in Japan. When rank variables are included, there is a statistically significant gender salary gap of 7.3%.¹⁴

Disaggregating pay disparities

Results so far show clear evidence that, in aggregate, there is a gender salary gap of about 7%. To illustrate why there is a gap, we examine whether the gap is concentrated among

particular groups of faculty. Although there are many ways to disaggregate the gap, we focus on the following characteristics: rank, experience, type of university and research orientation. In addition, we examine whether the gender salary gap is caused by the females' choice between household work and market

Discussions

We have found that, in aggregate, there is a gender salary gap of 7.3%. Now we ask why there is a gender salary gap. After examining the gender salary gap by characteristics, several explanations have been rejected. First, the aggregate gender salary gap does not arise from a concentration of the gender gap among more experienced workers. Second, the aggregate gap is not caused by a concentration of the gap in private universities or in teaching oriented universities. Third, there is little

Conclusion

By using a data set of academic economists from Japanese universities, we conducted one of the first and the most detailed study of the gender salary gap within Japanese academia. Despite the existence of rigid pay scales that emphasize experience, age and educational attainment, we found a 7.3% gender salary gap after controlling for rank and detailed personal, job, institutional and human capital characteristics. We tested several explanations as to the reasons for a gender salary gap in

Stop the Clock Policies and Career Success in Academia

Authors: Colleen Flaherty Manchester, Lisa M. Leslie and Amit Kramer

The American Economic Review

Vol. 100, No. 2, PAPERS AND PROCEEDINGS OF THE One Hundred Twenty Second Annual Meeting OF THE AMERICAN ECONOMIC ASSOCIATION (May 2010), pp. 219-223 (5 pages)

Published By: American Economic Association

<https://www.jstor.org/stable/27804993>